UNIVERSITY OF EASTERN FINLAND

Semantic Representation of Marine Monitoring Data

by

Markus Stocker

A dissertation submitted in partial fulfillment for the
degree of Master of Science

in the
Faculty of Science and Forestry
Department of Environmental and Biological Sciences

April 2017

Abstract

The large, and arguably increasing, number of sensing devices deployed into Earth's seas and oceans to monitor properties at high temporal and spatial resolution are forming the sensing infrastructure that is designed to generate the observational data required to answer current and future large-scale scientific questions.

As the global infrastructure is being constructed with mostly uncoordinated contributions from numerous research institutions and industrial partners, the infrastructure's interoperability—an essential requirement for its ability to support science—remains an important technical challenge.

We review a number of European, North American, and global collaboratories and observatories in marine monitoring. On the backdrop of the staggering heterogeneity and lack of metadata and data interoperability, we present a technological framework that we argue can address this concern. We demonstrate the application of the technological framework in experiments that integrate sensing data and metadata made available by heterogeneous and distributed resources, describing sensing device types on one hand and observatories acting as platforms for concrete sensing devices on the other hand.

The research question explores if the technological framework can increase the interoperability of metadata about marine observatories and of their observational data.

The work consists of three research objectives. First, we investigate the application of the technological framework for the representation of *metadata* about observatories and attached sensing devices utilized in marine monitoring. Second, we investigate the applicability of the framework for the representation of *observational data* collected from observatories. Third, we investigate the *linking* of metadata and data about these heterogeneous resources.

The small-scale experiments presented here merely scratch the surface of the problem. They involve only a few resources and of these we only consider little exemplary metadata and data. However, the work is already contributing to community discussions and efforts aiming at adoption of Semantic Technologies in marine monitoring.

Acknowledgements

This dissertation completes a challenge embarked back in 2010 while I had just started my postgraduate studies in environmental informatics at the University of Eastern Finland (back then, University of Kuopio). Postgraduate studies at the university required the completion of coursework totalling 60 ECTS. My plan was to complete those during the first year, 2009-2010. This meant that I had spent the better part of the first year in classrooms, learning about environmental science—and only marginally on my postgraduate work.

It turned out that the coursework in biogeochemistry, soil ecology, environmental health, plant ecology, aerosol science was in fact so interesting—and the Finnish approach to education so inspiring—that I started to wonder whether it may be possible to graduate with an additional degree in environmental science—while pursuing a postgraduate degree. The challenge was accepted and, following the completion of my postgraduate studies in environmental informatics last year and transition to the postdoctoral phase this year, I am now finishing up my graduate studies in environmental science.

I has been a long way, as well as an exciting one. As this dissertation shows, I clearly continue to lean toward information technology, albeit much more applied to environmental science than back in 2009. However, my understanding for some of the research—the methods, objectives, and motivation—scientists conduct in environmental science has much improved since 2009. This is thanks to my learning experience at the University of Eastern Finland, where I enjoyed some of the finest courses of my academic career.

Particularly memorable are the theory, laboratory and field courses in biogeochemistry and soil ecology by Emer. Prof. Pertti Martikainen, Dr. Christina Biasi, Dr. Narasinha Shurpali and the biogeochemistry team. With his lectures in aerosol science, Dr. James Smith, as well as colleagues and students in the Aerosol Physics group, have arguably been having the greatest and longest lasting impact on my own work. The list of people and groups I got to know and learned from goes on as many more have contributed to seeding and growing my understanding of interaction ecology, microscopy, environmental health, geographic information systems, exposure assessment.

I like to acknowledge Prof. Mikko Kolehmainen who made it all possible by supporting my postgraduate studies and never suggested that collecting five times more credits than required, or delaying postgraduate graduation by a few years, may be foolish. I am also grateful for his lectures in data-driven methods, machine learning, and time-series analysis which have allowed me to expand into new areas also in computer science.

I also like to acknowledge Dr. Christoph Waldmann for having accepted the role of external reviewer, for a thorough review of this work and valuable comments that have not just improved this dissertation but are also contributing to further thoughts, discussion, and future work. Thanks also to Prof. Kolehmainen for having acted as internal reviewer and for the positive evaluation of this work.

The dissertation could, and perhaps should, have been in experimental environmental science but this collided with the start of my research associate position at MARUM, the Center for Marine Environmental Sciences at the University of Bremen, Germany. Hence, this dissertation doesn't describe the results of months-long laboratory experiments in biogeochemistry but bridges some of my background in information science with marine environmental science.

Contents

List of Figures

Abbreviations

AWI	**A**lfred **W**egener **I**nstitute
CTD	**C**onductivity, **T**emperature, **P**ressure
DOI	**D**igital **O**bject **I**dentifier
EC	**E**uropean **C**ommission
EOOS	**E**uropean **O**cean **O**bserving **S**ystem
EOV	**E**ssential **O**cean **V**ariable
ESONET	**E**uropean **S**eas **O**bservatory **NET**work
FP6	**6**th **F**ramework **P**rogram
GMES	**G**lobal **M**onitoring for **E**nvironment and **S**ecurity
GEOSS	**G**lobal **E**arth **O**bservation **S**ystem of **S**ystems
GIS	**G**eographic **I**nformation **S**ystem
GOOS	**G**lobal **O**cean **O**bserving **S**ystem
IFREMER	**I**nstitut **F**rançais de **R**echerche pour l'**E**xploitation de la **MER**
IOC-UNESCO	**I**ntergovernmental **O**ceanographic **C**ommission of **UNESCO**
LTER	Long Term Ecological Research
ODAS	**O**cean **D**ata **A**cquisition **S**ystems
OGC	**O**pen **G**eospatial **C**onsortium
ONC	**O**cean **N**etworks **C**anada
OOI	**O**cean **O**bservatories **I**nitiative
ORCID	**O**pen **R**esearcher and **C**ontributor **ID**entifier
PID	**P**ersistent **I**dentifier
SOS	**S**ensor **O**bservation **S**ervice
SWE	**S**ensor **W**eb **E**nablement
T-SOS	**T**ransactional **SOS**

Chapter 1

Introduction

The number of instruments deployed into Earth's seas and oceans is large, and arguably increasing. Correspondingly, the number of institutions, people, and projects involved in the development, deployment, and maintenance of marine monitoring systems is large, too.

The proliferation of these systems is explained by the need to better understand marine dynamics, processes and events, and the effects of human activity on marine ecosystems and function. Institutions for marine science and technology research and development—such as the UK National Oceanography Centre (NOC) or the French Research Institute for Exploitation of the Sea (Ifremer)—have long been at the forefront in development, deployment, and maintenance of instrumentation designed to collect data about marine properties, such as water temperature and salinity, current velocity, or depth. The industry has also been having a key role in development and manufacturing of instruments used in marine science.

The many and largely uncoordinated efforts have resulted in a plethora of heterogeneous systems. In this dissertation, systems are observatories. An observatory is an infrastructure that can be utilized for different observational goals and provides some kind of trusted service to users. Observatories employ heterogeneous technologies, from heterogeneous sensing and transmission technologies to heterogeneous technologies utilized to encode, format, and manage transmitted data and metadata. This heterogeneity is currently severely limiting the interoperability between infrastructures, in particular the interoperability of collected data and metadata. The lack of interoperability is fueling redundant development efforts, inefficiencies in scientific workflows, and hinders scientific discovery. In fact, collaboratories continue building own registries and catalogues with information about relevant device models, about monitored features and their properties, or the units of collected observation values. Heterogeneity is also fueling inefficiencies in scientific workflows, as researchers continue spending large amounts of time only to harmonize datasets. Finally, lack of interoperability hinders scientific discovery.

With focus on sensing device based *in situ* measurement, this dissertation surveys a number of European and global observatories in marine monitoring as well as related collaboratories involved in development, deployment, and maintenance of observatories. The survey intends to provide non-expert readers a sense for the vast heterogeneity of existing observatories and collaboratories, including the data infrastructures maintained to manage collected data and metadata.

We propose that Semantic Technologies can address some of the highlighted interoperability concerns. Semantic Technologies are those of the W3C Semantic Web Activity, in particular the Resource Description Framework, the Web Ontology Language, and the SPARQL Protocol And RDF Query Language, as well as related software for creating, persisting, and retrieving RDF data. Key aspect is their support for machine readable descriptions of resources—such as sensing devices, observatories, or collected data—and for publishing, retrieving, and linking such descriptions on the Web. We will discuss Semantic Technologies in detail.

The research question explores if Semantic Technologies can increase the interoperability of metadata about marine observatories and of their observational data.

The dissertation consists of three research objectives. First, we investigate the semantic representation of *metadata* about observatories and attached sensing devices utilized in marine monitoring. Second, we investigate the semantic representation of *observational data* collected from observatories. Third, we investigate the *linking* of metadata and data about these heterogeneous resources.

We thus contribute to ongoing discussions and efforts on adopting Semantic Technologies for the representation of metadata and data about heterogeneous resources used in marine monitoring, and environmental monitoring more generally. Concretely, the dissertation is expected to contribute to developments within ENVRIplus,[1] "a Horizon 2020 project bringing together Environmental and Earth System Research Infrastructures, projects and networks together with technical specialist partners to create a more coherent, interdisciplinary and interoperable cluster of Environmental Research Infrastructures across Europe." As recent presentations and discussions at the 3rd ENVRIweek, Prague, Czech Republic, November 14-18, 2016, suggest, there is notable interest among partners for the approaches discussed in this dissertation. Indeed, enabling linking of metadata and data about heterogeneous resources used in marine and environmental monitoring is an important element toward the "interoperable cluster of Environmental Research Infrastructures across Europe."

The dissertation is structured as follows. Chapter 2 reviews some of the main collaboratories and concrete observatories in marine monitoring. The chapter presents examples for the kinds of sensing metadata and data that are relevant to marine monitoring. The chapter briefly presents some relevant data repositories. Chapter 3 introduces Semantic Technologies. These technologies provide a framework for creating or reusing formal vocabulary relevant for the semantic description of marine monitoring resources, and the management of such descriptions. Chapter 4 describes the conducted experiments and the obtained results. Chapter 5 discusses the results, highlights some of the key strengths and limitations of the approach, presents related work, and proposes some avenues for future work. Chapter 6 concludes with final remarks.

---

[1] http://www.envriplus.eu/

Chapter 2

Marine Monitoring

Marine research based on last century's era ship-based expeditions is "shifting to a permanent presence in the ocean".[1] Seas and oceans are monitored globally, increasingly using permanent observatories capable of continuous measurement of a large number of parameters, employing a wide range of sensing techniques.

Such monitoring is crucial for scientists, and excellence research, as it enables the study of natural processes, both rapid episodic events—such as earthquakes, tsunamis, or episodic releases of methane from the seabed—as well as long-term phenomena—such as variability in marine ecosystems due to warming. The understanding gained from data collected in marine monitoring are of fundamental importance to devise sensible policy.[2] Additional stakeholders in marine monitoring include educators, policymakers, and the general public.

The infrastructure utilized in ocean monitoring consists of more than sensing devices. Instruments are generally attached to some kind of platform, such as moorings. A mooring "consists of up to several kilometres of Kevlar rope, on which various instruments are mounted at certain intervals".[3] Moorings are kept vertically in the water column by buoyant floats attached to the rope. Also common are free-falling systems. Known as benthic landers, the platform is an "unmanned vehicle that falls to the seafloor unattached to any cable, and then operates autonomously on the bottom".[4] When a

---

[1] http://oceanobservatories.org/about/

[2] http://www.esonet-noe.org/About-ESONET

[3] https://www.awi.de/en/science/special-groups/deep-sea-research/technology/moorings.html#c35631

[4] https://www.awi.de/en/science/special-groups/deep-sea-research/technology/free-falling-systems.html#c35600

deployment comes to an end ballast weights are released and the observatory floats back to the surface.

This chapter surveys some of the main collaboratories (Section 2.1) and concrete observatories (Section 2.2) for marine monitoring. Furthermore, we survey some of the key relevant repositories for management and archive of observational data and metadata (Section 2.3). The chapter has no ambition to comprehensively survey collaboratories, observatories, and data repositories. Due to their staggering numbers, a comprehensive survey is arguably very difficult to achieve and is beyond the scope of this dissertation. The description here merely aims at providing readers an introduction to these systems and a sense for their heterogeneity.

2.1 Collaboratories

We begin by reviewing existing collaboratories for marine monitoring, i.e. collaboration networks for scientific research in marine sciences. Collaboratories are primarily enabled by research projects, research infrastructures, or research initiatives. In contrast to projects, research infrastructures and initiatives are designed for longer term operations. Research infrastructures are typically involved in a series of research projects.[5]

The primary aim of collaboratories is arguably to implement and maintain continuous monitoring and collection of data, curate observational data to ensure that persisted observational data meets stated quality standards, as well as supporting the access to, and to some extent the processing of, observational data.

Generally, collaboratories involve experts in roles of scientists, technicians, or students as well as technical infrastructure, including sensing devices, data communication links, and data storage and processing components. Collaboratories often involve numerous partner organizations geographically distributed over several countries and are funded by national or supranational funding agencies, such as the European Commission.

We describe the objectives of each collaboratory and provide some information about their size and operations. The compiled information is primarily drawn from the websites of the collaboratories. For the sake of brevity we only summarize the aspects that are

---

[5] Note that research infrastructures are socio-technical systems. Thus, they are not merely hardware infrastructures, which in this dissertation are observatories.

most relevant to this dissertation. Collaboratories enabled by projects typically include aspects such as project management, outreach activities in conferences and workshops, development of training materials and courses, dissemination of information via the project website or on social media, project evaluation using selected key performance indicators, and sustainability approaches to ensure continuity beyond the project. These are largely omitted in the presentation here. Instead, we focus on information regarding instrumentation and observed parameters as well as technological harmonization and innovation.

We first present collaboratories that focus on the development of next generation sensing devices (NEXOS and SenseNET). We then present European collaboratories (ESONET, FixO3, JERICO-NEXT, AtlantOS, and EMSO). Finally, we describe international collaboratories (GOOS, Argo, ONC, and OOI). In this chapter, quotations without citations are obtained from the corresponding project website, for which we provide the URL accessed at the time of writing.

2.1.1 NEXOS

Monitoring of a changing ocean requires next generation Web-enabled sensors. The NeXOS[6] project aims "to develop new cost-effective, innovative and compact integrated multifunctional sensor systems [that] can be deployed from mobile and fixed ocean observing platforms, as well as to develop downstream services for the Global Ocean Observing System (GOOS), Good Environmental Status (GES) of European marine waters (Marine Framework Strategy Directive) and the European Common Fisheries Policy (CFP)."

Coordinated by the Consorcio para el Diseño, Construcción, Equipamiento y Explotación de la Plataforma Oceánica de Canarias (PLOCAN), Spain, the FP7 project started in October 2013, lasts over 4 years, involves 22 participants from 6 countries, and has a total budget of 8 million Euros.[7]

NeXOS highlights several challenges, specifically sensor reliability, by addressing the biofouling problem to reduce maintenance costs and improve sensor performance; add value to sensor data, by adopting OGC's SWE standards; and the need for multifunctional

---

[6] http://www.nexosproject.eu/

[7] http://cordis.europa.eu/project/rcn/111405_en.html

multiplatform sensor systems. NEXOS supports the “development of a truly integrated and sustainably funded European Ocean Observing System (EOOS) to monitor key ocean processes.”

The project is organized in 11 work packages with various activities of interest to this dissertation. NEXOS engineers cost-efficient and reliable sensor systems and develops a biofouling protection system that is applicable to all developed sensor systems. The project develops a hardware and software architecture to enable interoperable Web access to marine sensors. This is to facilitate the rapid integration of data into portals. It develops, calibrates, tests and optimizes innovative, compact and cost efficient multi-functional sensor systems for optical measurement of marine environmental parameters, including contaminants, dissolved substances such as polycyclic aromatic hydrocarbons and organic matter (matrix-fluorescence sensor), particulate matter, phytoplankton (hyperspectral cavity absorption sensor) and variables relevant to the carbon cycle (Carbon sensor for pH, CO_2 and CH_4). In addition to optical measurement, NEXOS also develops sensor systems for passive acoustics, including software for post-processing of acoustic information, and variables relevant to fisheries management. The project is expected to demonstrate the new sensor systems on selected marine platforms in the Central Atlantic and Mediterranean Sea, in operational scenarios, on mobile and fixed platforms.

2.1.2 SenseNET

SenseNET[8] was a Marie Curie Action in Networks for Initial Training (MC-ITN) aimed at training young investigators and enable significant advances *in situ* sensor technology. The ITN brought together 15 research groups, 2 industrial partners, and provided training for 16 post-graduate students and 1 post-doctoral researcher.

Coordinated by the Natural Environment Research Council (NERC), United Kingdom, the FP7 Marie Curie Action started in June 2009, ended in May 2013, involved 10 participants from 5 countries, and had a total budget of 3.8 million Euros.[9]

SenseNET was organized in 3 work packages with focus on optical sensor development, sensors for chemical monitoring, and issues relevant to infrastructure and interfaces.

---

[8] http://www.eu-sensenet.net/

[9] http://cordis.europa.eu/project/rcn/92237_en.html

Each work package was structured in several research tasks. SenseNET consisted of 17 research tasks.

Optical sensor technology is used to detect inorganic and organic analytes as well as dissolved oxygen and nutrient concentrations, and pH. Oxygen is an important parameter in marine environmental science because “most chemical and biological processes result in changes in dissolved oxygen concentrations.”

For a limited number of important chemical parameters of marine environmental science, sensor prototypes based on wet chemistry and electrochemistry techniques have been developed. However, these sensors are physically too large and power consuming to be used autonomously on a wide range of observatories.

Because sensors have typically been designed for usage on a specific observatory or platform (e.g. buoys, autonomous underwater or remotely operated vehicles), they come with great heterogeneity in technical specifications, sensor interface and communication, and data output formats.

SenseNET has tackled these issues in new sensor technology development, optimization, and harmonization with research tasks assigned to post-graduate or post-doctoral students. Specifically for infrastructure and interface issues, students have developed antifouling coatings (antimicrobial materials, mechanical cleaning, chemical flushing) for long-term sensor deployments; evaluated the viability of selected materials and developed technologies in different environmental conditions, e.g. varying salinity, temperature and pressure; integrated heterogeneous sensor data into GIS systems to support storage, visualization and analysis of spatio-temporal referenced data; and integrated heterogeneous sensors, including data protocols, in platforms such as autonomous underwater vehicles.

2.1.3 ESONET

The European Seas Observatory NETwork[10] (ESONET) is a Network of Excellence that aims at maintaining and expanding a network consisting of “institutions, persons, tools and know-how on deep sea observatories” in order to “promote the implementation and the management of a network of long-term multidisciplinary ocean observatories in deep

---

[10] http://www.esonet-noe.org

waters around Europe.” ESONET involves 12 observatories ranging from the Arctic Ocean to the Marmara Sea. These sites “have been identified and selected for their scientific, technological, and socio economical interests.”

Coordinated by the Institut Français de Recherche pour l’Exploitation de la Mer (IFREMER), France, the FP6 project lasted between March 2007 and February 2011, involved 43 participants from 14 countries, had a total budget of 14 million Euros, and involved approximately 300 scientists, engineers and technicians.[11]

ESONET has scientific, operational, technical, political, and long-term governance objectives. Scientific objectives include to advance our understanding of the interactions between ocean, biosphere and geosphere; the provision of information about global change, natural hazards, and sustainable management; the contribution to Global Monitoring for Environment and Security[12] (GMES) and Global Earth Observation System of Systems (GEOSS) initiatives; and scientific advances in submarine geology and sea ecosystem. Operational objectives include “the comprehension and prevention of seismic and tsunami hazards, the development of physical oceanography and the registration of time series data to enhance ocean monitoring and forecasting.” Technical objectives include hardware infrastructure advances and commercial opportunities. Political objectives include “the protection, conservation and sustainable use of the marine environment.” Long-term governance objectives include financial and expertise continuity for a sustainable observatory network.

The activities of main interest to this dissertation are on data infrastructure and on standardization and interoperability. ESONET underscores concerns such as the heterogeneity of generated data, as well as the heterogeneity in quality control, by distributed observatories; limited data archiving support for selected data types; lack of integration among spatial data infrastructures; insufficient storage and computational resources, including network bandwidth; as well as challenges in long-term preservation and publication of data. ESONET aims at addressing some of these issues by adopting information standards such as SensorML and O&M as well as by organizing data capture, long-term archiving and publication of data, metadata, and data products using infrastructures for marine data management. ESONET also underscores the high

---

[11] http://cordis.europa.eu/project/rcn/84696_en.html

[12] The Global Monitoring for Environment and Security is now Copernicus

level of standardization needed especially in marine monitoring due to hardly accessible underwater observatories. Standardization is argued to enable better testing before deployment. ESONET highlights that standardization efforts are needed especially for sensors, systems engineering, quality assurance, and underwater intervention.

ESONET originally developed the ESONET Yellow Pages[13] aimed at organizing "the information concerning on-the-shelf products for the development and maintenance of Deep-Sea Observatories [including] a range of equipments, from simple, isolated sensors or parts, to communication systems or even integrated Observatories."

2.1.4 FixO3

The Fixed point Open Ocean Observatory network[14] (FixO3) is a European project that aims at integrating European marine observatories, harmonize them, and improve access to oceanographic data for the scientific and non-scientific communities.

Coordinated by the National Oceanography Centre, UK, the project started in September 2013, lasts over 4 years, involves 29 partners from 12 countries, and has a budget of 7 million Euros.[15]

The network consists of 23 observatories, located in the Mediterranean Sea, the Atlantic Ocean and the Arctic and Antarctic regions, with a total of over 400 instruments. The network collects data for parameters ranging from physical to biochemical, spanning from shallow infrastructure (20 m) to deep sites ($>$ 5000 m) located at hundreds kilometers from land.

The project is organized in 12 work packages with various activities of interest to this dissertation. FixO3 collects and synthesizes technical information about the hardware, software and middleware of fixed observing sites and creates a knowledge base. The project evaluates existing and novel sensor technology and assesses their applicability for fixed stations. Furthermore, it develops and enhances measurement techniques, in particular for *CO2*, pH, and sound. On software infrastructure, FixO3 develops the Open Ocean Observatories sensor registry using the OGC SensorML standard and integrates the ESONET Yellow Pages to streamline sensor registration by linking Yellow

---

[13] http://www.esonetyellowpages.com

[14] http://www.fixo3.eu

[15] http://cordis.europa.eu/project/rcn/110489_en.html

Pages sensor metadata with deployment site reference data. The project harmonizes data management standards and workflows. It also develops a marine observatory profile for the OGC SOS standard in order to address the heterogeneity of SOS servers. Furthermore, the project develops a platform to display open ocean fixed-point data on the project website. It provides real-time, near real-time and delayed mode access to data products as well as derived information, whereby real-time data are for operational oceanography and near real-time, and delayed mode, data are used in both scientific research and in numerical modelling.

The overall network has various scientific objectives, determined by the individual observatories and the characteristics of their location. Objectives include investigating deep or intermediate convection phenomena; monitoring water masses interactions; monitoring deep-ocean animal community and biomass production; collect data for climate and greenhouse gas studies; collect data for investigating dust impacts on marine ecosystems; studying seismicity and ground deformation; act as a platform for testing new instruments.

2.1.5 JERICO-NEXT

With focus on coastal observatories of a European Ocean Observing System, JERICO-NEXT[16] strengthens and enlarges "a solid and transparent European network in providing operational services for the timely, continuous and sustainable delivery of high quality environmental data and information products related to marine environment in European coastal seas."

Coordinated by Institut Français de Recerche pour l'Exploitation de la Mer (IFREMER), the project started in September 2015, lasts over 4 years, involves 33 partners from 15 countries, and has a budget of 10 million Euros.[17]

The network consists of 35 facilities—consisting of observatories, buoys, and gliders as well as numerous ferry boxes—located in 9 countries. Each facility operates a number of instruments. The network collects data for a wide range of parameters, including classical temperature, salinity, depth, conductivity, pH as well as seabed vibrations, chlorophyll fluorescence, wave characteristics.

---

[16] http://www.jerico-ri.eu/

[17] http://cordis.europa.eu/project/rcn/194965_en.html

The project is organized in 9 work packages with various activities of interest to this dissertation. JERICO-NEXT analyses the adequacy of present observation strategies to meet key scientific and societal challenges in the coastal ocean. It harmonizes technologies and methodologies to provide uniform access modes and interfaces. The key technology areas are fixed platforms, ferryboxes, and gliders. The network develops best practices for processing and analyzing data generated by new sensor technologies, such as HF-radar for surface currents and sea state monitoring, and for the application of cabled networks to coastal observatories. JERICO-NEXT develops, harmonizes, deploys, and evaluates existing and new sensor technology, in particular for biogeochemical measurements including for nutrients, e.g. nitrate; biological parameters, e.g. phytoplankton detection, classification and characterization; marine carbonate system, e.g. pCO_2, pH, CO_3-ion, alkalinity; assessment of the structure of benthic communities and benthic biogeochemical processes, specifically organic matter remineralization at the sediment-water interface. It integrates data from multiple components of coastal observing systems, such as surface current with water column data. It develops techniques to profile coastal waters, e.g. for internal waves, temperature, oxygen. It develops sensors for molecular (qPCR analysis of seawater DNA samples) detection of phytoplankton, harmful algae blooms, and pollutants (through their effect on microorganisms). The project determines how data will be made accessible, the data policy, INSPIRE compliance, data citation, data discovery; harmonizes sensor data and improves best practices for quality control; and improves the management of metadata about deployed observatories and standardized data management, from real-time to validated datasets.

2.1.6 AtlantOS

AtlantOS[18] is a research and innovation project “that proposes the integration of ocean observing activities across all disciplines for the Atlantic, considering European as well as non-European partners”. Its vision is “to obtain an international, more sustainable, more efficient, more integrated, and fit-for-purpose system” for Atlantic Ocean observation. The project aims at delivering “an advanced framework for the development of an Integrated Atlantic Ocean Observing System” (IAOOS).

---

[18] https://www.atlantos-h2020.eu/

Coordinated by the GEOMAR Helmholtz Centre for Ocean Research Kiel, Germany, the project started in April 2015, lasts over 4 years, involves 62 partners from 18 countries, and has a budget of 20 million Euros.[19]

The project is organized into 11 work packages with various activities of interest to this dissertation. AtlantOS analyzes the IAOOS high-level requirements and identifies gaps in existing Atlantic observing networks. It improves, expands, integrates and innovates ship-based observations for "high-quality, high spatial and vertical resolution measurements of a suite of physical, chemical, and biological parameters over the full water column on an ocean basin scale." It improves the systematic collection of ocean observations recorded *in situ* using autonomous ocean observation technologies (e.g. fixed moorings, drifters, gliders and profiling floats). The project integrates *in situ* autonomous ocean observation with other observational systems, e.g. remote sensing. It links ocean observing activities to initiatives in coastal ocean observing and integrates activities beyond the local scale toward regional observing systems. AtlantOS harmonizes workflows, data processing and distribution, and develops products for decision support and resource assessment to address issues of societal concern.

AtlantOS aims to enhance data acquisition capabilities; provide consistent standards for measurement; compile best practices on sampling techniques and data processing; increase spatiotemporal coverage; ensure timely and free data delivery to stakeholders, in particular the science community; improve access to, usability and integration of, data systems; test new techniques and instruments; improve the integration of observational data; identify societal needs relevant to ocean observation.

Of specific interest to AtlantOS are data on essential ocean variables, including temperature, salinity, currents, transient tracers and carbon parameters. Further data are on plankton abundance and biomass data; fish biomass, distribution and community structure; as well as sea floor mapping.

---

[19] http://cordis.europa.eu/project/rcn/193188_en.html

2.1.7 EMSO

The European Multidisciplinary Seafloor and water-column Observatory,[20] EMSO, is a European-scale distributed research infrastructure of seafloor & water-column observatories. Its scientific objective is "long-term monitoring, mainly in real-time, of environmental processes related to the interaction between the geosphere, biosphere, and hydrosphere, including natural hazards."

The EMSO research infrastructure includes a network of observatories "deployed at key locations in European seas, from the Arctic to the Atlantic, through the Mediterranean to the Black Sea." Currently, the network consists of eleven deep-sea observatories and four shallow water test observatories.

EMSO is being developed by EMSODEV, the EMSO Development project expected to "catalyse the full implementation and operation of [EMSO] through the development, testing and deployment of an EMSO Generic Instrument Module (EGIM)." The module is expected to support measurement of "a specific set of variables at any EMSO site irrespective of depth and environmental conditions" and thus "ensure accurate, consistent, comparable, regional scale, longterm measurements of ocean parameters" and increase the "interoperability of EMSO [observatories] thanks to the harmonized collection of ocean essential variable time series." Measured variables include temperature, pressure, salinity, dissolved oxygen, turbidity, chlorophyll fluorescence, currents, and passive acoustics. Novel sensors (e.g. for pH, pCO_2, and nutrients) will also be considered.

Coordinated by the Istituto Nazionale di Geofisica e Vulcanologia, Italy, EMSODEV started in September 2015, lasts over 3 years, involves 11 partners from 9 countries, and has a budget of 4.5 million Euros.[21]

EMSODEV is organized into 8 work packages with with various activities of interest to this dissertation. EMSODEV defines the technical requirements and evaluates the performance of the EGIM. The module should easily connect to any EMSO observatory, cabled or stand-alone, pelagic or seafloor. The project focuses on innovative technologies, methodologies and approaches, as well as on quality control, harmonization, and standardization. It implements and tests the EGIM prototype and driver software. Tests

---

[20] http://www.emso-eu.org/

[21] http://cordis.europa.eu/project/rcn/197997_en.html

for quality, operability and stability are performed against various environmental conditions and operational configurations, primarily at a shallow water test site on a cabled observatory. The project considers the extendibility of the EGIM to support adding more than the 6 basic sensors, and deploys the EGIMs at several EMSO observatories to demonstrate their operational capability *in situ*. It also defines, models and implements the EMSO Data Management Platform (DMP) following the ENVRI Reference Model (Chen et al., 2013). The DMP will "ingest, consolidate, process and archive data coming from the [EGIM], integrate the data management architectures of regional EMSO [observatories] and will make data available to the EMSO portal and to other international initiatives."

2.1.8 GOOS

The Global Ocean Observing System[22] (GOOS) "is a permanent global system for observations, modelling and analysis of marine and ocean variables to support operational ocean services worldwide". The observing system aims to provide a global view of the body of water that is the global ocean system. The system provides "accurate descriptions of the present state of the oceans, including living resources" and "continuous forecasts of the future conditions of the sea". It also supports monitoring, understanding and predicting weather and climate, the "management of marine and coastal ecosystems and resources," mitigation of "damage from natural hazards and pollution," and enables scientific research.

GOOS is a Programme of the Intergovernmental Oceanographic Commission of UNESCO (IOC-UNESCO),[23] the organization for marine science within the UN system established in 1960, and is the oceanographic component of GEOSS. The observation systems serves a number of stakeholders, including oceanographic researchers and agencies, marine and coastal industries, policy makers and the general public.

GOOS develops by executing focused and finite lifetime development projects. It is thus a system of programmes. Such projects may aim at sustaining and strengthening the mature aspects of the observing system or at expanding the system into new areas. The development of GOOS is guided by a Framework for Ocean Observing (Lindstrom et al.,

---

[22] http://www.ioc-goos.org/

[23] http://www.ioc-unesco.org/

2012) that supports "the various ocean observing communities in establishing requirements for a sustained global ocean observing system, the essential variables (EOVs) to be measured, the approach to measuring them, and how their data and products will be managed and made widely available."

As summarized by Lindstrom et al., the key framework concepts are: to "deliver an observing system that is fit for purpose"; to "apply a systems approach for sustained global ocean observing"; to "recognize and develop interfaces among all actors in the Framework for their mutual benefit"; and to "provide the basis for and promote transformation of observational data organized in EOVs into information (syntheses, analyses, assessments, forecasts, projections, and scenarios) that serve a wide range of science and societal needs, and enable good management of the human relationship with the ocean."

GOOS consists of many observatories, including the almost 4000 Argo floats (Section 2.1.9), 1250 drifting buoys, 350 embarked systems on commercial or cruising ships, 100 research vessels, 200 marigraphs and holographs enabling tsunami detection, 200 moorings in open sea used as long-term observatories. Together, these observatories measure a wide range of parameters, including temperature, salinity, currents, atmospheric pressure, oxygen and carbon dioxide in ocean and atmosphere, as well as physical, chemical and biological parameters.

EuroGOOS[24] is a pan-European ocean observing network registered as international non-profit association and operating within the context of GOOS. The association consists of 40 members from 19 European countries and coordinates five regional operational systems. EuroGOOS provides operational oceanographic services and supports marine research. Working groups and networks of marine observing platforms "deliver strategies, priorities and standards towards an integrated European Ocean Observing System (EOOS)."

2.1.9 Argo

Argo[25] is a global array of currently almost 4000 "free-drifting profiling floats that measures the temperature and salinity of the upper 2000 m of the ocean".[26] The array

---

[24] http://eurogoos.eu/

[25] http://www.argo.net/

[26] http://www.argo.ucsd.edu/

"allows, for the first time, continuous monitoring of the temperature, salinity, and velocity of the upper ocean, with all data being relayed and made publicly available within hours after collection." The array provides 100,000 profiles per year and most profiles consist of about 200 data points, or more depending on their communication bandwidth. The floats are distributed roughly every 3 degrees (300 km) and cycle to 2000 m depth every 10 days, with 4-5 year lifetime for individual instruments.

Argo began in 1999, with deployments since 2000, and is dedicated to "greatly improve the collection of observations inside the ocean through increased sampling of old and new quantities and increased coverage in terms of time and area." Increased spatio-temporal coverage are needed to improve the development and validation of climate models, which ultimately "guide international actions, to optimize governments' policies and to shape industrial strategies." To maintain the array, today Argo deploys floats at the rate of about 800 per year.

Argo floats perform 'park and profile' missions, whereby that at typically 10 days intervals floats rise from 2000 m depth to the surface over about 6 hours while measuring temperature and salinity. At the surface, satellites determine their position and transmit data. Floats then sinks to approximately 1000 m, where they drift. For the next cycle, they first descent to 2000 m and then begin a new profile. Each float makes about 150 such cycles.

Argo maintains two separate data streams: real-time and delayed mode. The real-time data stream delivers 90% of profiles as quality controlled data to users within 24 hours. The delayed mode data stream includes an additional quality control system.

Argo data and metadata can be obtained via Global Data Assembly Centers (GDACs). These "offer access to the complete Argo data collection, including float metadata, detailed trajectory data, profile data and technical data all in NetCDF format." Data files can be retrieved via HTTP or FTP. HTTP access merely enables browsing the directory structure, which is equivalent to the FTP directory structure. In addition, the Coriolis GDAC interface[27] provides Web-based search functionality. To support citing Argo data, DOIs are assigned to monthly snapshots.

---

[27] http://www.argodatamgt.org/Access-to-data/Argo-data-selection

2.1.10 ONC

Ocean Networks Canada[28] (ONC) was created in 2007 to "manage the cabled observatory being constructed on the sea floor off the [British Columbia] coast" to monitor "the west and east coasts of Canada and the Arctic to continuously gather data in real-time for scientific research that helps communities, governments and industry make informed decisions about our future".

ONC collects "long-term, continuous scientific data from the ocean environment." The data are "made available through Oceans 2.0—a powerful online data management system." Combined with high-performance computing, Oceans 2.0 "allows ONC to provide ocean analytics that assist researchers, communities, industry, and policy-makers in making evidence-based decisions in Canada and globally."

ONC's scientific objectives cover four themes, namely understanding human-induced change in the Northeast Pacific Ocean; life in the environments of the Northeast Pacific Ocean and Salish Sea; interconnections among the seafloor, ocean, and atmosphere; and seafloor and sediment in motion. Each theme "poses several key scientific questions, describes why each question is important, and explains how Ocean Networks Canada can contribute to answering the question." For instance, the key questions and aims surrounding the first theme are to quantify the rate of possibly accelerating changes in the "timing, intensity, and chemical properties of upwelled [deeper] waters, nutrient availability, and primary production," how ecosystems respond to increasing acidification, or ecosystem response to changes in oxygen availability.

ONC supports research projects, including in arctic studies, in particular physical and biogeochemical processes; the variation and dynamics of gas hydrate processes and related benthic communities; marine sediment ecosystems, studied using camera observations, interactive sampling with sediment traps, as well as data from sensors monitoring temperature, oxygen, and nitrate; the acoustic ecology of marine fauna and the effect of underwater sound from human activities on their physiology and behaviour; detect tsunamis using bottom pressure recorders.

The Innovation Centre division within ONC has developed a "suite of products and services called Smart Ocean Systems$^{\text{TM}}$." The suite includes advanced ocean sensor and

---

[28] http://www.oceannetworks.ca/

instrument technologies as well as undersea power and communication technology, the Oceans 2.0 “digital infrastructure and data management systems for sensor interfacing, data capture, storage and archiving, manipulation and annotation, and internet presentation,” and Ocean Analytics as a “new way to distill and exploit the vast amount of data [from] cabled observatories.”

ONC collects and curates data from the Arctic, Atlantic and Pacific Oceans. Each observatory hosts numerous scientific instruments, connected to the Internet for near real-time data collection. Similarly to FixO3 and the ESONET Yellow Pages, ONC provides registries with metadata about sensing device types and individual sensing devices hosted by observatories. Furthermore, similarly to other collaboratories surveyed here, ONC provides data for download and tools, e.g. for data visualization.

2.1.11 OOI

The Ocean Observatories Initiative[29] (OOI) is a United States National Science Foundation funded “integrated infrastructure project composed of science-driven platforms and sensor systems that measure physical, chemical, geological and biological properties and processes from the seafloor to the air-sea interface.” The collaboratory is designed to “address critical science-driven questions that will lead to a better understanding and management of our oceans, enhancing our capabilities to address critical issues such as climate change, ecosystem variability, ocean acidification, and carbon cycling.”

At the time of writing, the OOI infrastructure involves 89 platforms hosting over 830 instruments providing over 100,000 data products. Data can be accessed via a Web-based data portal.[30] The portal lists the available OOI Arrays and individual sites within Arrays. Selecting an individual site shows information about the site on a map. Information includes coordinates as well as metadata about the instruments hosted by the site. The portal also supports plotting and downloading data, as JSON, CSV, or NetCDF.

In addition to the data portal, OOI also maintains a raw data archive,[31] i.e. data as they are received directly from the instrument. Raw data are in instrument-specific format

---

[29] http://oceanobservatories.org/

[30] http://ooinet.oceanobservatories.org

[31] https://rawdata.oceanobservatories.org/files/

and may contain data for multiple sensors, be in native sensor units, or have processing steps already performed within the instrument. The data is not archived indefinitely. Rather, as disk utilization reaches a given threshold, older data is purged. The archive is organized in an elaborate directory structure with an own vocabulary for folder names. OOI also derives over 200 unique data products using software available in a repository on GitHub.

2.2 Observatories

Observatories are hardware and data infrastructures, in particular platforms that host, primarily, sensing devices and hardware for data communication as well as data infrastructures (permanently) linked to them. They are deployed *in-situ* into a marine environment and observe one or more parameters over time at fixed location or along a horizontal or vertical trajectory.

This section describes a number of observatories of various type, in particular fixed point, coastal, seafloor, water column, and open ocean. We concisely describe some of the main characteristics of *selected* observatories. The purpose here is not to provide a comprehensive survey of existing observatories. Due to the large number of observatories, this would be a major effort and not useful here, as we merely aim at providing readers an overview. Furthermore, there is considerable overlap among observatories, in particular in regard to monitored parameters.

Naturally, observatories may be classified into multiple categories. Indeed, a coastal observatory may be deployed on the seafloor or an open ocean observatory may monitor the water column. The Expandable Seafloor Observatory OBSEA is an example coastal *and* seafloor observatory. Furthermore, in some cases it is unclear whether a platform hosting one or more sensing devices is an observatory or part of a larger system, which may be better characterized as being the observatory. This is the case of Argo, where a single float is arguably merely a subsystem of a larger system. Finally, observatories may be part of multiple networks, and receive funding through serveral projects.

2.2.1 Fixed Point

ANTARES is a multidisciplinary permanent marine observatory of type Cabled Multiple Arrays located in the Ligurian Sea nearby the continental slope and operated by IFREMER. It is part of the Mediterranean Ocean Observing System for the Environment (MOOSE) network, was first deployed on January 21, 2014, and hosts 17 instruments. The observatory monitors physical and biogeochemical parameters and provides real-time data transmission through two deep cabled moorings. Monitored parameters include ground motion, currents, dissolved oxygen, conductivity, temperature, and pressure.

The Ocean-meteorological buoy Augusto Gonzlez de Linares (Biscay AGL) is a fully-equipped Ocean Data Acquisition Systems (ODAS) buoy of type Single Array located in the Southern Bay of Biscay, North of Santander and operated by the Instituto Español de Oceanografía. It was first deployed on July 27, 2007, and hosts 9 instruments. The observatory monitors meteorological, physical, biogeochemical, and ecological parameters and provides real-time data transmission. Monitored parameters include wind speed and direction, relative humidity, wave height, dissolved oxygen, water temperature, conductivity, chlorophyll, currents, atmospheric pressure at sea level, and air temperature.

The Central Irminger Sea observatory (CIS) is a multidisciplinary mooring of type Single Array located in Central Irminger Sea, Subpolar North Atlantic operated by the GEOMAR Helmholtz Centre for Ocean Research Kiel. The site is of interest for the deep winter mixed layer depth. It was first deployed on September 21, 2002, and hosts 20 instruments. More recently, the observatory was redeployed at a slightly different location to accommodate a configuration that includes an OOI observatory. CIS monitors meteorological, physical, and biogeochemical parameters, including air temperature, currents, dissolved oxygen, conductivity, temperature, and pressure.

The Southern Adriatic Interdisciplinary Laboratory for Ocean Research (E2-M3A) is a deep-sea and continuous monitoring station that provides the longest oceanographic time series in the Eastern Mediterranean Sea. The observatory is of type Multiple Arrays composed by two moorings, a surface buoy and a subsurface mooring line. It is operated by the Istituto Nazionale di Oceanografia e di Geofisica Sperimentale. E2-M3A monitors physical and biogeochemical processes in the water column. Specifically, the surface buoy monitors air and sea meteorological and physical parameters in the surface layer. The subsurface mooring line monitors currents, conductivity, temperature, pressure, turbidity, and dissolved oxygen. The observatory also hosts optical sensors. Furthermore, biochemical sensors monitor CO_2, pH, pCO_2. The observatory was first deployed on November 16, 2006, and hosts 14 instruments.

The Frontiers in Arctic marine Monitoring (FRAM) is an array of moorings for multidisciplinary long-term observations across the Fram Strait. The observatory is of type Multiple Arrays and is designed to monitor the exchange of Atlantic and Arctic waters. It is operated by the Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research. FRAM hosts 100 instruments and was first deployed on July 1, 1997. It provides partial near real-time data access.

2.2.2 Coastal

The Acqua Alta Oceanographic Tower is a research platform consisting of an instrument house supported by a steel pipe structure. It was deployed in January 1970[32], is located in Gulf of Venice, Mediterranean Sea, is part of the Long Term Ecological Research (LTER) network, and is operated by the Institute of Marine Sciences (ISMAR), National Research Council (CNR) of Italy. The platform allows for specific dedicated campaigns as well as long-term measurement of biological, chemical, physical oceanography parameters. Meteorological stations and oceanographic instruments enable monitoring atmospheric and hydrological parameters. The four instruments measure wind speed and direction, air temperature, humidity, currents, waves, water temperature, dissolved oxygen, turbidity, conductivity, and salinity. Three high resolution cameras allow for direct view of sea conditions and two underwater cameras are installed to "observe biological populations and to monitor potentially critical phenomena such as jellyfish swarms and mucilaginous macro aggregates." It can host two technicians and three scientists for several days, has energy supply and 10 Mbit/s data communication rate.[33]

The Coastal Observation System for Northern and Arctic Seas (COSYNA) "is an operational coastal monitoring, forecasting and information system for the North Sea composed by fixed platforms, FerryBoxes, gliders and HF-radar systems".[34] The observatory is operated by the German Marine Research Consortium. The construction phase

---

[32] http://www.ismar.cnr.it/infrastructures/piattaforma-acqua-alta

[33] http://www.jerico-ri.eu/infrastructure/acqua-alta-oceanographic-tower/

[34] http://www.jerico-ri.eu/infrastructure/cosyna-one-glider/

started in 2007 and the operational phase in 2012. The glider is equipped with 4 sensors measuring various parameters including conductivity, temperature, depth, fluorescence, turbidity. Data from the gliders is collected every 3-4 hours. FerryBoxes are attached to vessels and monitor various parameters along different routes in the southern North Sea. For each FerryBox, a total of 8 sensors measure parameters such as alkalinity, turbidity, pCO_2, pH, dissolved oxygen, chlorophyll-a fluorescence, temperature, conductivity, salinity. Data is collected after each cruise.

The Utö Atmospheric and Marine Research Station is an observing site of the Finnish Meteorological Institute located on Utö Island at the outer edge of the Archipelago Sea. The observatory monitors surface waves, temperature, currents, chlorophyll, salinity, dissolved oxygen, turbidity, sea ice. The observatory also monitors greenhouse gases as well as aerosol and trace gases.

Operated by ONC, the Victoria Experimental Network Under the Sea (VENUS) observatory network[35] is deployed in the coastal waters of southern British Columbia. The network "provides long-term oceanographic data on physical, chemical, biological, and sediment conditions."

2.2.3 Seafloor

The Expandable Seafloor Observatory OBSEA is an underwater cabled observatory operated by the Polytechnic University of Catalonia.[36] The observatory is located on the Catalan Coast near Barcelona at a depth of 20 m and is cabled over 4 km with the coast—including a 1 GBit/s fiber optic link and electric power. The underwater observatory is also connected to a surface buoy. The buoy measures oceanographic and environmental parameters, specifically wind speed, air temperature and atmospheric pressure as well as GPS position, orientation and pitch and roll movements. Furthermore, a video camera provides images of the buoy. The observatory hosts 8 sensors monitoring 15 parameters, including currents and waves, seabed vibration, temperature, and pressure. Data collection is in near real-time.

The Porcupine Abyssal Plain (PAP) observatory "is a sustained, multidisciplinary observatory in the North Atlantic coordinated by the National Oceanography Centre,

---

[35] http://www.oceannetworks.ca/installations/observatories/venus-salish-sea

[36] http://www.jerico-ri.eu/infrastructure/expandable-seafloor-observatory/

Southampton".[37] Located in the Northeast Atlantic, the aim of the observatory is to acquire time-series monitoring data "for analysing the effect of climate change on the open ocean and deep-sea ecosystems." It is a fixed point observatory consisting of an array of moorings covering the entire water column and seafloor. The observatory was first deployed on October 1, 2002, though some of the instrumentation has delivered data for two decades. At 4850 m depth a time-lapse camera records images of the seafloor every 8 hours.

NEMO-SN1 is a multidisciplinary deep-sea real-time multi-parameter observatory located in the Western Ionian See, offshore Catania, Sicily, at 2100 m depth. It is a cabled observatory first deployed on September 20, 2002, with real-time data transmission since 2005, and is operated by the Istituto Nazionale di Geofisica e Vulcanologia (INGV), Italy. Data and electric power are transmitted from the shore to the observatory via a 28 km long electro-optical cable. The observatory hosts 13 instruments and monitors a number of geophysical and environmental parameters, including temperature, conductivity, pressure, currents, seismics, acoustics, magnetics, and gravity.

2.2.4 Water Column

Argo floats carry out water column profile measurements while ascending from 2000 m depth to the surface, from where collected data are transmitted. Several types of floats are deployed. The Provor, developed by IFREMER, can be configured to perform measurements also during descent and while parking. The float measures temperature, conductivity, pressure, salinity. These parameters are measured by sampling water and pumping it through an instrument. The Provor also measures dissolved oxygen. Further biogeochemical parameters that can be monitored by Argo floats include fluorescence and turbidity. New technologies will enable the measurement of zooplankton using acoustic sensors.

Described earlier as a seafloor observatory, the Porcupine Abyssal Plain (PAP) observatory is equipped with a sub-surface sediment trap mooring designed to monitor particle flux and currents between 3000 and 4800 m depth. The particle of living and dead material is known as "marine snow" and falls from the upper ocean parts downwards.

---

[37] http://noc.ac.uk/pap

Described earlier as a fixed point observatory, the Southern Adriatic Interdisciplinary Laboratory for Ocean Research (E2-M3A) monitors a number of physical and biogeochemical processes in the water column.

These examples show that an observatory may be classified in multiple categories, and funded by various projects.

2.2.5 Open Ocean

Open Ocean observatories are also known as Deep Ocean. The Argo floats network is an example observatory monitoring parameters in the open ocean.

Operated by ONC, the North East Pacific Time-series Underwater Networked Experiments (NEPTUNE) observatory network[38] is located off the west coast of Vancouver Island, British Columbia. The network consists of six observatories, most of which are located in the open and deep ocean, at depths up to 2660 m. The observatories are connected by 840 km fibre optic cable. The NEPTUNE system hosts over 130 instruments. Data collection began in 2009 and is in real time.

Operated by OOI, the Cabled Continental Margin Array[39] is a network of 8 observatories installed at the base of the continental slope and performs water column and seafloor measurements. Data is collected in near real-time. The Oregon Slope Base Shallow Profiler Mooring is one of the 8 observatories of the Array, hosts 10 instruments, and provides “a wide variety of opportunities for observing coastal phenomena, including cross-shelf and along-shelf variability.”

The Alfred Wegener Institute (AWI) operates the HAUSGARTEN Long-Term Ecological Research (LTER) observatory[40] in the deep Arctic Ocean. The observatory is located in the eastern Fram Strait. It enables “the detection of expected changes in abiotic and biotic parameters in a transition zone between the northern North Atlantic and the central Arctic Ocean.” The observatory is in fact a network 21 permanent observatories (stations). The covered water depth range is 300-5500 m. The observatory was deployed in summer 1999.

---

[38] http://www.oceannetworks.ca/installations/observatories/neptune-ne-pacific

[39] http://oceanobservatories.org/array/cabled-continental-margin/

[40] https://www.awi.de/en/science/special-groups/deep-sea-research/observatories/lter-observatory-hausgarten.html

2.3 Data Repositories

As the observatories described here underscore, state of the art infrastructures tend toward near (quasi) real-time data collection. Acquired data are curated in data management systems. Data is accessed through data portals, a Web based interface that supports search and download of data. Some systems may support plotting (real-time) data.

In this section we briefly review some data repositories and their data portals through which data can be accessed. Some of the presented repositories are developed for marine monitoring data while others are more general purpose data repositories, e.g. for a discipline such as earth and environmental science.

SeaDataNet[41] is a “standardized system for managing the large and diverse data sets collected [in marine monitoring].” The system attempts to address the problem of hardly accessible and not standardized data collected by the large number of scientific data collecting laboratories. SeaDataNet is a virtual data management system that aims at networking existing national oceanographic data centers of 35 countries. Funded by the European Commission, the project is currently in its second phase and aims at “an operationally robust and state-of-the-art Pan-European infrastructure for providing up-to-date and high quality access to ocean and marine metadata, data and data products.” Realizing *semantic* interoperability among data management systems is, among others, a goal of the second phase.

The SeaDataNet portal presents metadata, data, and data product services. For metadata, SeaDataNet presents various registries compiled from national contributions, including registries for European marine organizations, environmental research projects, cruise report, and environmental datasets. The registries are maintained by organizations in selected European countries—namely the private company MARIS in The Netherlands, the British Oceanographic Data Centre, and the German Maritime and Hydrographic Agency. However, through SeaDataNet the “directories have been harmonised in use of syntax, semantics and tools.”

---

[41] http://www.seadatanet.org/

SeaDataNet also provides an integrated and harmonized overview and access to data resources managed in 90 national data centres from 35 European countries. This integration and harmonization is “achieved by developing, implementing and operating the Common Data Index [CDI] service.” The CDI provides an ISO-based index to individual data sets, and an interface to online data access. Data requests are forwarded from the portal to the relevant data centers and are managed by a specialized service. Users can use this service to request information about the status of their data requests. Access may be granted automatically, upon consideration by the data center, or denied, depending on user roles. Data are available in Ocean Data View format.

SeaDataNet also made attempts at providing data products for download. For instance, the ‘aggregated datasets’ are collections of temperature and salinity measurements by sea basins (such as the Baltic Sea or the Mediterranean Sea). SeaDataNet has produced and published two versions of such aggregated datasets, for January 2014 and March 2015.

The European Marine Observation and Data Network[42] (EMODnet) “consists of more than 100 organisations assembling marine data, products and metadata.” Its primary goal is to make “fragmented resources more available to public and private users relying on quality-assured, standardised and harmonised marine data which are interoperable and free of restrictions on use.”

EMODnet provides 8 thematic sub-portals, namely for bathymetry, geology, seabed habitats, chemistry, biology, physics, human activities, and coastal mapping. These portals provide a range of services and functionalities for data, data product, and metadata discovery and access. Some sub-portals, such as EMODnet Physics, provide machine-to-machine communication by offering OGC-compliant Web Feature Services (WFS) and the Web Map Services (WMS). Such services provide spatial information about features, such as the locations at which data are available for selected parameters. Sub-portals may also provide Web services for communication of the near real-time data streams. EMODnet provides a query tool that supports the aggregated access to data of different thematic sub-portals.

---

[42] http://www.emodnet.eu/

The collaboratories described in this section generally also develop and maintain data repositories and portals to enable access and download of data collected from observatories that constituent systems of the collaboratory. For instance, the FixO3 Web portal supports visualization of data for parameters monitored by observatories. Data can be downloaded and metadata can be retrieved. Another example is the Ocean Observatories Initiative, which includes *cyberinfrastructure* technology[43] dedicated to OOI data management. Technologies include “computing servers, data storage and backup, and front-facing [cyberinfrastructure] portal access point.”

---

[43] http://oceanobservatories.org/cyberinfrastructure-technology/

Chapter 3

Semantic Technologies

Over the past decade, Semantic Technologies and their application have matured from early specifications to production systems and applications. The adoption has been remarkable especially in academia and research-oriented organizations, while businesses have emerged with Semantic Technologies at their core or integrated in their products.

The popularity of Semantic Technologies can arguably be attributed to the activities within the Semantic Web Initiative of the World Wide Web Consortium (W3C). The idea underlying the Semantic Web, namely to evolve "objects from being principally human-readable documents to contain more machine-oriented semantic information" (Berners-Lee et al., 1994), had already been conceived in the early days of the Web. The ideas gained momentum with the Web's growth rate during the first decade of its existence. The steadily growing amount of data made it "increasingly difficult to locate, organize, and integrate the available information" and increasingly obvious that computers needed to do better at these tasks (Heflin and Hendler, 2001). As computers had not succeeded in processing natural language, researchers sought to make the Web more understandable to computers by giving data well-defined meaning (Berners-Lee et al., 2001).

This required the development of several new technologies, including standards and tools. Specifically, the standardization of the syntactic form of data achieved with the Extensible Markup Language (Bray et al., 1998, XML) was to be attained also for the *semantic* content of data (Decker et al., 2000).

The following sections briefly present the W3C recommendations that are of particular interest in this dissertation. Section 3.1 introduces the Resource Description Framework (RDF), which is the data model underlying the Semantic Web. Section 3.2 presents the *de facto* standard query language for RDF data, the SPARQL Protocol and RDF Query Language (SPARQL). Section 3.3 introduces the notion of ontology, as understood in information science. Ontologies, and languages for creating ontologies, are central to the Semantic Web as they allow for the definition of machine readable (i.e. formal) semantics of the concepts and relations relevant to Semantic Web applications.

Having introduced the notion of ontology, we present the two core Semantic Web ontology languages, namely RDF Schema (Section 3.4) and the Web Ontology Language (Section 3.5). These ontology languages have been utilized to construct numerous ontologies. Some are of interest to this dissertation. Hence, Section 3.6 presents an ontological framework consisting of key ontologies of interest to this work, namely OWL-Time (Section 3.6.1), GeoSPARQL (Section 3.6.2), ontologies for quantities (Section 3.6.3) and units (Section 3.6.4), the Semantic Sensor Network ontology (Section 3.6.5), the RDF Data Cube Vocabulary (Section 3.6.6), the Data Quality Vocabulary (Section 3.6.7), and the PROV ontology (Section 3.6.8). These ontologies form a framework for vocabulary relevant to environmental monitoring, as it addresses time, space, quantities, units, sensing, observation, datasets, data quality, and provenance. This framework is for environmental monitoring in general, and can be specialized for marine monitoring.

3.1 Resource Description Framework

The Resource Description Framework (Lassila and Swick, 1999; Klyne and Carroll, 2004; Cyganiak et al., 2014b) is a model of metadata, specifically a model of data about Web resources (Lassila and Swick, 1999). At its core, the model consists of resources, properties, and statements.

A *resource* is primarily a *Web* resource, such as a Web page or an image linked to a Web page. However, resources do not need to be accessible on the Web. Physical objects, such as sensing devices, or abstract concepts, such as time instant can be resources as well. Generally, any entity that can be named by a Uniform Resource Identifier (Berners-Lee et al., 2005, URI) is a resource and a member of the set of RDF resources.

```
@prefix ex: <http://example.org#> .

ex:aThermometer ex:observes ex:temperature .
ex:temperature ex:isPropertyOf ex:water .
```

Listing 3.1: Example RDF statements in the namespace `http://example.org#` with intelligible fragments suggesting that a thermometer observes the temperature of water.

A *property* is a specific "relation used to describe a resource" (Lassila and Swick, 1999) and is a member of the set `rdf:Property`, which is a subset of the set of RDF resources.

A *statement* is a triple consisting of a resource, a property, and the value for the property of the resource. Statements are members of the set `rdf:Statement`. The three elements of the triple are called, respectively, the subject, the predicate, and the object of the RDF statement. The object of a statement can be a resource or a literal. A literal is a value of primitive data type, in particular XML data type, and is a member of the set of RDF literals. Listing 3.1 includes two RDF statements. The six resources are in the namespace `http://example.org#` for which we define the prefix `ex:`. The subject of the first statement is named `ex:aThermometer`, which is equivalent to the URI

`http://example.org#aThermometer`

composed of the namespace and a fragment, i.e. `aThermometer`. The six fragments in the two statements suggest that a thermometer observes the temperature of water. Note that the statements are intelligible only to human readers. To computer systems, the two statements mean no more than, e.g., `(ex:a ex:o ex:t), (ex:t ex:i ex:w)`, mean to human readers.

The property `rdf:type` is an additional important feature of RDF. It is a member of the set `rdf:Property` and enables primitive typing in RDF. RDF requires that the subject and the object in a statement with `rdf:type` predicate are members of the set of RDF resources.

An RDF statement is represented graphically as two nodes connected by a directed arc. The two nodes are for the subject and the object of the statement, respectively. The arc is for the property, and is directed from the subject to the object. A set of RDF statements thus forms an RDF graph. Figure 3.1 displays the RDF graph corresponding to the RDF statements of Listing 3.1. To exchange RDF statements, Beckett (2004)

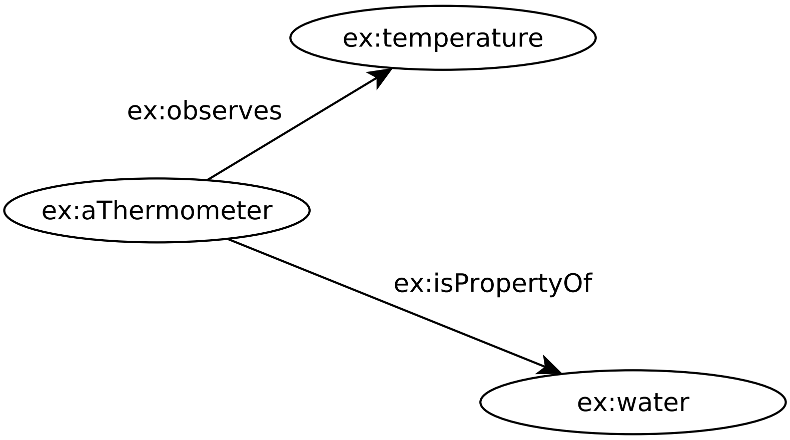


FIGURE 3.1: RDF graph corresponding to the RDF statements of Listing 3.1.

proposed an XML syntax for RDF. RDF/XML was designed for the exchange of RDF statements between computer systems. Various other RDF syntaxes have been proposed, such as Turtle (Prud'hommeaux and Carothers, 2014) and N-TRIPLES (Carothers and Seaborne, 2014), some of which are easier to parse, stream, or read by humans.

3.2 SPARQL Protocol and RDF Query Language

Having provided an overview of RDF, we now turn to the SPARQL Protocol and RDF Query Language (Prud'hommeaux and Seaborne, 2008; Harris and Seaborne, 2013). SPARQL is of interest here as query language for RDF.

The core SPARQL construct is arguably the *triple pattern*. A triple pattern follows the structure of RDF triples and allows for the subject, predicate, and object to be variable. A triple pattern *matches* RDF triples when variables can be substituted. For instance, given a set of RDF triples, a triple pattern with variables in the subject, predict, and object position matches all triples in the set. In SPARQL, a set of triple patterns is called a *basic graph pattern*.

SPARQL is somewhat reminiscent of the Structured Query Language (Chamberlin and Boyce, 1974, SQL). In fact, the declarative query language adopts several of the well-known SQL clauses, including `SELECT`, `WHERE`, and `ORDER BY`. A (group) graph pattern specifies the `WHERE` clause. Filters can be declared in order to restrict the solutions of a graph pattern according to a filter expression. Parts of the graph pattern may be

```
PREFIX ex: <http://example.org#> .

SELECT ?sensor ?property
WHERE {
  ?sensor ex:observes ?property .
  ?property ex:isPropertyOf ex:water .
}
```

Listing 3.2: Example SPARQL query for sensors that observe properties of water.

optional. Listing 3.2 is an example SPARQL SELECT query with a basic graph pattern consisting of two triple patterns. The query matches sensors that observe properties of water. For the RDF statements in Listing 3.1, the query returns the bindings `?sensor = ex:aThermometer` and `?property = ex:temperature`. Note that it is more accurate to speak of resources that `ex:observes` resources that are `ex:isPropertyOf` the resource named `ex:water`. This is because the chosen variable names `?sensor` and `?property` do not carry meaning for computer systems.

SPARQL supports query forms other than `SELECT`. Of particular interest is the `CONSTRUCT` query form. The `SELECT` query form returns variables and their bindings. In contrast, the `CONSTRUCT` query form returns an RDF graph specified by a graph template defined by the query. The `CONSTRUCT` query form is often useful in applications because the returned RDF graph can be further processed, e.g. by a subsequent query.

3.3 Ontology

A human reader may interpret the statements in Listing 3.1 to mean that a thermometer observes the temperature property of water. For computer systems, however, the URIs are little more than strings. That the URI `ex:aThermometer` refers to a particular *thermometer* is impossible for computer systems to conclude—and for human readers the fragment could be misleading.

Therefore, we need a mechanism to describe intended semantics explicitly. The descriptions should be both human and machine readable. The mechanism of choice here is the development of ontologies using formal ontology languages.

Ontology has recently received considerable attention in various computational fields of study (Gruber, 1993; Guarino and Giaretta, 1995; Studer et al., 1998). Gruber (1993) defines ontology as:

Definition 3.1 (Ontology)**.** Ontology is an explicit specification of a conceptualization.

Gruber's definition has been prevalent in the literature for over two decades. Unfortunately, it is not intuitive. What is a conceptualization and when is a specification explicit? Guarino et al. (2009) provides a succinct analysis of the notions of conceptualization and explicit specification. We briefly summarize Guarino et al.'s analysis of Gruber's definition, and then present some alternative definitions.

3.3.1 Conceptualization

Guarino et al. propose the intensional relational structure

$$(\mathcal{D}, \mathcal{W}, \mathcal{R})$$

as mathematical representation of a conceptualization. The universe of discourse $\mathcal{D}$ is a "set of objects about which knowledge is being expressed" (Genesereth and Nilsson, 1987). An object is "anything about which we want to say something" (Genesereth and Nilsson, 1987). $\mathcal{W}$ is a set of possible worlds and $\mathcal{R}$ is a set of intensional relations on $< \mathcal{D}, \mathcal{W} >$. An intensional relation of arity n on $< \mathcal{D}, \mathcal{W} >$ is a total function from the set $\mathcal{W}$ into the set of all n-ary extensional relations on $\mathcal{D}$. The intensional relational structure allows for different state of affairs (worlds) to be described by a single conceptualization.

Example 3.1 (Intensional relational structure)**.** *For the sensing domain, we demonstrate an intensional relational structure and how it supports the description of different state of affairs. Our universe of discourse* $\mathcal{D}$ *contains sensors, observed properties, and features, each identified by a code. The set* $\mathscr{R}$ *contains the unary relations Sensor, Property, and FeatureOfInterest, as well as the binary relations observes and isPropertyOf. The intensional relational structure* $(\mathcal{D}, \mathcal{W}, \mathcal{R})$ *is:*

- $\mathcal{D} = \{s01,\ s02,\ \ldots,\ p01,\ \ldots,\ f01,\ \ldots\}$
- $\mathcal{W} = \{w_1,\ w_2,\ \ldots\}$
- $\mathcal{R} = \{Sensor^1,\ Property^1,\ FeatureOfInterest^1,\ observes^2,\ isPropertyOf^2\}$

The intensional relations may thus map to different extensions in different worlds, as shown for the two binary relations:

- *for all worlds* w *in* $\mathcal{W}$: $Sensor^1(w) \cup Property^1(w) \cup FeatureOfInterest^1(w) = \mathcal{D}$
- *for all worlds* w *in* $\mathcal{W}$: $Sensor^1(w) = \{s01, s02, \ldots\}$
- *for all worlds* w *in* $\mathcal{W}$: $Property^1(w) = \{p01, \ldots\}$
- *for all worlds* w *in* $\mathcal{W}$: $FeatureOfInterest^1(w) = \{f01, \ldots\}$
- $observes^2(w_1) = \{\ldots, (s01, p01), \ldots\}$
- $observes^2(w_2) = \{\ldots, (s02, p01), \ldots\}$
- $observes^2(w_3) = \ldots$
- $isPropertyOf^2(w_1) = \{\ldots, (p01, f01), \ldots\}$
- $isPropertyOf^2(w_2) = \ldots$

3.3.2 Explicit Specification

The *explicit specification* of a conceptualization rests on a language, one that enables reference to the elements of a conceptualization—in other words, one that enables us to talk about a conceptualization.

Of particular interest are logical languages, with vocabulary consisting of a set of constant and predicate symbols. The symbols in the vocabulary of a language obtain meaning through interpretations. Specification occurs by means of axioms that constrain the possible interpretations for the symbols.

An ontology is a set of axioms. Guarino et al. (2009) underscore that an ontology is, strictly speaking, an *approximate* specification of a conceptualization, in other words a partial account of a conceptualization. This is because the degree of specification depends on various factors, e.g. the purpose of the specification (Guarino and Giaretta, 1995).

3.3.3 Alternative Definitions

Alternative definitions have been proposed in the literature. Building on Gruber's definition, Borst (1997) defines ontology as a "formal specification of a shared conceptualization." In this definition, the specification must be formal, i.e. machine readable (Guarino et al., 2009). Formal languages such as logical languages meet this requirement, while natural language does not. Furthermore, the conceptualization must be shared, i.e. it must reflect a consensus among ontology stakeholders. Indeed, specifications of a conceptualization that lack consensus are arguably hard to reuse and are thus considered useless (Borst, 1997; Guarino et al., 2009). Studer et al. (1998) merge these definitions and define ontology as a "formal, explicit specification of a shared conceptualization."

Other authors have proposed alternative definitions that avoid the notion of conceptualization. According to Neches et al. (1991) "an ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary." According to Swartout et al. (1996), an ontology is "a hierarchically structured set of terms for describing a domain that can be used as a skeletal foundation for a knowledge base." Thus, Swartout et al. explicate, "an ontology provides a skeletal structure for a knowledge base." According to Hendler (2001), an ontology is "a set of knowledge terms, including the vocabulary, the semantic interconnections, and some simple rules of inference and logic for some particular topic." Hendler notes that this definition reflects (or reflected, at the time of his writing) how the term ontology is used within the Semantic Web community. Berners-Lee et al. (2001), and co-author Hendler, note that to Artificial Intelligence and Web researchers "an ontology is a document or file that formally defines the relations among terms."

Some elements of these alternative definitions—such as term, vocabulary, interconnection, and domain—are reminiscent of Gruber's definition, while others are new. Swartout et al. introduce the notion of Knowledge Base as an entity distinct from ontology. Swartout and Tate (1999) clarify that a Knowledge Base uses the set of terms provided by an ontology "to represent what is true about some real or hypothetical world." By introducing the notion of rule, Hendler's definition acquires reasoning as additional characteristic.

3.4 RDF Schema

In RDF, the property of a statement represents a relationship between two resources. RDF does not provide a mechanism to *describe* such relationships, for instance describe the particular groups (classes) of resources a property relates. This is addressed by RDF Schema (Brickley and Guha, 2004, 2014). RDF Schema (RDFS) is a data-modelling vocabulary for RDF data. The RDFS vocabulary consists of defined sets of resources and is a language for constructing basic ontologies.

Example 3.2 (Relationship description). *Consider the property `ex:observes` in Listing 3.1. Given the domain of sensing, the property intuitively relates things that observe, i.e. sensors, with things that are observed, i.e. properties. However, RDF does not provide a mechanism to describe the groups of things that the property `ex:observes` relates, i.e. sensors and properties. Such semantics can be expressed using RDFS. The RDFS vocabulary supports defining the semantics of the group of things that observe and the group of things that are observed, as well as specifying that the property `ex:observes` relates these two groups.*

Any entity described in RDF is a resource, and instance of `rdfs:Resource`. Thus, `rdfs:Resource` includes "everything." Resources may be divided into groups, i.e. classes. All classes are thus subclasses of `rdfs:Resource`. Among classes, `rdfs:Class` is the class of *resources* that are RDF classes. Clearly, `rdfs:Class` is a subclass of `rdfs:Resource`. A class, including `rdfs:Resource`, is an instance of `rdfs:Class`. Being the class of all RDF properties, `rdf:Property` is an instance of `rdfs:Class`. RDFS defines `rdfs:Literal`, the class of all RDF literals. Being a class, `rdfs:Literal` is an instance of `rdfs:Class` and a subclass of `rdfs:Resource`.

Example 3.3 (Class). *The property `ex:observes` intuitively relates things that observe, i.e. sensors, with things that are observed, i.e. properties. With RDFS we can define a class for sensors by stating that `ex:Sensor` is a resource instance of `rdfs:Class`. Concrete sensors related by `ex:observes` with concrete properties are thus instances of `ex:Sensor`. Concrete properties may be instances of the class `ex:Property`.*

In addition to classes, RDFS defines a particular set of properties, instances of the class `rdf:Property`. Specifically, the property `rdfs:subClassOf` enables the construction of class hierarchies. The statement

<C, `rdfs:subClassOf`, D>

states that all instances of `C` are instances of `D`, and that the classes `C`, `D` are instances of `rdfs:Class`.

Example 3.4 (Subclass). *Being resources capable of observing properties, some devices are sensors. Given that, for instance, humans are well capable of observing properties, sensors are, however, not necessarily devices. We can group the resources that are devices capable of observing properties as class `ex:SensingDevice`. Using the RDFS property `rdfs:subClassOf`, we can specify that:*

<`ex:SensingDevice`, `rdfs:subClassOf`, `ex:Sensor`>

<`ex:SensingDevice`, `rdfs:subClassOf`, `ex:Device`>

On a similar line, the property `rdfs:subPropertyOf` enables the construction of property hierarchies. Given the statement

<P, `rdfs:subPropertyOf`, Q>

pairs of resources related by `P` are also related by `Q`. The statement also implies that `P`, `Q` are instances of `rdf:Property`.

Two further properties defined by RDFS are of particular interest, namely `rdfs:domain` and `rdfs:range`. The statements

<P, `rdfs:domain`, C> <P, `rdfs:range`, D>

state that for statements <`r`, `P`, `s`> the resources `r` and `s` are instances of the classes `C` and `D`, respectively, that `P` is an instance of `rdf:Property`, and that `C`, `D` are instances of `rdfs:Class`.

Example 3.5 (Domain and range). *Earlier we stated that the property `ex:observes` intuitively relates the class of sensors with the class of properties. We can specify such semantics by using the RDFS properties `rdfs:domain` and `rdfs:range` as follows:*

<`ex:observes`, `rdfs:domain`, `ex:Sensor`>

<`ex:observes`, `rdfs:range`, `ex:Property`>

For the statement <`ex:aThermometer`, `ex:observes`, `ex:temperature`> we can thus conclude that the resources `ex:aThermometer` and `ex:temperature` are instances of the classes `ex:Sensor` and `ex:Property`, respectively. RDFS thus provides a mechanism to describe relationships.

3.5 Web Ontology Language

RDFS supports the construction of basic ontologies. The construction of ontologies with richer semantics is supported by the Web Ontology Language (Bechhofer et al., 2004; Motik et al., 2012). A Web Ontology Language (OWL) ontology consists of a set of axioms and, often, a set of assertions (i.e. "facts about individuals"). The set of axioms consists of class axioms and property axioms. The set of assertions consists of concept and role assertions, i.e. class membership and property values of individuals. The following paragraphs describe the core features of OWL, in particular how the language supports the definition of axioms and assertions.

Example 3.6 (Axiom)**.** *With OWL we can define sensors as physical objects that observe properties by means of the following class axiom:*

```
Class: ex:Sensor
  SubClassOf: ex:PhysicalObject that ex:observes only ex:Property
```

OWL supports the description of classes by means of six types of so-called class descriptions. An `owl:Class`, which is defined as a subclass of `rdfs:Class`, can be described through (1) a class name (as in RDFS); (2) an exhaustive enumeration of individuals, instances of the described class; (3) a property restriction; the (4) intersection or (5) union of two or more class descriptions; or (6) the complement of a class description. A property restriction describes the class of all individuals that satisfy the restriction. There exist two types of property restrictions: value constraints and cardinality constraints. A value constraint restricts the range of the property when applied to the particular class description (which is thus different from `rdfs:range`). This type of property restriction includes constraints analogous to universal and existential quantifiers of Predicate logic. The axiom for the class `ex:Sensor` in Example 3.6 includes a (universal) value constraint that restricts the range of the property `ex:observers` to instances of the class `ex:Property`. A cardinality constraint restricts the number of values a property can take in the context of the particular class description. An instance of a class may have an arbitrary number of values for a particular property. Cardinality constraints can make a property required, allow only a specific number of values for a property, or specify that a property must not occur.

In addition to `rdfs:subClassOf`, a property inherited from RDFS, OWL includes two further constructs for the definition of class axioms, i.e. `owl:equivalentClass` and `owl:disjointWith`. Subclass axioms represent necessary conditions for establishing class membership of an individual. In contrast, equivalent class axioms represent necessary *and* sufficient conditions.

Example 3.7 (Necessary and sufficient conditions)**.** *In Example 3.6, sensors are defined as physical objects that observe properties. It is necessary for a sensor to be a physical object and observe properties. Not being equivalent classes, it is, however, not sufficient for physical objects to observe properties to be sensors.*

OWL distinguishes between object and data type properties. An object property is an instance of the class `owl:ObjectProperty` and relates two individuals. A data type property is an instance of the class `owl:DatatypeProperty` and relates an individual and a literal. Both are subclasses of `rdf:Property`. OWL defines several constructs for property axioms in addition to those inherited from RDFS, such `owl:equivalentProperty` and `owl:TransitiveProperty`.

Example 3.8 (Object property)**.** *The property `ex:observes` is an OWL object property.*

Facts about individuals are defined in OWL with axioms about individuals (i.e. assertions). Of particular interest are axioms that specify the class membership of an individual (concept assertions) and axioms that specify the property values of individuals (role assertions). OWL also supports stating that two individuals are same or are different.

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix time: <http://www.w3.org/2006/time#> .

<> rdf:type time:Interval ;
  time:hasBeginning [
    rdf:type time:Instant ;
    time:inXSDDateTime "2016-07-20T00:00:00Z"^^xsd:dateTime
  ] ;
  time:hasEnd [
    rdf:type time:Instant ;
    time:inXSDDateTime "2016-07-21T00:00:00Z"^^xsd:dateTime
  ] .
```

Listing 3.3: Example OWL-Time interval with beginning and end instants represented in XSD `dateTime` strings.

Example 3.9 (Assertion)**.** *Following the class axiom for sensors in Example 3.6 we can state the following assertions about the individual thermometer `ex:aThermometer` being a member of the class `ex:Sensor` with value `ex:temperature` for property `ex:observes`.*

```
Individual: ex:aThermometer
  Types: ex:Sensor
  Facts: ex:observes ex:temperature
```

3.6 Ontological Framework

We present an ontological framework with design patterns for modelling time and space; quantities and units; sensor descriptions, including metadata for, e.g., sensor capabilities; observations and their relations to sensor, property, feature, and value; datasets with their structure and data elements; data quality; and provenance. The selected ontologies provide relevant generic vocabulary for environmental monitoring, metadata and data. For each ontology, we present the key classes and properties. The framework can be specialized with vocabulary defined by ontologies relevant to marine monitoring.

It is important to note that the ontologies mentioned here are a selection. There are many alternative ontologies that provide relevant generic vocabulary for environmental monitoring. However, for brevity and focus we survey only those selected for this framework. Alternatives and comparisons are discussed in Chapter 5.

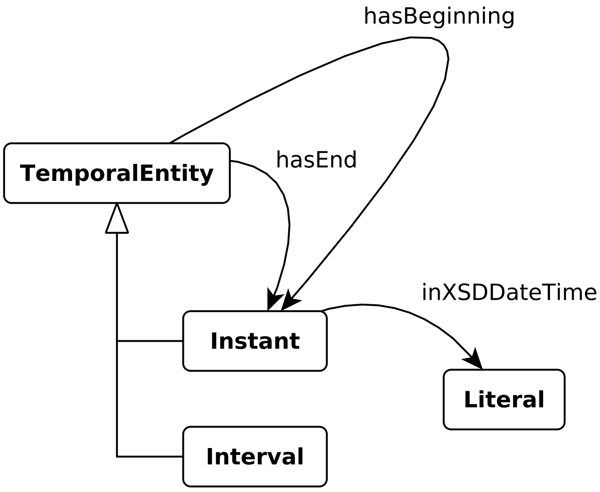


FIGURE 3.2: Relations between OWL-Time temporal entity with beginning and end instant with XSD textual representation.

3.6.1 Time

OWL-Time (Hobbs and Pan, 2006) is an ontology of temporal concepts, for describing temporal content and properties. It supports the description of topological relations among instants and intervals as well as information about durations and timestamps.

The ontology defines the class `time:TemporalEntity` and the subclasses `time:Instant` and `time:Interval`. It also defines the two object properties `time:hasBeginning` and `time:hasEnd` used to relate a temporal entity with an instant. Finally, it defines the data type property `time:inXSDDateTime` used to relate an instant with a literal of type `xsd:dateTime`. Beyond these basic classes and properties, OWL-Time allows for the explicit representation of temporal descriptions (e.g. durations) and topological relations (e.g. before). Figure 3.2 provides a graphical overview of the relations between OWL-Time temporal entity and the XSD textual representation of instants. Listing 3.3 demonstrates an example OWL-Time interval with beginning and end instants.

The Spatial Data on the Web Working Group,[1] a joint activity of W3C and OGC, is currently revisiting the earlier draft by Hobbs and Pan and has proposed a first working draft (Cox and Little, 2016). Compared to the earlier draft, the new working draft aims at supporting calendars, or Temporal Reference Systems, other than the Gregorian calendar, such as Unix date, Carbon Date, Geological Date.

---

[1] https://www.w3.org/2015/spatial/

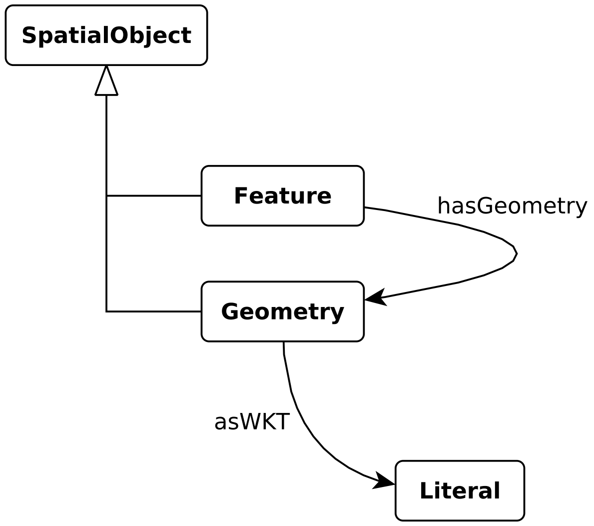


FIGURE 3.3: Relations between GeoSPARQL feature and geometry with WKT textual representation.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix geo: <http://www.opengis.net/ont/geosparql#> .
@prefix sf: <http://www.opengis.net/ont/sf#> .

<> rdf:type geo:Feature ;
  geo:hasGeometry [
    rdf:type sf:Point ;
    geo:asWKT "POINT (0 0)"^^geo:wktLiteral
  ] .
```

LISTING 3.4: Example GeoSPARQL feature with geometry represented as WKT.

3.6.2 Space

GeoSPARQL (Perry and Herring, 2012) is a vocabulary for representing spatial information. The ontology defines the class `geo:SpatialObject`, as well as its subclasses `geo:Feature` and `geo:Geometry`. It defines the object property `geo:hasGeometry` used to relate a feature with a geometry. Finally, it defines the data type property `geo:asWKT` used to relate a geometry with a literal of type `geo:wktLiteral` (a GeoSPARQL data type) to allow for text representation of geometries. Beyond these most relevant classes and properties, GeoSPARQL supports the explicit representation of topological relations, in particular also those of the Region Connection Calculus (Randell et al., 1992). Figure 3.3 provides a graphical overview of the relations between GeoSPARQL feature and geometry with WKT textual representation. Listing 3.4 demonstrates an example GeoSPARQL feature with geometry represented as Well-Known Text (WKT).

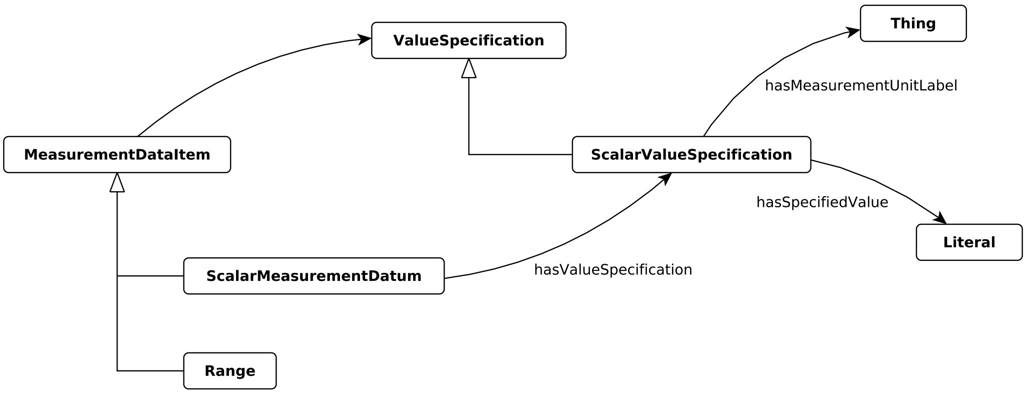


FIGURE 3.4: Relations between a `MeasurementDataItem` and a `ValueSpecification`.

3.6.3 Quantities

There exist several ontologies of potential interest for the semantic representation of quantities. For instance, the Statistical Methods Ontology[2] (STATO) is a general-purpose statistics ontology with an extensive vocabulary for statistical terms. The ontology provides formal definitions of the most common statistical tests; supports communicating and reporting scientific results; supports annotating statistical methods used in scientific investigations.

However, in this dissertation we adopt schema.org terms to represent quantitative values. While schema.org is not an ontology in the formal sense, the vocabulary addresses a limitation we found in our review of other ontologies, namely the support for quantitative value *ranges*.

STATO utilizes the term `MeasurementDataItem` of the Information Artifact Ontology[3] (IAO). A `MeasurementDataItem` (or `MeasurementDatum`) is, specifically, either a `ScalarMeasurementDatum` or a `Range`. A data item relates to a `ValueSpecification`. Of specific interest here is the `ScalarValueSpecification`, which relates a value with a unit. Figure 3.4 provides a graphical overview of the relations between the IAO measurement data item and the value specification. Unfortunately, it is not immediately clear how to represent the value specification of a `Range`.

Schema.org proposes a more straightforward pattern that is also explicit regarding the representation of quantitative value ranges. The schema.org description for the term

---

[2] http://stato-ontology.org/

[3] https://github.com/information-artifact-ontology/IAO/

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix unit: <http://qudt.org/vocab/unit#> .
@prefix schema: <http://schema.org/> .

<> rdf:type schema:QuantitativeValue ;
   schema:minValue "-5.0"^^xsd:float ;
   schema:maxValue "45.0"^^xsd:float ;
   schema:unitCode unit:DegreeCelsius .
```

Listing 3.5: Example schema.org quantitative value range in degree Celsius.

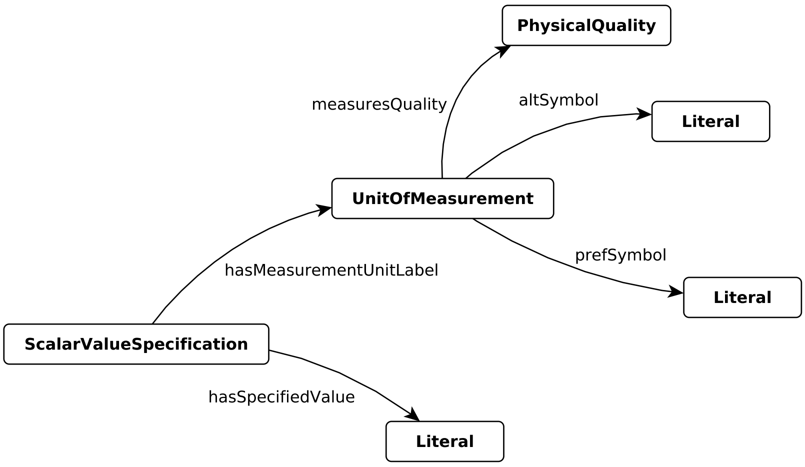


Figure 3.5: Relations between a scalar value specification and related MUO unit of measurement as a specialization of `owl:Thing`.

`QuantitativeValue`[4] lists a number of properties, whereby the most relevant here are `value`, `minValue`, `maxValue`, and `unit`. Listing 3.5 provides an example schema.org quantitative value range in degree Celsius.

3.6.4 Units

With the property `unitCode`, the schema.org `QuantitativeValue` already provides for relating values to units. For the purpose here, we adopt the Quantities, Units, Dimensions and Data Types Ontologies (QUDT) (Hodgson et al., 2014), specifically the URIs provided by the units ontology. These identifiers are the values of the schema.org `unitCode` property of quantitative values.

Other, including more formal, options exist. For instance, the representation of quantities suggested earlier in Figure 3.4 suggests that the unit needs to be specialized to a class of an ontology specifying units. Alternatively to QUDT, one could also

---

[4] http://schema.org/QuantitativeValue

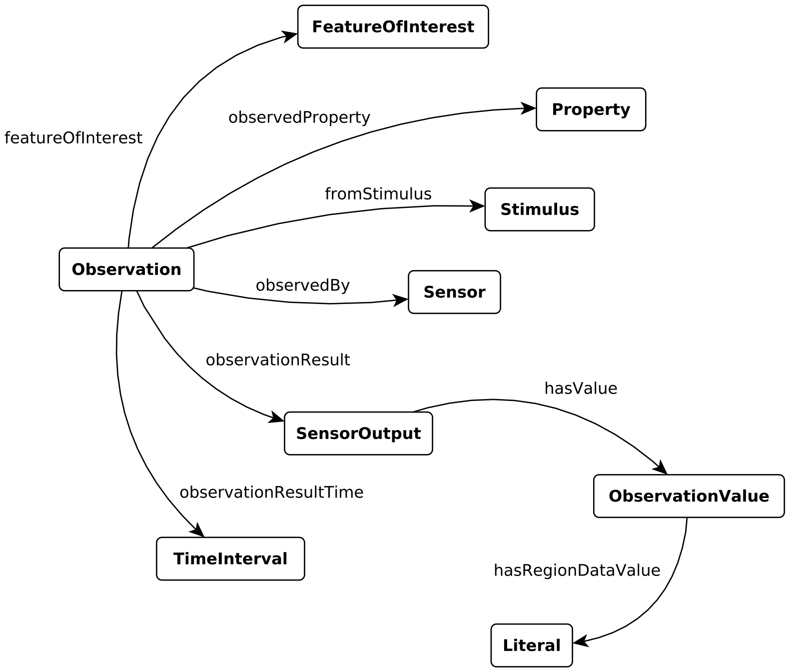


FIGURE 3.6: Relations between SSN observation and the sensor that made the observation, the observed property of the feature, the sensor output and observation value, and the time at which the observation was made.

choose the Units of Measurements (MUO) ontology.[5] Of particular interest is the `muo:UnitOfMeasurement` class which relates to preferred and alternative symbols and the measured physical quality. Interesting is the adoption of the Unified Code for Units of Measure (UCUM) in MUO, as MUO provides a set of URIs for UCUM instances. Figure 3.5 provides a graphical overview for how the measurement unit in a scalar value specification may be specialized to the MUO unit of measurement and the relations to symbols and measured physical quality.

3.6.5 Sensing

The Semantic Sensor Network (SSN) ontology (Compton et al., 2012) is designed to "describe the capabilities and properties of sensors, the act of sensing and the resulting observations." The SSN ontology aims at providing semantic interoperability of sensor data, on top of syntactic interoperability addressed in particular also by Open Geospatial Consortium (OGC) standards such as SensorML (Botts and Robin, 2007) and Observations and Measurements (Cox, 2011, O&M).

---

[5] http://idi.fundacionctic.org/muo/muo-vocab.html

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dul: <http://www.loa-cnr.it/ontologies/DUL.owl#> .
@prefix ssn: <http://purl.oclc.org/NET/ssnx/ssn#> .
@prefix ex: <http://example.org#> .

ex:water rdf:type ssn:FeatureOfInterest .

ex:temperature rdf:type ssn:Property .

ex:ethanolVolumeExpansion rdf:type ssn:Stimulus .

ex:aThermometer rdf:type ssn:Sensor .

<> rdf:type ssn:Observation ;
  ssn:featureOfInterest ex:water ;
  ssn:observedProperty ex:temperature ;
  ssn:fromStimulus ex:ethanolVolumeExpansion ;
  ssn:observedBy ex:aThermometer ;
  ssn:observationResult [
    rdf:type ssn:SensorOutput ;
    ssn:hasValue [
      rdf:type ssn:ObservationValue ;
      dul:hasRegionDataValue "19.4"^^xsd:double
    ]
  ] ;
  ssn:observationResultTime [
    rdf:type time:Instant ;
    time:inXSDDateTime "2016-07-21T00:00:00Z"^^xsd:dateTime
  ] .
```

Listing 3.6: Example SSN observation with feature of interest, property, stimulus, sensor, sensor output, and time inteval.

Though descriptions about the capabilities and properties of sensors are useful in applications, of most interest here are the observations resulting in the act of sensing, i.e. the observation perspective of the SSN ontology. To model observations, the SSN ontology defines the class `ssn:Observation`. Closely aligned with OGC standards and modelling of observations, an SSN observation is *for* a particular property *of* a feature, is *from* a stimulus, and is *observed* by a sensor that implements some sensing method. Sensor is understood broadly to include physical devices as well as other entities that can implement a sensing method to observe a property, such as computational methods or laboratory set-ups. Naturally, in addition to descriptions for what was sensed, what made the observation and how it was made, SSN observations also describe the sensor output, which is often a numerical observation value. Finally, SSN observations can describe other metadata, in particular spatio-temporal data for where and when the observation was made. Ontological modelling of time and space are, however, not part of the SSN ontology. Figure 3.6 provides a graphical overview of the main relations between SSN observation and sensor, property, feature, stimulus, observation value, and time. Listing 3.6 demonstrates an example SSN observation observed by a thermometer

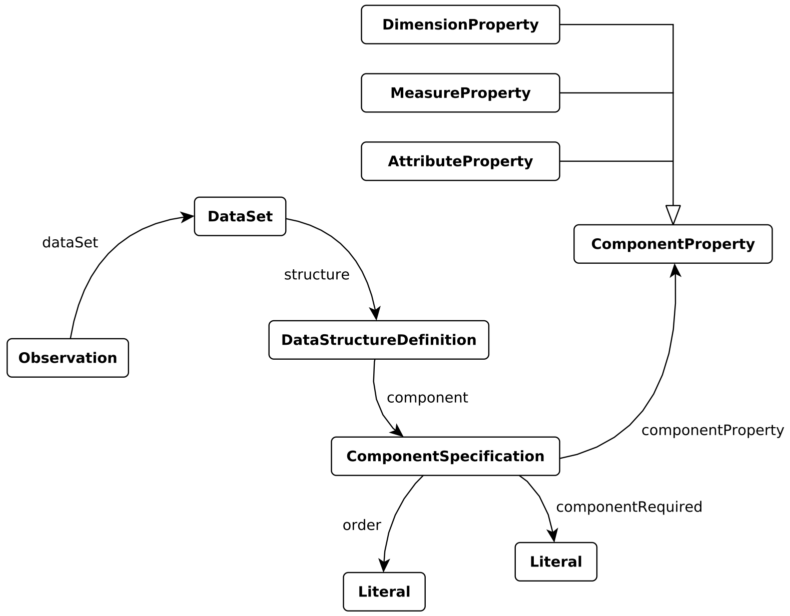


Figure 3.7: Relation between QB observation and the dataset with its data structure definition consisting of a set of component specifications. Component properties are RDF properties available to observations to relate property values.

for water temperature of 19.4 at midnight of July 21, 2016.

In this framework, SSN observation values are represented as Schema.org quantitative values. This enables the representation of scalar or range values as well as scalar value units in SSN observations.

3.6.6 Dataset

The RDF Data Cube (QB) Vocabulary (Cyganiak et al., 2014a) is designed to represent multi-dimensional (multi-variate) data in RDF. Fundamental to multi-dimensional data "is a set of observed values organized along a group of dimensions, together with associated metadata" (Cyganiak et al., 2014a). Observed values are represented as `qb:Observation`. A QB observation relates to a `qb:DataSet`, which is thus a collection of observations. Datasets are generally structured. Accordingly, QB supports the definition of structures as `qb:DataStructureDefinition`. One or more datasets may relate to the same data structure definition. A structure is described by a set of `qb:ComponentSpecification`. A component specification determines the `qb:ComponentProperty` as well as other metadata about the component, such as whether or not it is required and its order within the structure. QB supports three types of component properties,

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix qb: <http://purl.org/linked-data/cube#> .
@prefix sdmx-dimension: <http://purl.org/linked-data/sdmx/2009/dimension#> .
@prefix ex: <http://example.org#> .

sdmx-dimension:timePeriod rdf:type qb:DimensionProperty .
ex:waterTemperature rdf:type qb:MeasureProperty .

ex:d1 rdf:type qb:DataSet ;
  qb:structure [
    rdf:type qb:DataStructureDefinition ;
    qb:component [
      rdf:type qb:ComponentSpecification ;
      qb:componentProperty sdmx-dimension:timePeriod ;
      qb:componentRequired "true"^^xsd:boolean
    ] ;
    qb:component [
      rdf:type qb:ComponentSpecification ;
      qb:componentProperty ex:waterTemperature ;
      qb:componentRequired "true"^^xsd:boolean
    ] ;
  ] .

<> rdf:type qb:Observation ;
  qb:dataSet ex:d1 ;
  sdmx-dimension:timePeriod [
         rdf:type time:Instant ;
         time:inXSDDateTime "2016-07-21T00:00:00Z"^^xsd:dateTime
  ] ;
  ex:waterTemperature "19.4"^^xsd:double .
```

Listing 3.7: Example QB observation with components according to the data structure definition of the related dataset.

namely dimension, measure, and attribute properties. Component properties are RDF properties and are used to relate observations with values. Figure 3.7 provides a graphical overview of the relation between QB observation and dataset with data structure definition. Listing 3.7 demonstrates an example QB observation of dataset `ex:d1` with two components, one for time and one for water temperature, and their respective values.

As an example, consider a typical comma-separated values file consisting of n labels on the first line and m lines with n numbers starting on the second line and ending on line $m+1$ of the file. The first line of the file can be translated into a QB data structure definition. Each of the n labels is translated to a component specification. The label itself maps to a component property while the position of the label in the list determines the value of the order property in the component specification. The m lines of the file form a $m \times n$ multi-dimensional dataset. This dataset relates to the described data structure definition. Each of the $2 \ldots m+1$ lines in the file can be translated into a QB observation. Each line consists of n numbers. The QB observation relates thus to the dataset and to the n numbers via the component properties as defined by the data structure definition.

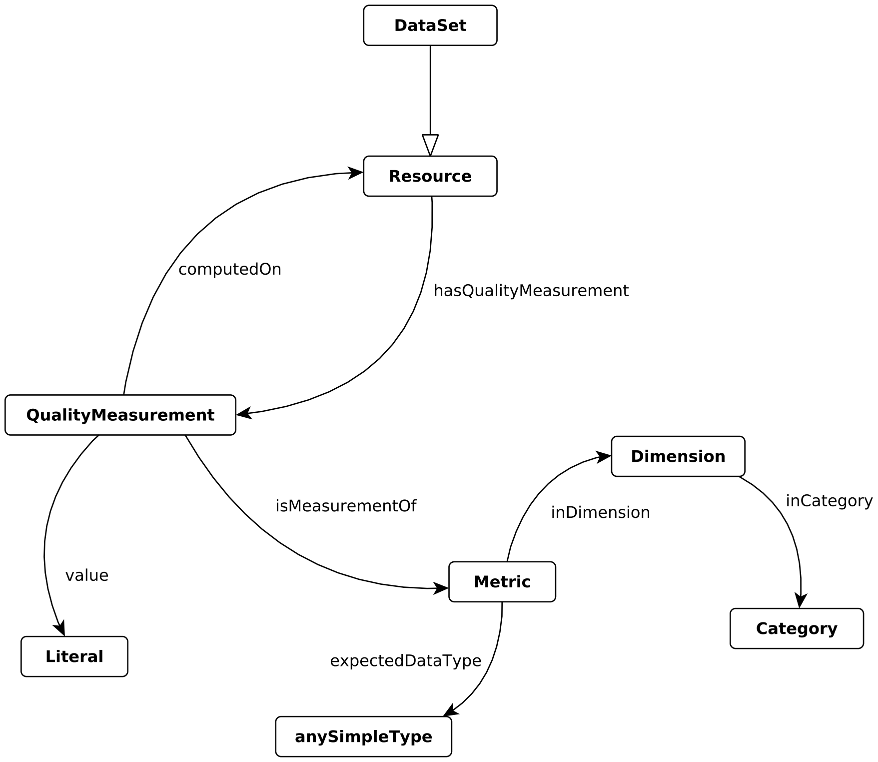


Figure 3.8: Relations between quality measurement and the resource it is computed on. The QB dataset is a resource type of interest here. A quality measurement specifies computed value for the quality metric.

In this framework, values of components of QB observations are represented as Schema.org quantitative values. This enables the representation of scalar or range values as well as scalar value units in QB observations.

3.6.7 Quality

The Data Quality Vocabulary (DQV) "provides a framework in which the quality of a dataset can be described" (Albertoni and Isaac, 2016). The vocabulary does not provide a formal definition of quality. Rather, it supports users in deciding whether or not a dataset is fit for a purpose.

Central to the DQV is the concept `dqv:QualityMeasurement`. A quality measurement represents a quantitative or qualitative metric value computed on a resource. A resource can, in principle, be any RDF resource. Practically, it is a dataset, a graph, a set of triples. In DQV, a resource is generally a dataset. In this framework, quality

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix qb: <http://purl.org/linked-data/cube#> .
@prefix iso: <http://iso25000.com#> .
@prefix dqv: <http://www.w3.org/ns/dqv#> .
@prefix ex: <http://example.org#> .

iso:inherentDataQuality rdf:type dqv:Category .

iso:completeness rdf:type dqv:Dimension ;
  dqv:inCategory iso:inherentDataQuality .

ex:missingValuesFraction rdf:type dqv:Metric ;
  dqv:expectedDataType xsd:double ;
  dqv:inDimension iso:completeness .

ex:d1 rdf:type qb:DataSet ;
  dqv:hasQualityMeasurement [
    rdf:type dqv:QualityMeasurement ;
    dqv:computedOn ex:d1 ;
    dqv:isMeasurementOf ex:missingValuesFraction ;
    dqv:value "0.045"^^xsd:double ;
  ] .
```

Listing 3.8: Example DQV quality measurement computed on dataset `ex:d1` with 4.5% missing values.

measurements are computed in particular on `ssn:Observation`, `qb:Observation`, or `qb:DataSet`.

A quality measurement represents a metric value as a quality metric `dqv:Metric` and a qualitative or quantitative value of some primitive datatype. A quality metric "gives a procedure for measuring a data quality dimension, which is abstract, by observing a concrete quality indicator." A dimension can have multiple metrics. A quality dimension `dqv:Dimension` is a "quality-related characteristic of a dataset relevant to the consumer." Finally, a category `dqv:Category` "represents a group of quality dimensions."

Figure 3.8 provides a graphical overview of the main relations between quality measurement, the resource for which the quality measurement is computed, the metric value the measurement represents. Listing 3.8 demonstrates an example DQV quality measurement computed on QB dataset `ex:d1`. The quality measurement expresses that the fraction of missing values to all values is 0.045, or 4.5%. The fraction of missing values is a metric in `iso:completeness`, a dimension in the category of `iso:inherentDataQuality`.

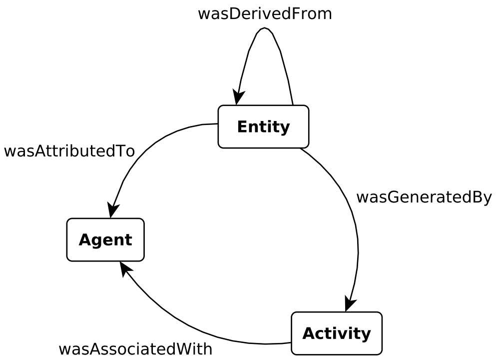


FIGURE 3.9: Relations between PROV entity, activity, and agent. The PROV ontology supports the representation of information about the provenance of SSN observations and QB datasets and observations in environmental monitoring systems, as well as information about the involved (software) agents and (algorithmic) activities.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix ex: <http://example.org#> .

ex:aThermometer rdf:type prov:Agent ;
  rdf:type ssn:Sensor .

ex:obs rdf:type prov:Entity ;
  rdf:type ssn:Observation .

ex:sensing rdf:type prov:Activity .

ex:obs prov:wasAttributedTo ex:aThermometer ;
  prov:wasGeneratedBy ex:sensing .

ex:sensing prov:wasAssociatedWith ex:aThermometer .
```

LISTING 3.9: Example PROV provenance relationships between entity, agent, and activity.

3.6.8 Provenance

PROV is a specification for provenance designed for the representation of the origins of digital objects in form of descriptions about “the entities and activities involved in producing and delivering or otherwise influencing a given object” (Gil and Miles, 2013). In PROV, provenance is, generally, of entities, which can be physical, digital, or conceptual. Entities can be derived from other entities and they are generated by activities. Activities are the processes through which entities come into existence. Associated with activities are agents, which can be, e.g., persons or, of most interest here, software. Figure 3.9 provides a graphical overview of the relations between PROV entity, activity, and agent. Listing 3.9 demonstrates an example PROV provenance relationships between entity, agent, and activity.

Chapter 4

Experiments and Results

This chapter describes the experiments we conducted and the obtained results. Section 4.1 describes the conducted experiments and Section 4.2 presents the results.

The goal of the experiments is to demonstrate the semantic representation of marine monitoring metadata and data. We utilize the ESONET Yellow Pages as a resource for metadata about sensing device types and FixO3 as a resource for metadata about observatories, including sensing devices, as well as observational data collected from sensing devices. These resources have been described in Chapter 2. Furthermore, we utilize Semantic Technologies for the semantic representation of metadata and data. These technologies are described in Chapter 3.

The experiments and results support the research question and objectives of this dissertation. As we will argue, Semantic Technologies provide a compelling framework for the semantic representation of data and metadata about observatories and their hosted sensing devices—in marine monitoring and arguably beyond. The framework is compelling because of the flexible graph data model; the straightforward linking of (distributed) resources; the existence of reusable ontologies with formal term definitions; and the reasoning capabilities enabled by the technologies.

4.1 Experiments

We first describe the conducted experiments. They are organized in experiments on sensor metadata (Section 4.1.1), experiments on sensor data (Section 4.1.2), and experiments on embedding semantics into sensors (Section 4.1.3). For each experiment, we describe the aims as well as the utilized materials and methods.

The experiments are small scale. We do not attempt to create semantic resources of scale equivalent or comparable to the ESONET Yellow Pages and FixO3. However, the few selected resources are described with some degree of formal semantics, in order to demonstrate the potential of the technologies.

The FixO3 and ESONET Yellow Pages resources are relatively comprehensive. While certainly an interesting and potentially useful effort, creating equivalent semantic resources is beyond the scope of this dissertation. Indeed, this needs be a community effort with broad agreement.

4.1.1 Sensor Metadata

The aim of the first experiment is to demonstrate the semantic representation of observatory and sensing device *metadata* using Semantic Technologies. ESONET Yellow Pages serve as the resource for metadata about sensing device types (models) and FixO3 for metadata about observatories, including attached sensing devices. Instances of the selected sensing device types are part of FixO3 observatories deployed in marine environments.

We consider two sensing device *types*, namely the Teledyne RD Instruments Workhorse Quartermaster 150 kHz ADCP and the PRO OCEANUS Submersible pCO_2 Sensor. Human readable descriptions for these sensing device types can be found at the ESONET Yellow Pages. In particular, the Yellow Pages provide a table with measurement capabilities of these sensing device types, such as operating depth or sampling frequency. We utilize such metadata in creating formal, machine readable, descriptions for the two sensing device types, subclasses of SSN `SensingDevice`. Selected measurement capabilities are represented as value constraints on the SSN `hasMeasurementCapability` property for the respective sensing device type class definition. The value of the property restriction is an individual SSN `MeasurementCapability` having the property SSN `hasMeasurementProperty` to an individual SSN `MeasurementProperty`, itself having the property SSN `hasValue` to a Schema.org `QuantitativeValue`.

We consider two observatories, namely the Southern Adriatic Interdisciplinary Laboratory for Ocean Research (E2-M3A) and the Porcupine Abyssal Plain Observatory (PAP). These observatories are part of the FixO3 collaboratory and observatory network and are described in Chapter 2. Observatories are described with a label, title, description, point coordinates for location, and, most importantly, the hosted sensing devices. Abstractly, observatories are modelled as SSN `Platform`. More concretely, they are instances of `FixedPointOceanObservatory`, which is a kind of `OceanObservatory`. Fixed point ocean observatories relate to sensing devices via the property SSN `attachedSystem`, an inverse of the property SSN `onPlatform`. The E2-M3A observatory has attached an individual of the Teledyne RD Instruments Workhorse Quartermaster 150 kHz ADCP as well as an individual of the PRO OCEANUS Submersible pCO_2 Sensor. In contrast, PAP has only an individual of the PRO OCEANUS Submersible pCO_2 Sensor attached to the observatory.

4.1.2 Sensor Data

The aim of the second experiment is to demonstrate the semantic representation of sensing device observational data using Semantic Technologies. FixO3 is utilized as the resource for observational data.

Observational data are modelled as individuals of the class SSN `Observation`. Following the pattern shown in Figure 3.6, observations are observed by individual sensing devices, here of the Teledyne RD Instruments Workhorse Quartermaster 150 kHz ADCP or the PRO OCEANUS Submersible pCO_2 Sensor. They are for an observed property, such as speed, of a feature of interest, such as water current, from a stimulus, such as the Doppler effect. Observations are made in time and result in an observation value.

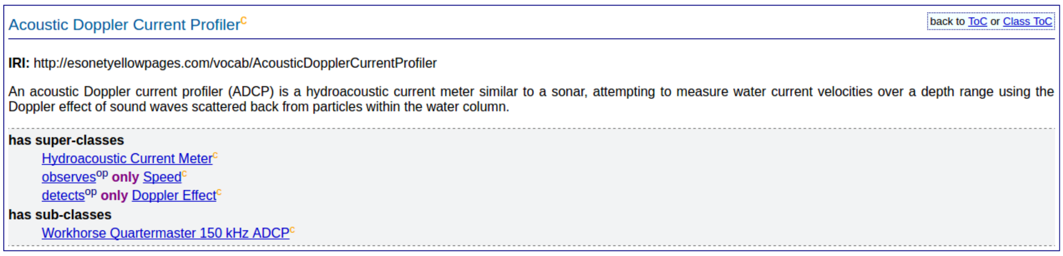
4.1.3 Embedded Semantics

Currently, standardization of data and metadata encoding and format generally occurs after data collection, as part of data acquisition in a computing infrastructure. The encoding and format of collected data is typically determined by vendors of sensing devices. As a consequence, there is large heterogeneity in the encoding and format of data collected from observatories, and even more so across observatories in different earth and environmental science domains.

Following data collection, systems typically try to harmonize the various encodings and formats of collected data by means of conversions. Harmonizing the different syntaxes of collected data as well as the semantics of terms utilized in the representation of collected data and metadata can be laborious, complex, expensive, and error prone. Such conversions are, at least to some extent, also potentially unnecessary, as collected data could in principle be standardized at the source, i.e. by the sensing device or within the observatory.

The aim of the third experiment is to demonstrate the embedding of Semantic Technologies into selected sensing devices, and test the possibility of representing and transmitting device data and metadata using RDF. We test the possibility using a Wave Glider and in collaboration with MARUM, the Center for Marine Environmental Science at the University of Bremen. Wave Glider is an unmanned surface vehicle. It is a platform that hosts one or more sensing devices. Once deployed into a marine environment, the platform propagates by wave motion.

These experiments are part of the ENVRIplus Implementation Case 14 (IC). ENVRIplus is a European Project that brings together experts of 20 environmental research infrastructures as well as experts in information and communication technologies. The IC investigates the standardization of data and metadata transmitted from sensing devices to collecting computer infrastructure. It considers OGC SensorML and O&M as well as RDF and SSN ontology semantic technologies.



Acoustic Doppler Current Profiler^{c}

back to ToC or Class ToC

IRI: http://esonetyellowpages.com/vocab/AcousticDopplerCurrentProfiler

An acoustic Doppler current profiler (ADCP) is a hydroacoustic current meter similar to a sonar, attempting to measure water current velocities over a depth range using the Doppler effect of sound waves scattered back from particles within the water column.

has super-classes

Hydroacoustic Current Meter^{c}
observes^{op} **only** Speed^{c}
detects^{op} **only** Doppler Effect^{c}

has sub-classes

Workhorse Quartermaster 150 kHz ADCP^{c}

FIGURE 4.1: Semantic description of the class `AcousticDopplerCurrentProfiler`.

4.2 Results

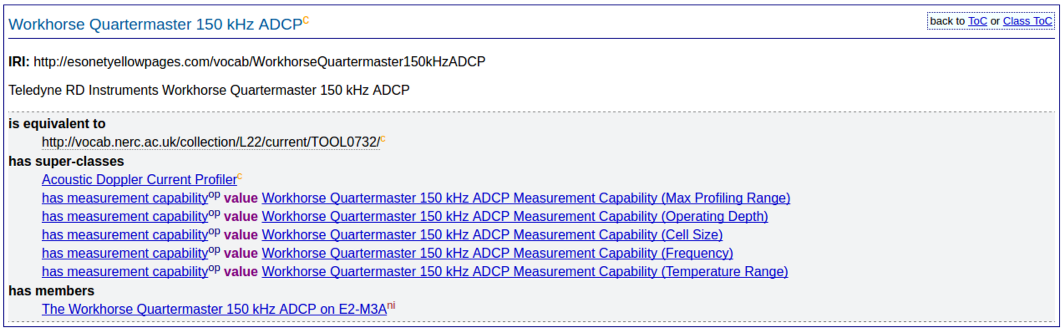
We now present the results of the experiments on semantic representation of metadata about observatories and sensing device types (Section 4.2.1); observational data resulting in marine monitoring (Section 4.2.2); and the embedding of semantic technologies in sensing devices (Section 4.2.3).

4.2.1 Representing Metadata

We describe two sensing device types, namely the Teledyne RD Instruments Workhorse Quartermaster 150 kHz ADCP and the PRO OCEANUS Submersible pCO_2 Sensor. Abstractly, these sensing device types are subclasses of SSN `SensingDevice`, the class of entities that are both SSN `Sensor` and `Device`. The `SensingDevice` class acts as a conceptual anchor for domain specific sensing device types, here in marine monitoring.

The two selected sensing devices types can be further described. For instance, the Teledyne RD Instruments Workhorse Quartermaster 150 kHz ADCP is in fact a type of Acoustic Doppler Current Profiler (ADCP), which is “a hydroacoustic current meter similar to a sonar, attempting to measure water current velocities over a depth range using the Doppler effect of sound waves scattered back from particles within the water column” (Wikipedia, 2016). This definition reveals that ADCPs are hydroacoustic current meters that observe the speed (property) of water current (feature of interest) and detect the Doppler effect (stimulus). Hydroacoustic current meters are SSN sensing devices.

As an example for the semantic description of a sensing device type (class), Figure 4.1 visualizes the semantic description of the acoustic doppler current profiler sensing device type. The visual representation is created using the Live OWL Documentation



Workhorse Quartermaster 150 kHz ADCP^{c}

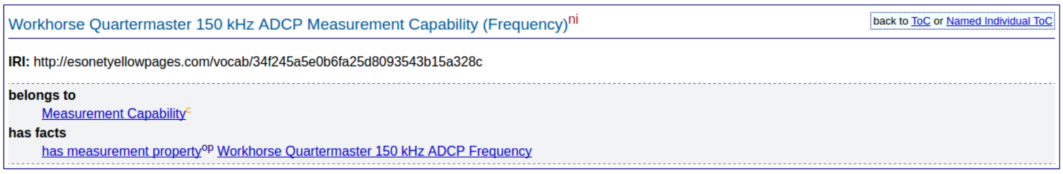
back to ToC or Class ToC

IRI: http://esonetyellowpages.com/vocab/WorkhorseQuartermaster150kHzADCP

Teledyne RD Instruments Workhorse Quartermaster 150 kHz ADCP

is equivalent to
http://vocab.nerc.ac.uk/collection/L22/current/TOOL0732/^{c}
has super-classes
Acoustic Doppler Current Profiler^{c}
has measurement capability^{op} **value** Workhorse Quartermaster 150 kHz ADCP Measurement Capability (Max Profiling Range)
has measurement capability^{op} **value** Workhorse Quartermaster 150 kHz ADCP Measurement Capability (Operating Depth)
has measurement capability^{op} **value** Workhorse Quartermaster 150 kHz ADCP Measurement Capability (Cell Size)
has measurement capability^{op} **value** Workhorse Quartermaster 150 kHz ADCP Measurement Capability (Frequency)
has measurement capability^{op} **value** Workhorse Quartermaster 150 kHz ADCP Measurement Capability (Temperature Range)
has members
The Workhorse Quartermaster 150 kHz ADCP on E2-M3A^{ni}

Figure 4.2: Semantic description of the class `WorkhorseQuartermaster150kHzADCP`.



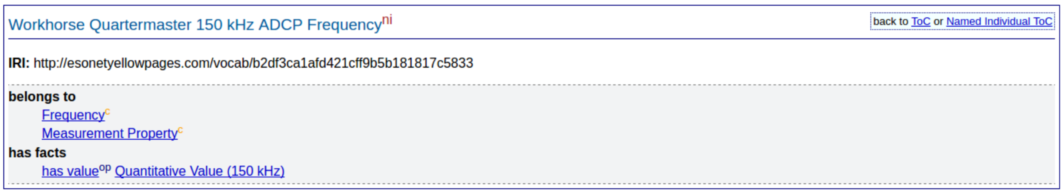
Workhorse Quartermaster 150 kHz ADCP Measurement Capability (Frequency)^{ni}

back to ToC or Named Individual ToC

IRI: http://esonetyellowpages.com/vocab/34f245a5e0b6fa25d8093543b15a328c

belongs to
Measurement Capability^{c}
has facts
has measurement property^{op} Workhorse Quartermaster 150 kHz ADCP Frequency

Figure 4.3: Semantic description of the Workhorse Quartermaster 150 kHz ADCP frequency measurement capability.



Workhorse Quartermaster 150 kHz ADCP Frequency^{ni}

back to ToC or Named Individual ToC

IRI: http://esonetyellowpages.com/vocab/b2df3ca1afd421cff9b5b181817c5833

belongs to
Frequency^{c}
Measurement Property^{c}
has facts
has value^{op} Quantitative Value (150 kHz)

Figure 4.4: Semantic description of the Workhorse Quartermaster 150 kHz ADCP frequency measurement property.

Quantitative Value (150 kHz)^{ni}

back to ToC or Named Individual ToC

IRI: http://esonetyellowpages.com/vocab/7979e8f4e8211a09a88e511ec1b593c9

belongs to
Quantitative Value^{c}
Observation Value^{c}
has facts
value^{dp} *"150.0"^^float*
unit code^{op} kilo hertz

Figure 4.5: Semantic description of the quantitative value 150 kHz.

Environment (Peroni et al., 2012). Figure 4.2 visualizes the semantic description of the Teledyne RD Instruments Workhorse Quartermaster 150 kHz ADCP sensing device type. The description captures formal, i.e. machine readable, semantics of the sensing device type. Specifically, the type is identified by IRI and has a human intelligible label. For machine intelligibility, the description states a number of super-class axioms, in particular value constraints on the SSN `hasMeasurementCapability` property. The description also states that the Teledyne RD Instruments Workhorse Quartermaster 150 kHz ADCP is a kind of acoustic Doppler current profiler and specifies equivalence to

```
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix eyp: <http://esonetyellowpages.com/vocab/> .

eyp:WorkhorseQuartermaster150kHzADCP
  a owl:Class ;
  rdfs:label "Workhorse Quartermaster 150 kHz ADCP"@en ;
  owl:equivalentClass <http://vocab.nerc.ac.uk/collection/L22/current/TOOL0732/> ;
  rdfs:subClassOf eyp:AcousticDopplerCurrentProfiler, [
    a owl:Restriction ;
    owl:onProperty <https://www.w3.org/ns/ssn/hasMeasurementCapability> ;
    owl:hasValue eyp:409fda4bcf98680cfea84e736c533f72
  ], [
    a owl:Restriction ;
    owl:onProperty <https://www.w3.org/ns/ssn/hasMeasurementCapability> ;
    owl:hasValue eyp:75b8b7fbfaa48dedd3cad8e2908e0805
  ], [
    a owl:Restriction ;
    owl:onProperty <https://www.w3.org/ns/ssn/hasMeasurementCapability> ;
    owl:hasValue eyp:3ee84ecb7a3713d2995ea7498f40fee9
  ], [
    a owl:Restriction ;
    owl:onProperty <https://www.w3.org/ns/ssn/hasMeasurementCapability> ;
    owl:hasValue eyp:34f245a5e0b6fa25d8093543b15a328c
  ], [
    a owl:Restriction ;
    owl:onProperty <https://www.w3.org/ns/ssn/hasMeasurementCapability> ;
    owl:hasValue eyp:bc278331051ab7f26cd9a38ab5e8d159
  ] ;
  rdfs:comment "Teledyne RD Instruments Workhorse Quartermaster 150 kHz ADCP"@en .
```

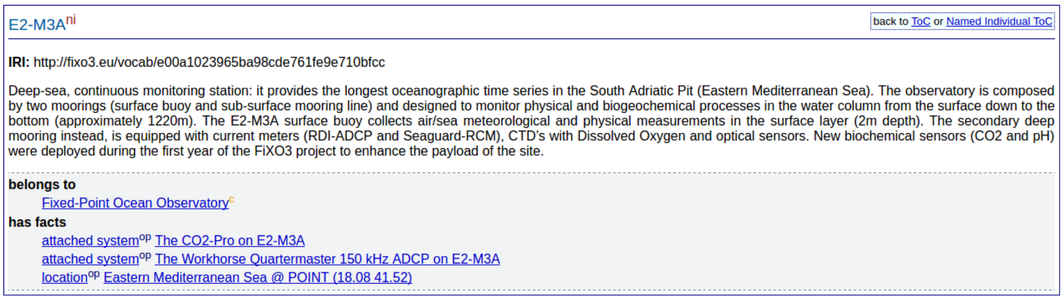
Listing 4.1: Semantic description of the class WorkhorseQuartermaster150kHzADCP (RDF).

the sensing device type description in the NERC vocabulary.[1] Figure 4.3 visualizes the semantic description of the Teledyne RD Instruments Workhorse Quartermaster 150 kHz ADCP *frequency* measurement capability with corresponding measurement property shown in Figure 4.4. The measurement property relates to the quantitative value 150 kHz (Figure 4.5) via the SSN `hasValue` property. Listing 4.1 presents the corresponding description in RDF. These statements can be processed and interpreted by software agents. Note that the OWL `hasValue` property relates to concrete instances of SSN `MeasurementCapability`.

Furthermore, we describe two observatories, namely the Southern Adriatic Interdisciplinary Laboratory for Ocean Research (E2-M3A) and the Porcupine Abyssal Plain Observatory (PAP). Abstractly, observatories are subclasses of SSN `Platform`. As for sensing devices, the class `Platform` acts as a conceptual anchor for domain specific observatories in marine monitoring.

---

[1] http://vocab.nerc.ac.uk/collection/L22/current/TOOL0732/


E2-M3A^{ni}

back to ToC or Named Individual ToC

IRI: http://fixo3.eu/vocab/e00a1023965ba98cde761fe9e710bfcc

Deep-sea, continuous monitoring station: it provides the longest oceanographic time series in the South Adriatic Pit (Eastern Mediterranean Sea). The observatory is composed by two moorings (surface buoy and sub-surface mooring line) and designed to monitor physical and biogeochemical processes in the water column from the surface down to the bottom (approximately 1220m). The E2-M3A surface buoy collects air/sea meteorological and physical measurements in the surface layer (2m depth). The secondary deep mooring instead, is equipped with current meters (RDI-ADCP and Seaguard-RCM), CTD's with Dissolved Oxygen and optical sensors. New biochemical sensors (CO2 and pH) were deployed during the first year of the FiXO3 project to enhance the payload of the site.

belongs to
    Fixed-Point Ocean Observatory^{c}
has facts
    attached system^{op} The CO2-Pro on E2-M3A
    attached system^{op} The Workhorse Quartermaster 150 kHz ADCP on E2-M3A
    location^{op} Eastern Mediterranean Sea @ POINT (18.08 41.52)

Figure 4.6: Semantic description of the E2-M3A fixed-point ocean observatory.

In contrast to the two sensing device types, the two observatories are modelled as individuals, instances of the class `Platform`. E2-M3A and PAP are thus not observatory types but concrete physical entities deployed in a marine environment. However, the class `Platform` is rather abstract. Hence, we specialize this class for our domain specific marine observatories, namely `OceanObservatory` and `FixedPointOceanObservatory`, as further specialization for fixed point platforms. Both E2-M3A and PAP are individuals of the class `FixedPointOceanObservatory`. These individuals, just as any other *resource* described using Semantic Technologies, are identified by URI. Furthermore, ocean observatories are described with spatial location and description. Finally, and most importantly, platforms are described for the attached systems, which are, specifically, sensing devices (individuals).

The fixed-point ocean observatory E2-M3A has attached an individual of the Workhorse Quartermaster 150 kHz ADCP and an individual of the CO_2-Pro sensing device types. In contrast, PAP only has attached an individual of CO_2-Pro. These relations are obtained from the corresponding observatory descriptions made available online by FixO3. Figure 4.6 visualizes the semantic description of the E2-M3A observatory and the corresponding Listing 4.2 presents the description in RDF machine processable and interpretable statements.

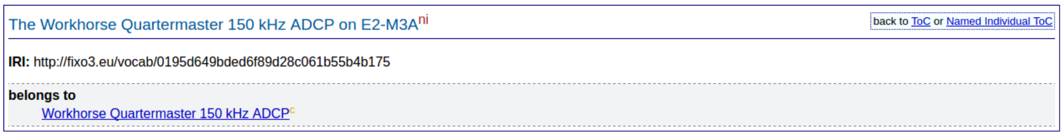
As suggested in Figure 4.6 and Listing 4.2, attached to observatories are concrete instances of sensing device types. Naturally, these resources are also described formally. For instance, the individual of sensing device type Workhorse Quartermaster 150 kHz ADCP attached on the E2-M3A observatory is semantically described as shown in Figure 4.7 and Listing 4.3.

An interesting highlight in descriptions for individual sensing devices attached to observatories is the application of automated reasoning. Only the *direct* type assertion (i.e.

```
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix ssn: <https://www.w3.org/ns/ssn/> .
@prefix schema: <http://schema.org/> .
@prefix fixo3: <http://fixo3.eu/vocab/> .

fixo3:e00a1023965ba98cde761fe9e710bfcc
  a fixo3:FixedPointOceanObservatory ;
  rdfs:label "E2-M3A"@en ;
  dc:source "http://www.fixo3.eu/observatory/e2-m3a/"^^xsd:anyURI ;
  dc:title
    "Southern Adriatic Interdisciplinary Laboratory for Ocean Research"@en ;
  ssn:attachedSystem fixo3:e2c9180abc695bc67abf04f811a2ff6b,
    fixo3:0195d649bded6f89d28c061b55b4b175 ;
  schema:location fixo3:31357ead1ff26f860373676bdf3dc36a .
  rdfs:comment "Deep-sea, continuous monitoring station: it provides the longest
    oceanographic time series in the South Adriatic Pit (Eastern Mediterranean
    Sea). The observatory is composed by two moorings (surface buoy and
    sub-surface mooring line) and designed to monitor physical and biogeochemical
    processes in the water column from the surface down to the bottom
    (approximately 1220m). The E2-M3A surface buoy collects air/sea meteorological
    and physical measurements in the surface layer (2m depth). The secondary deep
    mooring instead, is equipped with current meters (RDI-ADCP and Seaguard-RCM),
     CTD  s with Dissolved Oxygen and optical sensors. New biochemical sensors
    (CO2 and pH) were deployed during the first year of the FiXO3 project to
    enhance the payload of the site."@en ;
```

Listing 4.2: Semantic description of the E2-M3A fixed-point ocean observatory (RDF).


The Workhorse Quartermaster 150 kHz ADCP on E2-M3A ni

back to ToC or Named Individual ToC

IRI: http://fixo3.eu/vocab/0195d649bded6f89d28c061b55b4b175

belongs to
Workhorse Quartermaster 150 kHz ADCP c

Figure 4.7: Semantic description of the Workhorse Quartermaster 150 kHz ADCP attached to the E2-M3A observatory. Includes only explicit assertions.

sensing device of type `WorkhorseQuartermaster150kHzADCP`), as well as the statement asserting the label, are stated explicitly (see Figure 4.7). The other assertions shown in Listing 4.3, are inferred consistently and automatically using a reasoner. As we can see, reasoning annotates all implicit abstract classes, such as `HydroacousticCurrentMeter` or SSN `System`. Moreover, the SSN `onPlatform` relationship with the E2-M3A observatory is inferred by the explicit axiom stating that the property `onPlatform` is an inverse of the property SSN `attachedSystem`. Finally, and perhaps most interestingly, the individual sensing device is automatically related to individual measurement capabilities, specified as value constraints on the sensing device type class description (the identifiers for measurement capabilities in Listing 4.3 are equivalent with those in Listing 4.1).

Finally, Listing 4.4 shows the RDF description of the temperature range measurement property of the Workhorse Quartermaster 150 kHz ADCP measurement capability. The

```
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix ssn: <https://www.w3.org/ns/ssn/> .
@prefix eyp: <http://esonetyellowpages.com/vocab/> .
@prefix fixo3: <http://fixo3.eu/vocab/> .
@prefix nerc: <http://vocab.nerc.ac.uk/collection/L22/current/> .

fixo3:0195d649bded6f89d28c061b55b4b175
  a
    eyp:WorkhorseQuartermaster150kHzADCP
    eyp:AcousticDopplerCurrentProfiler,
    eyp:HydroacousticCurrentMeter,
    ssn:SensingDevice,
    ssn:Device,
    ssn:Sensor,
    ssn:System,
    nerc:TOOL0732;
  rdfs:label "The Workhorse Quartermaster 150 kHz ADCP on E2-M3A"@en ;
  ssn:observes eyp:c656387b92a29bb14edd50267ec1dec6 ;
  ssn:detects eyp:6fd421abc092d7996a4bed6f50e582f6 ;
  ssn:hasMeasurementCapability
    eyp:409fda4bcf98680cfea84e736c533f72,
    eyp:75b8b7fbfaa48dedd3cad8e2908e0805,
    eyp:3ee84ecb7a3713d2995ea7498f40fee9,
    eyp:34f245a5e0b6fa25d8093543b15a328c,
    eyp:bc278331051ab7f26cd9a38ab5e8d159 ;
  ssn:onPlatform fixo3:e00a1023965ba98cde761fe9e710bfcc .
```

Listing 4.3: Semantic description, including inferred assertions, of the Workhorse Quartermaster 150 kHz ADCP attached to the E2-M3A observatory (RDF).

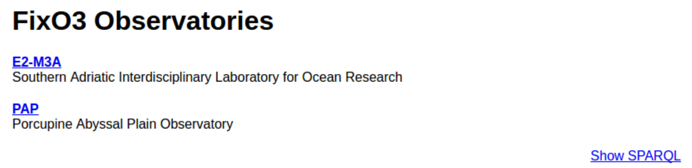
```
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix schema: <http://schema.org/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix ssn: <https://www.w3.org/ns/ssn/> .
@prefix eyp: <http://esonetyellowpages.com/vocab/> .
@prefix unit: <http://qudt.org/vocab/unit#> .

eyp:75b8b7fbfaa48dedd3cad8e2908e0805
  a ssn:MeasurementCapability ;
  rdfs:label "Workhorse Quartermaster 150 kHz ADCP
    Measurement Capability (Temperature Range)"@en ;
  ssn:hasMeasurementProperty eyp:8148568235658792efe60fe4fd7638ec .

eyp:8148568235658792efe60fe4fd7638ec
  a eyp:TemperatureRange, ssn:MeasurementProperty ;
  rdfs:label "Workhorse Quartermaster 150 kHz ADCP Temperature Range"@en ;
  ssn:hasValue eyp:bc78d25c3356a22d5d08ddf1c7ea12ff .

eyp:bc78d25c3356a22d5d08ddf1c7ea12ff
  a schema:QuantitativeValue, ssn:ObservationValue ;
  rdfs:label "Quantitative Value (-5  C to 45  C)"@en ;
  schema:minValue "-5.0"^^xsd:float ;
  schema:maxValue "45.0"^^xsd:float ;
  schema:unitCode unit:DegreeCelsius .
```

Listing 4.4: Semantic description of the temperature range measurement capability of the Workhorse Quartermaster 150 kHz ADCP, including the one attached to the E2-M3A observatory (RDF).



FixO3 Observatories

E2-M3A
Southern Adriatic Interdisciplinary Laboratory for Ocean Research

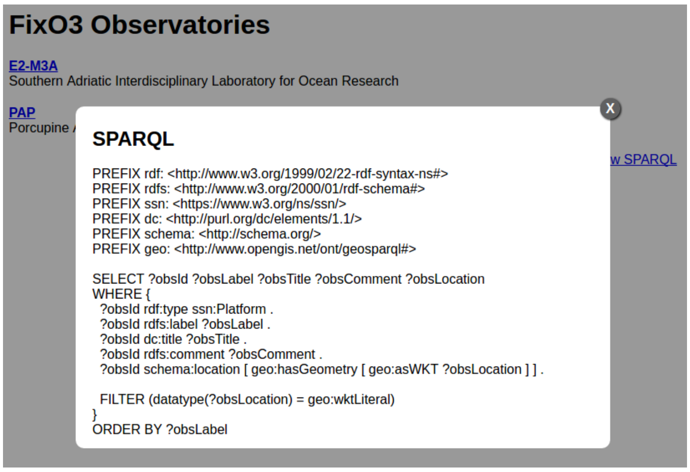
PAP
Porcupine Abyssal Plain Observatory

FIGURE 4.8: Web application that enables browsing observatories and their descriptions managed by the RDF database.

temperature range measurement property relates to a schema.org `QuantitativeValue` with `minValue` and `maxValue` as well as a unit, namely the QUDT Unit `DegreeCelsius`. Naturally, with this link to the QUDT Unit ontology we also gain the QUDT statements about `DegreeCelsius`, such as the `symbol degC` or the type `SIUnit`.

The RDF presented here can be loaded into an arbitrary RDF database (triple store) in order to manage and retrieve RDF. Given such a database, it is straightforward to create a Web application that supports browsing observatories, metadata about them and about the attached sensing devices. Figure 4.8 is a visualization of a Web listing for our observatories. The Web application supports the visualization of the SPARQL query used to retrieve information about the listed observatories (Figure 4.9). Observatories can be selected to obtain a Web visualization of information about them. Figure 4.10 is an example for the E2-M3A observatory. The Web visualization shows the label, description, and the location on a Google Map. Furthermore, it describes the sensors attached to the observatory. Visible in Figure 4.10 is the Workhorse Quartermaster 150 kHz. In addition to the measurement capabilities, we are also provided with information about the observed property (speed), monitored feature (water current), and detected stimulus (Doppler effect). The units are active links (and RDF resources). We can thus obtain further information about them (Figure 4.11).

Listing 4.5 demonstrates how linked ESONET Yellow Pages and FixO3 resource descriptions can be queried using SPARQL. The query selects metadata about FixO3 sensing devices, including the FixO3 platform on which they are attached. Most importantly, the query retrieves the measurement capabilities of sensing devices as they are described for corresponding ESONET Yellow Pages sensing device *types*. Recall that the measurement capabilities are automatically inferred from sensing device type descriptions to sensing devices. The query also highlights the linking to QUDT unit descriptions. Table 4.1 lists the results matching the query.



FixO3 Observatories

E2-M3A
Southern Adriatic Interdisciplinary Laboratory for Ocean Research

PAP
Porcupine /

w SPARQL

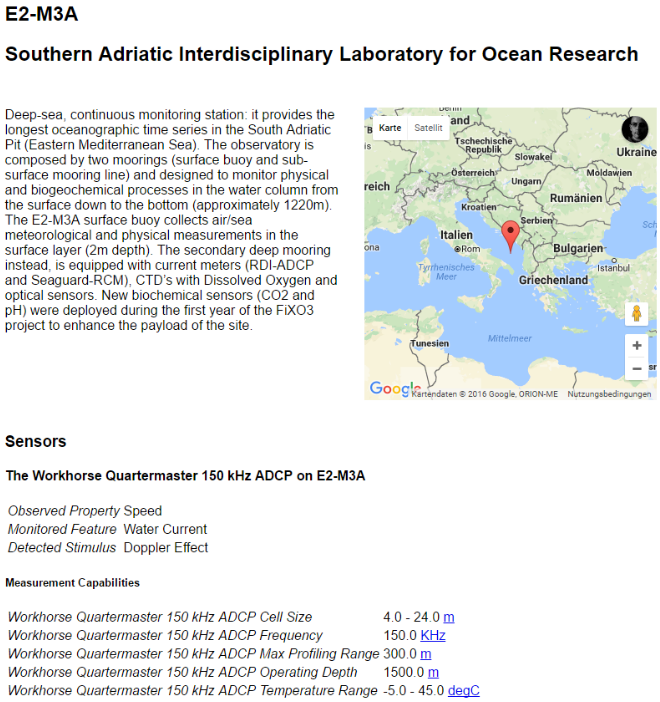
SPARQL

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ssn: <https://www.w3.org/ns/ssn/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX schema: <http://schema.org/>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>

SELECT ?obsId ?obsLabel ?obsTitle ?obsComment ?obsLocation
WHERE {
  ?obsId rdf:type ssn:Platform .
  ?obsId rdfs:label ?obsLabel .
  ?obsId dc:title ?obsTitle .
  ?obsId rdfs:comment ?obsComment .
  ?obsId schema:location [ geo:hasGeometry [ geo:asWKT ?obsLocation ] ] .

  FILTER (datatype(?obsLocation) = geo:wktLiteral)
}
ORDER BY ?obsLabel
```

FIGURE 4.9: The Web application supports the visualization of the SPARQL query used to retrieve information about the observatories and their descriptions managed by the RDF database.



E2-M3A

Southern Adriatic Interdisciplinary Laboratory for Ocean Research

Deep-sea, continuous monitoring station: it provides the longest oceanographic time series in the South Adriatic Pit (Eastern Mediterranean Sea). The observatory is composed by two moorings (surface buoy and sub-surface mooring line) and designed to monitor physical and biogeochemical processes in the water column from the surface down to the bottom (approximately 1220m). The E2-M3A surface buoy collects air/sea meteorological and physical measurements in the surface layer (2m depth). The secondary deep mooring instead, is equipped with current meters (RDI-ADCP and Seaguard-RCM), CTD's with Dissolved Oxygen and optical sensors. New biochemical sensors (CO2 and pH) were deployed during the first year of the FiXO3 project to enhance the payload of the site.

Sensors

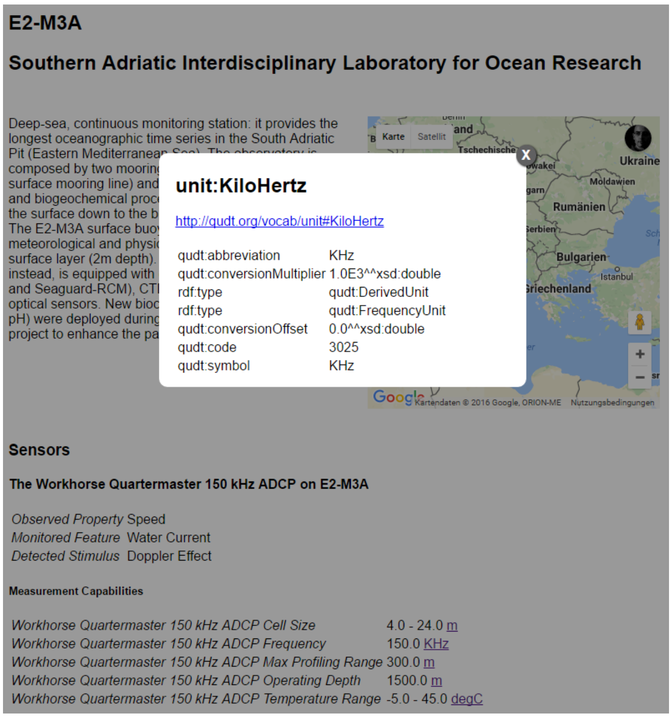
The Workhorse Quartermaster 150 kHz ADCP on E2-M3A

Observed Property Speed
Monitored Feature Water Current
Detected Stimulus Doppler Effect

Measurement Capabilities

Workhorse Quartermaster 150 kHz ADCP Cell Size	4.0 - 24.0 m
Workhorse Quartermaster 150 kHz ADCP Frequency	150.0 KHz
Workhorse Quartermaster 150 kHz ADCP Max Profiling Range	300.0 m
Workhorse Quartermaster 150 kHz ADCP Operating Depth	1500.0 m
Workhorse Quartermaster 150 kHz ADCP Temperature Range	-5.0 - 45.0 degC

FIGURE 4.10: Web visualization of information about the E2-M3A observatory and attached Workhorse Quartermaster 150 kHz ADCP. Units are active links and further information about these resources can be visualized.



E2-M3A

Southern Adriatic Interdisciplinary Laboratory for Ocean Research

Deep-sea, continuous monitoring station: it provides the
longest oceanographic time series in the South Adriatic
Pit (Eastern Mediterranean
composed by two mooring
surface mooring line) and
and biogeochemical proc
the surface down to the b
The E2-M3A surface buo
meteorological and physic
surface layer (2m depth).
instead, is equipped with
and Seaguard-RCM), CTI
optical sensors. New bioc
pH) were deployed during
project to enhance the pa

unit:KiloHertz

http://qudt.org/vocab/unit#KiloHertz

qudt:abbreviation	KHz
qudt:conversionMultiplier	1.0E3^^xsd:double
rdf:type	qudt:DerivedUnit
rdf:type	qudt:FrequencyUnit
qudt:conversionOffset	0.0^^xsd:double
qudt:code	3025
qudt:symbol	KHz

Sensors

The Workhorse Quartermaster 150 kHz ADCP on E2-M3A

Observed Property	Speed
Monitored Feature	Water Current
Detected Stimulus	Doppler Effect

Measurement Capabilities

Workhorse Quartermaster 150 kHz ADCP Cell Size	4.0 - 24.0 m
Workhorse Quartermaster 150 kHz ADCP Frequency	150.0 KHz
Workhorse Quartermaster 150 kHz ADCP Max Profiling Range	300.0 m
Workhorse Quartermaster 150 kHz ADCP Operating Depth	1500.0 m
Workhorse Quartermaster 150 kHz ADCP Temperature Range	-5.0 - 45.0 degC

FIGURE 4.11: Web visualization of information about the kHz unit.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ssn: <https://www.w3.org/ns/ssn/>
PREFIX schema: <http://schema.org/>
PREFIX qudt: <http://qudt.org/schema/qudt#>

SELECT ?observatory ?sensor ?property ?value ?minValue ?maxValue ?unit
WHERE {
  [
    rdf:type ssn:SensingDevice ;
    rdfs:label ?sensor ;
    ssn:onPlatform [ rdfs:label ?observatory ] ;
    ssn:hasMeasurementCapability [
      ssn:hasMeasurementProperty [
        rdfs:label ?propertyLabel ;
        ssn:hasValue ?ov
      ]
    ]
  ] .
  ?ov rdf:type ?vt .
  ?ov schema:unitCode [ qudt:symbol ?unit ] .
  OPTIONAL { ?ov schema:value ?value } .
  OPTIONAL {
    ?ov schema:minValue ?minValue .
    ?ov schema:maxValue ?maxValue
  } .
  FILTER(?vt = ssn:ObservationValue)
}
ORDER BY ?observatory ?sensor ?property
```

LISTING 4.5: SPARQL query demonstrating how linked ESONET Yellow Pages and FixO3 resource descriptions can be queried.

TABLE 4.1: Results for the SPARQL query shown in Listing 4.5.

observatory	sensor	property	value	minValue	maxValue	unit
E2-M3A	The CO2-Pro on E2-M3A			800.0	2000.0	mbar
E2-M3A	The CO2-Pro on E2-M3A			0.0	300.0	m
E2-M3A	The Workhorse Quartermaster 150 kHz ADCP on E2-M3A		150.0			KHz
E2-M3A	The Workhorse Quartermaster 150 kHz ADCP on E2-M3A		1500.0			m
E2-M3A	The Workhorse Quartermaster 150 kHz ADCP on E2-M3A		300.0			m
E2-M3A	The Workhorse Quartermaster 150 kHz ADCP on E2-M3A			-5.0	45.0	degC
E2-M3A	The Workhorse Quartermaster 150 kHz ADCP on E2-M3A			4.0	24.0	m
PAP	The CO2-Pro on PAP			800.0	2000.0	mbar
PAP	The CO2-Pro on PAP			0.0	300.0	m

4.2.2 Representing Data

Having presented the representation of metadata about sensing device types, observatories, and sensing devices attached to observatories, we now present the representation of observational data collected from sensing devices.

Relevant to the representation of observational data is the ontology pattern shown in Figure 3.6. While experts continue to disagree on whether an observation is an activity, an event, or an information object, the pattern suggests that observations are entities

```
@prefix ssn: <https://www.w3.org/ns/ssn/> .
@prefix time: <http://www.w3.org/2006/time#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix schema: <http://schema.org/> .
@prefix fixo3: <http://fixo3.eu/vocab/> .
@prefix unit: <http://qudt.org/vocab/unit#> .
@prefix eyp: <http://esonetyellowpages.com/vocab/> .

fixo3:ca7ff781-3325-4302-8bf3-14a142f27552
  a ssn:Observation ;
  ssn:observationResultTime fixo3:5918b3cb-903b-4b62-82f6-3dd8fba36680 ;
  ssn:observationResult fixo3:1de465d2-3252-4666-aecd-3325ae943c2c ;
  ssn:fromStimulus eyp:6fd421abc092d7996a4bed6f50e582f6 ;
  ssn:featureOfInterest eyp:f9211947db29a7c0590ab410a5c5111b ;
  ssn:observedProperty eyp:da9d4f1ed93c43acc135d6daba0cfe26 ;
  ssn:observedBy fixo3:0195d649bded6f89d28c061b55b4b175 .

fixo3:5918b3cb-903b-4b62-82f6-3dd8fba36680
  a time:Instant ;
  time:inXSDDateTime "2011-06-01T00:00:00.000+02:00"^^xsd:dateTime .

fixo3:1de465d2-3252-4666-aecd-3325ae943c2c
  a ssn:SensorOutput ;
  ssn:hasValue fixo3:393ecdbf-f8ad-4d89-a395-2429dab505ec .

fixo3:393ecdbf-f8ad-4d89-a395-2429dab505ec
  a schema:QuantitativeValue ;
  schema:unitCode unit:MeterPerSecond ;
  schema:value 6.372222e-2 .
```

Listing 4.6: Semantic description of an observation by the Workhorse Quartermaster 150 kHz ADCP on the E2-M3A observatory (RDF).

that relate observation values (here numbers) with time and the property of the feature observed by the sensing device using the stimulus as a proxy.

We have utilized the FixO3 resource to obtain observational data for the described sensing devices and observatories. FixO3 provides observational data as CSV file exports. These are converted into RDF with observational data conforming to the SSN `Observation` pattern. Listing 4.6 is an example SSN `Observation` by the Workhorse Quartermaster 150 kHz ADCP on the E2-M3A observatory observing the speed (property) of water current (feature of interest) by Doppler effect (stimulus).

Naturally, observational data represented as SSN `Observation` RDF resources can be visualized as time series. As the resources are managed in RDF databases, we can express requests for observational data as SPARQL queries and visualize data as time series plots. Figure 4.12 shows an example for observational data by the Workhorse Quartermaster 150 kHz ADCP on the E2-M3A observatory. Listing 4.7 shows the corresponding SPARQL query.

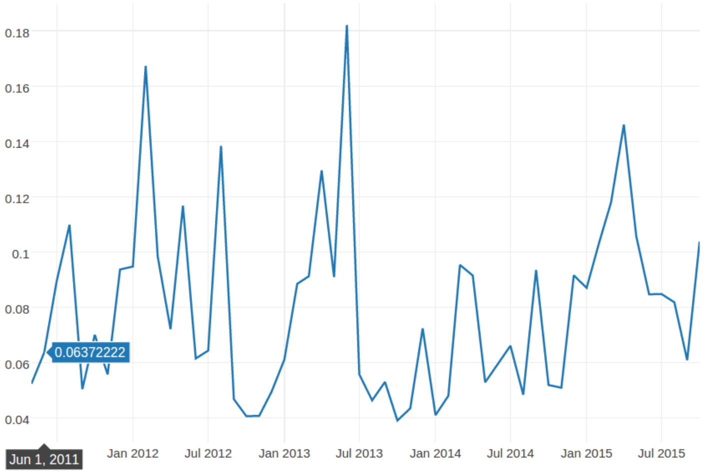


FIGURE 4.12: Web visualization of time series for observational data by the Workhorse Quartermaster 150 kHz ADCP on the E2-M3A observatory.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ssn: <https://www.w3.org/ns/ssn/>
PREFIX schema: <http://schema.org/>
PREFIX time: <http://www.w3.org/2006/time#>
PREFIX fixo3: <http://fixo3.eu/vocab/>

SELECT ?time ?value
WHERE {
  [
    rdf:type ssn:Observation ;
    ssn:observedBy fixo3:0195d649bded6f89d28c061b55b4b175 ;
    ssn:observationResultTime [ time:inXSDDateTime ?time ] ;
    ssn:observationResult [ ssn:hasValue [ schema:value ?value ] ]
  ]
}
ORDER BY ?time
```

LISTING 4.7: SPARQL query used to retrieve the time and value of observational data by the Workhorse Quartermaster 150 kHz ADCP on the E2-M3A observatory. The query may state observed property and feature, as well as stimulus, explicitly (omitted here). This is necessary if a sensing device observes multiple properties.

4.2.3 Embedding Semantics

The experiment on embedding semantic technologies in sensing devices, and transmission of semantic data and metadata from observatories to curation systems, is currently work in progress.

Relevant to experiments on embedding semantics was a workshop organized in at Fraunhofer Institut in Itzehoe, Germany, on May 27, 2016. The workshop brought together representative from industry and academia to discuss future collaborations on marine communication systems. We have presented the application of Semantic Technologies for syntactic and semantic interoperability among heterogeneous marine communication systems. We also highlighted the possibility of embedding semantics into sensing devices. Unfortunately, the workshop has so far not lead to follow-up activities.

Of further and ongoing relevance to experiments on embedding semantics is an Implementation Case (IC) we submitted to the ENVRIplus project. The IC has undergone project internal review. The development of the IC is work in progress but outside of the scope of this dissertation.

The IC aims at standardizing the transmission of data from sensing devices using OGC SWE as well as using the described Semantic Technologies and approaches. Additionally, the IC aims at the implementation of near real-time quality control algorithms suitable for multiple environmental research infrastructures.

The envisioned result will be a demonstration of data and metadata transmission of sensing devices attached to a deployed Wave Glider. Transmission is over IRIDIUM. Transmitted data is expected to be in RDF conformant to the SSN ontology. The resulting streams of data will thus be standardized. The streamed triples can be directly consumed by triples stores for persistence and management as well as by software agents designed to perform quality control over RDF triple streams.

By moving the standardization level into sensing devices we can reduce the required data encoding and format translations at higher levels in environmental research infrastructures. With Semantic Technologies, we can potentially also achieve a harmonization of terms and term semantic utilized in data and metadata descriptions across sensing devices and observatories. Such standardization and harmonization should considerably simplify data acquisition in environmental research infrastructure, and in marine monitoring specifically.

The IC involves several institutions, including MARUM and IFREMER, and over a dozen collaborators. As its further development is planned for 2017, the results on embedding semantics cannot be part of this dissertation and will be available after completing this work. Nonetheless, we included the discussion of this experiment here as embedding semantics into sensing devices is arguably an important next step following the representation of sensing device data and metadata using Semantic Technologies.

Chapter 5

Discussion

For marine monitoring, we have presented how Semantic Technologies can be applied to represent metadata about sensing device types and observatories with attached sensing devices, and to represent observational data collected by sensing devices so that the representation—in particular entity semantics—is machine readable and interpretable.

We have reviewed some of the existing collaboratories, observatories, and data repositories in marine monitoring. While not comprehensive, the review highlights the multitude and heterogeneity of existing networks, technical as well as social.

We have presented Semantic Technologies, in particular the Resource Description Framework, the SPARQL Protocol and RDF Query Language, the notion of Ontology, the ontology languages RDF Schema and Web Ontology Language, as well as an Ontology Framework with defined terminology relevant to marine monitoring, including for time, space, quantities, units, sensing, dataset, quality, and provenance.

Finally, we have utilized some online resources in marine monitoring, specifically the ESONET Yellow Pages and FixO3, and applied the presented Semantic Technologies to demonstrate how these technologies can be utilized to represent metadata about sensing device types and marine monitoring observatories with attached sensing devices, and to represent observational data collected by sensing devices. We have presented the results of this work.

In Section 3.6 we have presented a number of ontologies that each formally describe the semantics of a design pattern for a concept relevant to marine monitoring, e.g.

OWL-Time for the concept of time. Naturally, the set of presented ontologies reflects a particular selection. For most, if not all, of the covered concepts (time, space, quantity, unit, sensing, dataset, quality, provenance) there exist alternative ontologies to the one selected here.

Selecting an ontology over another, or deciding to create a new one, is arguably not a straightforward task. On one hand, for most ontologies there are alternatives with a more or less different formalization of a concept. To take a few examples for themes relevant to marine monitoring, the Basic Geo (WGS84 lat/long) Vocabulary[1] may be considered instead of GeoSPARQL. Supporting only few relations for latitude, longitude, and altitude, the Basic Geo Vocabulary supports representing only point locations, while GeoSPARQL has a richer vocabulary with separation between feature and geometry, as well as representation of complex geometries. Yet, the minimal commitment the Basic Geo Vocabulary requires may be more suited for certain applications, or the representation of certain entities. Another prominent example are ontologies for quantities and units. There exist several alternatives here, too. QUDT supports the representation of both quantities and units.

On the other hand, to experts each of the alternative ontologies generally has deficiencies. As a case in point, in marine monitoring quantities often are value *ranges*, such as the temperature operating range of a sensing device. Applications thus need ontology support for the representation of such quantities. Yet, many of the popular choices, such as QUDT or WURVOC OM, seem to lack support for the representation of value ranges, or it may not be obvious how to represent such quantities. For instance, QUDT seems to entirely lack support for the representation of value ranges. STATO provides the class `Range` as subclass of `MeasurementDataItem` as a pendant to `ScalarMeasurementDatum`. These entities relate to a `ValueSpecification` via the relation `hasValueSpecification`. The ontology provides for the specification of scalar values. However, it is not obvious how to represent ranges. One option could be to introduce has minimum and has maximum value specification properties (sub properties of `hasValueSpecification`). Another option may be to introduce a `RangeValueSpecification` (subclass of `ValueSpecification`). As we can see, even fairly authoritative ontologies can lack support for representing very common entities.

---

[1] https://www.w3.org/2003/01/geo/

One of the primary motivations for selecting a design pattern proposed and formalized by an existing and published ontology, and possibly improve it as necessary, is that, never mind the deficiencies, a team developing an application, e.g. in marine monitoring, is not guaranteed to propose a better pattern. This is at least partially due to the fact that many of these ontologies are the result of shorter or longer processes of deliberation by experts in the respective field. Another strong motivation is that developing applications with commitment to an existing and published pattern improves the interoperability of the application in an ecosystem.

Our review of existing collaboratories and observatories in Chapter 2 highlights the general emphasis on OGC SWE technologies, in particular SensorML, O&M, and SOS. In contrast, the semantic technologies discussed here are rarely mentioned, even though semantic heterogeneity is recognized as a concern at par with syntactic heterogeneity in marine observatory data and metadata. Reasons for the relative popularity of OGC SWE technologies are arguably many. On one hand, EU directives such as INSPIRE endorse OGC SWE technologies with the effect of motivating adoption among Member State infrastructure as well as in research, especially research funded by the European Union. On the other hand, XML-based approaches are arguably more established compared to RDF-based ones. This means that both architects and developers are more familiar with XML technologies. Furthermore, there exist respected implementations, e.g. the suite of products by 52 North. In contrast, for RDF-based technologies frequently there merely exist ontology documents, without implementation support. For instance, Stocker et al. (2015) have argued that we lack a scalable database for time series observation data represented in RDF using the SSN ontology. Arguably, as long as we lack robust and trusted software systems to manage the trillion triples that observatories could generate, RDF-based approaches will probably remain interesting to a small community consisting largely of academics interested in exploring the comparative advantages of these technologies in prototype applications.

Semantic technologies indeed have a few interesting aspects worth highlighting. First, the RDF graph data model is flexible and can accommodate arbitrary connections between, and additions or removal of, resources. This flexibility is an advantage especially for data and metadata with highly variable attributes and with complex relationships. In marine monitoring, metadata about sensing devices or observatories can arguably portray such complexity, as these entities not only display varying attributes but they are also interrelated, and they relate to space, time, organizations, people, projects, funders.

Another interesting aspect of RDF is the use of URI to identify entities. Contrary to identifiers that are local to a database, such as incremental numbers, with URI the entities are "identified globally" and with HTTP URI—the scheme typically used in RDF—they are identified in the global address space of the Web and may be resolved to a Web location (URL). Locating RDF resources on the Web is intuitive for digital entities, such as images or videos. However, RDF is often utilized to describe physical entities that are not "living on the" Web. Rather, they are inhabiting the physical world. This is true also for marine monitoring, where RDF is used to describe entities such as individual sensing devices attached to observatories deployed in a physical environment, e.g. some location at the bottom of the Mediterranean Sea.

A further aspect worth highlighting is that with Semantic Technologies it is possible to standardize not just the syntax of data but also the semantics of terms utilized to describe data. As such, these technologies go well beyond XML or JSON, which are both languages designed to standardize the syntax of data. With Semantic Technologies, the semantics of terms are formal, i.e. machine readable. This opens interesting possibilities. First, we can communicate what a community means by a term to both human and software agents. Second, software agents can automatically reason and infer statements that are logically implied. Assuming consistent semantics, the resulting inferences should be meaningful, "make sense" to human agents.

Finally, with Semantic Technologies, and ontologies in particular, we are given the opportunity to reuse conceptualizations. Rather than to think hard about how to sensibly represent data about concepts relevant to marine monitoring, we can reuse conceptualizations that have been published and are thus available for reuse. This is, at least in principle, powerful because creating ontologies is hard. It is a community driven effort involving a number of experts deliberating about the pros and cons of conceptualizing a theme one or the other way for, often, a very substantial amount of time. Reusing ontologies is thus appealing. Also appealing is that ontology requires a team or community to think hard about semantics. Thinking about semantics in collaboratories is important. Arguably, it is often largely omitted. Assuming definitions are at all discussed and not merely implied, the fact that experts involved in collaboratories often disagree on the definitions of concepts, and thus conceptualize the same domain differently, underscores the importance of ontology and the opportunity these technologies provide to think carefully about semantics. Arguably, the result of the effort pays off as ontologies can be utilized as a tool to communicate semantics to both human agents as well as software agents.

While working on our results, we first attempted to develop an automated translation of data and metadata from ESONET Yellow Pages and FixO3 online resources. This should be feasible as the data is fairly structured. Indeed, for ESONET Yellow Pages there exist XML descriptions corresponding to (some of the) sensing device types available in the registry. For FixO3 the data is not as structured but the resource presents fairly structured Web pages from which the data may be extracted automatically.

Unfortunately, we were too optimistic with automation. As it turns out, even though data and metadata are available (semi-) structured, automating the process of creating ontology axioms and assertions from ESONET Yellow Pages and FixO3 is not as straightforward as one may want it to be. As usual, the difficulties are in the details. For instance, even though ESONET Yellow Pages structures the metadata about sensing device types in tabular way (or in XML), extracting quantities and quantity values automatically will not work without issues. Unfortunately, values not always only consist of actual numeric values but the field may also contain some sort of annotation (e.g. 'Accuracy ±0.5°' as a value for the tilt accuracy measurement characteristic of the Teledyne RD Instruments Workhorse Quartermaster 150 kHz ADCP). Thus, even though these resources are fairly structured, there still is too much heterogeneity in the presentation to support automation reliably.

Incidentally, these difficulties underscore an advantage of ontologies. By using the ontology design patterns described in Section 3.6, there is no doubt about what the literal value of a numeric quantity is because the pattern specifies not just the semantics of numeric quantity but also the data type of the literal value. It is thus clear that such a resource accepts numbers, not numbers and some further annotation included to inform human agents. As software agents can consistency check data automatically, such short cuts are simply not allowed in systems that commit to formalized ontology design patterns.

In Section 3.6 we have presented a number of ontologies which we have not actually used in our experiments. This is because due to time constraints we have limited the experiments to a minimum. However, we discuss here the application of these ontologies to the semantic representation of marine monitoring data and metadata next.

We have utilized the SSN ontology to represent observation data resulting in marine monitoring implemented by sensing devices. Concretely, the data values are represented as literal objects of SSN `ObservationValue` instances, i.e. quantitative values. However, in practice such (multivariate) time series data are typically represented as *dataset*. Using the `Observation` pattern proposed by the SSN ontology results in a representation of the data that explicitly maintains the relationships to the sensing device, stimulus, property, feature, time, and space in an observation. In contrast, datasets often merely represent a series or a multivariate series of quantitative or qualitative values, or may maintain information about sensing device, property, etc. as attribute metadata of the dataset.

Naturally, it is possible to represent the same (multivariate) observation values as a classical dataset in RDF. For this, the ontological framework presented in Section 3.6 has introduced the RDF Data Cube Vocabulary. The RDF Data Cube Vocabulary (QB) introduces the concept QB `DataSet`. Datasets can be described for their structure definition and related QB `Observation` instances, which are the rows of a typical spreadsheet. Observations relate to one or more component property values, which are the values found in spreadsheet cells. Given SSN observation values by sensing devices of an observatory, we can organize the values as QB `Observation` instances, elements of QB `DataSet` instances. A QB `Observation` provides a different view on the same data values and would not maintain the relationships to sensing device, stimulus, property, and feature as do SSN observations. As such, the QB vocabulary is more suited for the representation of a generic dataset. Such representation of data is of interest to marine monitoring collaboratories and research infrastructure to represent processed (observation) data.

Quality of measurement is a central concept in environmental monitoring and marine monitoring, specifically. Indeed, following acquisition from observatories, observation values generally undergo a series of quality control checks. Some sensing devices may perform quality control to some extent. Various schemes designed to compute and assign quality values to observation values have been developed. Rönkkö et al. (2015) present an approach that builds on the quality flagging scheme introduced by the Nordic meteorological institutes. The ontological framework proposed here includes a concept and vocabulary for the representation of quality of measurement in RDF. The pattern supports representing the literal value of a quality measurement, of a metric in a dimension of a category, and computed on a resource. Resources of interest here are the SSN `Observation`, the QB `Observation`, and the QB `DataSet`. Indeed, in marine monitoring we may want to express the quality of measurement of individual sensor observations, processed observations, as well as collections of processed observations. Data quality is central to the ENVRIplus Implementation Case (IC) discussed in Chapter 4 which serves to test embedding semantics into sensing devices as well as implementing quality control on standardized data streams. The ontology design pattern for quality of measurement presented here will be used in development of the IC.

Provenance is another important concept for marine monitoring systems that acquire observation data from sensing devices attached to observatories, and process such data. Hence, the ontological framework introduced here also accounts for this concept. As the SSN observations are acquired from observatories, provenance can represent the relationships between the acquired entities, the activities that generated them, and the agents associated with activities as as well the relationships between the processed entities, and the activities and agents relevant to processing. Hence, provenance provides a view that emphasizes the life cycle of the entities acquired and processed in marine monitoring collaboratories.

A technical concern not addressed here is the management of the billions and potentially trillions of RDF triples that may be generated and processed in marine monitoring collaboratories. The large amount of triples, and the fact that triples acquired from observatories are streamed, are challenges for conventional triple stores (RDF databases). Some of these difficulties have been demonstrated by Stocker et al. (2015) where load and query performance of two conventional triple stores are compared to an Apache Cassandra powered database for SSN `Observation` time series. The experiments showed that at scale of a billion or more triples conventional triples stores perform increasingly poorly in both loading streamed triples as well as in querying time series. This is not surprising, because conventional triple stores are designed to support arbitrary complex graph pattern queries. Furthermore, they perform better at bulk loading large amounts of triples, rather than at streaming triples into a database. In the environmental monitoring domain, queries typically filter time series data for a particular interval and then utilize the resulting data for some purpose. Thus, queries tend to follow a fairly predictable pattern and are not arbitrary complex graph patterns.

The lack of a database that scales to billions and possibly trillions of SSN and QB observations is an important issue for the adoption of Semantic Technologies in marine monitoring. Clearly, a collaboratory cannot afford to start with developing of a database system. They lack the resources for such an effort and, more importantly, developing database systems is not of primary concern to marine monitoring and science collaboratories. Such collaboratories rely on a triple store that is capable of managing the generated trillion triples. If such a triple store is not available, then representing observation data in RDF is hardly a viable option.

We also discussed the possibility of embedding semantics into sensing devices. However, we could not report results here because the time horizon required for such developments by far exceeds the one available for this dissertation. To a large extent, this is due to the difficulty of interacting with vendors or to obtain the necessary hardware, software development kits, and skills to develop a prototype. The workshop at Fraunhofer Institut in Itzehoe, Germany, on May 27, 2016 attended by representatives of both academia and industry has also underscored the challenge of communicating the potential of these technologies to vendors. Adherence to standards seems not to be of priority to vendors who continue, perhaps for good reasons, to develop sensing devices with data output that does not conform with widely accepted standards, e.g. OGC SensorML and O&M. As the workshop highlighted, Semantic Technologies are of little interest, at least to the vendors who attended this workshop. There are probably several reasons, including the gap between these technologies and the concrete issues vendors currently deal with, the abstract character of Semantic Technologies and their non-palpable advantages, or the lack of expertise for such technologies.

5.1 Strengths

The approach for semantic representation of marine monitoring metadata and data discussed here has several interesting aspects and strengths worth highlighting.

Semantic Technologies require a team or community to think deeply about the relevant vocabulary and its semantics, more than is required using other technologies. Surely, developers of a system that uses JSON to structure data and metadata communicated between systems also need to decide on the vocabulary (keywords) used. However, in JSON the keywords used are merely strings. The same keyword may mean something different in different JSON document (possibly even in the same document). In short, with JSON the team or community that architects the system is prone to skip the effort of conceptualizing the domain first. It is thus less likely that the community realizes the different understanding members of the community often have for same concepts. The result of weak informal conceptualization is that the keywords used in structured documents are not well defined, certainly not to software agents. XML technologies are more advanced compared to JSON with respect to supporting specifications of vocabularies. However, XML documents need to be well-formed. Contrary to JSON, XML documents can be validated. Validation is performed by checking the conformance of an XML document with the associated schema. An XML schema describes a type of XML document by specifying constraints on the structure and content. Constraints may specify elements and attributes that must or may be included; the permitted structure of elements, specified by a regular expression syntax; and the data type of character data. In contrast to these technologies, Semantic Technologies and, specifically, RDFS and OWL go beyond the specification of the permitted structure (syntax) of documents to support the specification of semantics. Moreover, XML is a serialization format. In contrast, RDF is a data model. In other words, XML is designed to specify how to structure data while RDF is designed to represent the information content (semantics) of data. To draw an analogy, consider a book in printed and digital forms, and its information content. XML is concerned with the serialization, print or digital, while RDF is concerned with the meaning of the information content.

As a graph-based data model with resources identified globally on the Web, RDF is interesting for its support of distributed data linking. As we have shown, in RDF linking two resources only requires a relationship between the two resources. The resources may both reside in a local namespace, a local resource may link to a remote one, or we may relate two resources in (different) remote namespaces. For instance, in our experiments we linked a local resource for a quantitative value with a resource of a remote namespace (QUDT). The flexibility of the graph-based data model for creating or deleting arbitrary links between distributed resources is a powerful mechanism to link data, and thus reuse resources made available on the Web, in particular vocabularies published on the Web, such as a vocabulary for units.

Linking distributed resources is indeed central to marine monitoring data and metadata. As we have shown in our experiments, linking distributed resources is at the core of ESONET Yellow Pages and FixO3 resources, where FixO3 sensing device instances attached to observatories make reference to sensing device types described by ESONET Yellow Pages. As we have seen, resources in ESONET Yellow Pages and FixO3 namespaces can be related with a single RDF `type` relationship of the form

```
http://fixo3.eu/devices/d
  rdf:type http://esonetyellowpages.com/types/t
```

This simple mechanism is arguably powerful and its potential in marine observatories and collaboratories is surely unexplored. The possibilities are arguably many. As any RDF resource pair can be linked, observatories could not just link sensing device types and instances but also observed properties, their features, the data structure definitions of datasets, datasets themselves, or individual observation quantitative values.

The ubiquitous use of HTTP URIs to identify resources in RDF is also interesting in regard to Persistent Identifier (PID) infrastructure. A PID is a association between a character string and a resource (Hakala, 2010). Typical resources are digital copies of articles identified by Digital Object Identifier (DOI). DOI is increasingly used also for resources other than literature. A prominent example are datasets and the adoption of DOI by data publishers for the identification and citation of data publications. Beyond articles and datasets, DOI is utilized for all kinds of research objects, including figures and software. PANGAEA (Diepenbroek et al., 2002), the Data Publisher in Earth & Environmental Science, is an example data publisher that adopts DOI for the identification and citation of data. Other repositories such as Zenodo, figshare, and Dryad are further examples. DOIs are governed, meaning that DOIs need to be ‘minted’ via a registration agency and the association between the DOI and the URL that specifies the location of the digital object, as well as metadata about the digital object, are maintained by the agency. Crossref and DataCite are example registration agencies for articles and data, respectively.

In addition to digital objects, persistent identifiers are utilizes also for physical objects as well as agents and concepts. A prominent example is ORCID, the Open Researcher and Contributor Identifier. At ORCID, researchers can sign-up to obtain a unique identifier which researchers can use to unambiguously identify themselves as contributors to research artefacts, such as articles and data. ORCID is increasingly integrated in both article and data submission workflows. Publishers are increasingly often actively collecting, or even requiring, author ORCID iDs. PANGAEA has recently integrated ORCID (de Mello et al., 2016) to enable linking contributor ORCID iDs with data publication DOIs. PIDs are also utilized to identify instrumentation or physical platforms carrying instrumentation such as research vessels (cruises).

As PIDs are typically extended to actionable HTTP URIs to enable resolution on the Web, they fit the RDF data model nicely. Indeed, as HTTP URIs, there are several ways to integrate PIDs in RDF. HTTP PIDs can be used to name RDF resources. Rather than using `http://fixo3.eu/devices/d` as the URI for our FixO3 device `d`, we could utilize the DOI for the device, assuming one was minted for it. Alternatively, we could relate the DOI with the FixO3 device URI as 'related identifiers' (e.g. `swrc:doi`). This simple mechanism could be used to persistently identify all kinds of resources in marine observatories, including SSN sensing devices or QB data structure definitions and datasets. Sensing device types described by ESONET Yellow Pages as RDF resources could have a DOI related identifier and would be persistently identified by a trusted mechanism. We could thus cite sensing device types, or concrete instances.

5.2 Limitations

The dissertation, the discussed approaches, as well as many of the discussed collaboratories face a number of limitations.

The dissertation is very limited in the breath and depth of its experiments. We have merely scratched the surface. We have only considered two online resources, ESONET Yellow Pages and FixO3. As we have shown in Chapter 2, there are many more. Furthermore, of these resources we have just taken a few sensing device types and a few observatories. We have not provided in-depth analysis of what the technologies can do, especially not at scale, and we have not made significant progress on embedding semantics into sensing devices. Many issues remain thus unresolved.

The main reason for having pursued only limited experiments was the limited resources available for this dissertation. Broader experiments covering more observatories and more data and metadata clearly demand more resources. The development cycle for the experiment on embedding semantics are obviously longer, as it involves a number of third parties, including sensing device manufactures. Having only scratched the surface, this work is limited to contributing to ongoing community discussions and efforts aiming at adoption of Semantic Technologies in environmental monitoring, specifically marine monitoring.

The discussed approaches also have a number of limitations and challenges to face. Overall, the arguably steep learning curve hinders the adoption of Semantic Technologies. While RDF is conceptually straightforward and not complicated to use in practice, the technologies that enable the formal description of semantics—namely RDFS, OWL, rule languages, and reasoners—are both conceptually difficult and complicated to use in practice.

Furthermore, there is a lack of robust databases and other tools for managing the RDF data and metadata generated and used in marine monitoring, and environmental monitoring more generally. The obvious example is a database for time series data represented in RDF. We think this is a major concern for adoption of these technologies in marine and environmental monitoring. Aside databases, the domain also lacks of specialized software tools that support the description of observatories, attached sensing devices, collected data.

The field also lacks qualified man power and funding. Judging from the prolific literature as well as the many companies that specialize on the technologies, there certainly exist experts in Semantic Technologies. However, their numbers in earth and environmental informatics or marine monitoring and science is low, also compared to other sciences, such as life science and bioinformatics which have seen greater adoption of these technologies. These differences are at least partially probably also due to differences in available funding.

The EU INSPIRE Directive is arguably another hindrance to adoption of Semantic Technologies in earth and environmental science to represent geospatial data. The INSPIRE Directive "aims to create a European Union spatial data infrastructure for the purposes of EU environmental policies and policies or activities which may have an impact on the environment".[2] The Directive entered into force in 2007. While the INSPIRE Implementing Rules do not mandate particular standards or technologies, the INSPIRE Technical Guidance documents make clear reference to OGC standards by including "details of what aspects of specific OGC/ISO standards are to be implemented".[3] EU Member States and their organization are thus strongly guided toward the adoption of OGC standards, which continue to be based on XML and are thus not considered Semantic Technologies. INSPIRE guidance thus hinders the adoption of the technologies and approaches discussed here.

An important concern at the level of collaboratory and challenge many collaboratories face is the issue of sustainability. European observatories, and the collaboratories that build and maintain the observatories, are generally funded via project funding cycles—in particular initial development phases. With the European Research Infrastructure Consortium (ERIC), environmental research infrastructures have at their disposition a legal framework that also addresses long-term funding. EMSO is an example research infrastructure discussed in this dissertation that has attained ERIC status. Thanks to long-term funding commitments from Member States, environmental research infrastructures that have attained ERIC status are arguably more sustainable. However, the vast majority of small and large observatories and collaboratories continue to operate on funding horizons each lasting a few years. Such short time horizons are problematic for the planning, construction, and maintenance of complex projects, as well as for science itself, as it benefits greatly from long-term and *continuous* data acquisition, processing, and interpretation.

5.3 Related Work

Sensor data management is a research topic of continued and increasing importance. Balazinska et al. (2007) note that "we have placed too much attention on the networking

---

[2] http://inspire.ec.europa.eu/about-inspire/563

[3] http://www.opengeospatial.org/pressroom/marketreport/inspire ... cite this!

of distributed sensing and too little on tools to manage, analyse, and understand the data." The authors underscore that conventional database management systems "have several critical shortcomings that prevent using them directly to process live sensor data." Hence, alternative databases and systems have been proposed in the literature that address the challenges of sensor data management (Marascu et al., 2014; Chang et al., 2006; Hill et al., 2011; Horsburgh et al., 2011; Dow et al., 2015). Due to high heterogeneity, volume, and frequency sensor data are notoriously challenging to process. and various stream processing and querying engines have been proposed (Babu and Widom, 2001; Yao and Gehrke, 2002; Chandrasekaran et al., 2003; Madden et al., 2005; Abadi et al., 2005).

Even though Semantic Technologies compound the challenges of streamed data management and processing, the semantic description of sensors and their data using ontologies has received considerable attention in the literature. Sheth et al. (2008) discuss the semantic sensor web "in which sensor data [are] annotated with semantic metadata to increase interoperability as well as provide contextual information essential to situational knowledge."

The semantic sensor web extends the sensor web, which "refers to web accessible sensor networks and archived sensor data that can be discovered and accessed using standard protocols and application program interfaces" (Botts et al., 2007, 2008). Toward the sensor web vision, the Open Geospatial Consortium (OGC) develops "a suite of specifications related to sensors, sensor data models, and sensor web services" (Sheth et al., 2008), designed for syntactic interoperability between information systems (Egenhofer, 2002); they do not address semantic interoperability.

In an early attempt to tackle this limitation, Probst (2006) suggests to align key terms of OGC O&M to the DOLCE foundational ontology (Masolo et al., 2002). With the semantic sensor web, Sheth et al. (2008) extend the syntactic XML-based metadata standards of the OGC with OWL-based semantic metadata standards of the W3C. Sheth et al. propose a mechanism whereby semantics are added into XML documents by annotating (OGC) XML with terms defined in ontologies. The authors demonstrate the mechanism by annotating a timestamp encoded in O&M with the term OWL-Time `Instant`.

Ontologies were soon developed. Compton et al. (2009) provide a survey of early efforts in the semantic specification of sensors. Some ontologies are designed to primarily support the description of sensor types, others focus on sensor data, and some support the description of sensor systems, their components, structure, and processes. Today, the most notable result in the semantic specification of sensors is arguably the SSN ontology.

The adoption of Semantic Technologies for sensor data management has been advocated. Lewis et al. (2006) argue that "semantics can enhance data management in sensor networks." The system presented by Lewis et al. manages sensor data in (daily) RDF files, a practice that, in light of state of the art RDF databases, is arguably antiquate. However, the authors underscore the ability of the RDF (graph) data model to represent semantic associations between data, and the possibility of using such relations in the formulation of queries.

Le-Phuoc et al. (2011) present a Linked Stream Middleware for the collection of heterogeneous sensor data, their translation into RDF data that conforms with the SSN ontology, and the access of RDF data. The middleware is designed to support the collection and transformation of large volume data. The authors claim that their instance handles over 100,000 data sources.

Wang et al. (2011) describe a "semantic technology-based approach to ecological and environmental monitoring." The authors develop an upper ontology for monitoring, and deploy the approach in a system that "integrates environmental monitoring and regulation data from multiple sources" using Semantic Technologies.

Lefort et al. (2012) use the SSN ontology and the QB vocabulary to publish temperature data released by the Australian Bureau of Meteorology as tabular time series (tab-delimited data files) in RDF as a 'Linked Sensor Data Cube'.

Ahmedi et al. (2013) present an ontology for water quality management. The proposed ontology was "developed to support water quality classification based on different regulation authorities." The ontology is based on the SSN ontology and supports the modelling of sensor data, regulations published by authorities, sources of pollution, and expert knowledge about the water (quality) domain, in particular rules.

Abecker et al. (2014) present a sensor and semantic data warehouse “able to store and provide sensor, measurement and forecasting data, as well as semantic knowledge about the water-supply chain.” The software architecture separates sensor and semantic data into distinct stores. Specifically, sensor data in form of OGC WaterML 2.0 (Taylor, 2014) are managed by a conventional relational database management system whereas data with irregular and complex relationships are managed by a knowledge base. According to Abecker et al., to manage the sensor data with the RDF database “seemed not feasible and promising.” A drawback of the approach is the resulting ‘technology gap’ which means that it is not possible to evaluate sensor data and semantic data in a single query.

The semantic annotation of sensor data is a subtask of sensor data management with Semantic Technologies, and is frequently discussed in the literature. Studied for a building fire emergency scenario, Huang and Javed (2008) present an architecture for a system that enriches sensor data with semantic information. Wei and Barnaghi (2009) propose to annotate sensor data with concepts of existing knowledge bases, such as DBpedia (Auer et al., 2007), following the linked data principle, and utilize semantic reasoning over sensor data to infer new knowledge and answer complex user queries. Moraru and Mladenić (2012) present a framework for the semantic enrichment of sensor data. Calbimonte et al. (2012) propose to learn semantic properties of observations from sensor data.

Another subtask of sensor data management is data access. In order to abstract from the heterogeneity of devices in sensor networks, service oriented principles have been adopted to model sensors as services and thus enable access to sensor data through standard service technologies. With semantic sensor service networks, Wang et al. (2012) propose a generic framework that models sensors as services and supports the semantic description, seamless service-oriented connectivity, discovery and composition of sensor services. Of concern to sensor services are, among other issues, semantic registries for sensor metadata (Chaves et al., 2013) and the matching of sensor characteristics and service requirements for correct integration (Bröring et al., 2012). Some of the OGC service specifications have been extended with semantic features. An example is the semantically enabled SOS proposed by Henson et al. (2009).

Calbimonte et al. (2010) discuss the ontology-based access to data streams and present a SPARQL extension for streaming data that supports operators over RDF streams. An

RDF stream is a sequence of pairs consisting of an RDF triple and a timestamp. Similar extensions to SPARQL have been proposed in the literature (Bolles et al., 2008; Barbieri et al., 2009). The efficient transmission of RDF streams using data compression techniques has also been addressed (Fernández et al., 2014). Stream reasoning, i.e. logical reasoning on data streams, has also drawn interest recently (Della Valle et al., 2009; Barbieri et al., 2010; Margara et al., 2014) and Wetz et al. (2014) discuss the integration of RDF streams in environmental information systems, noting that the blending of static data sources and dynamic data streams "is non-trivial and major advances still need to be made in this area."

Chapter 6

Conclusion

A large and increasing number of instruments is deployed into Earth's seas and oceans by numerous collaboratories that lead the design, deployment and maintenance of marine observatories. The insufficient coordination among collaboratories means that there is little interoperability between their data resources. Specifically, it is not possible to (programmatically) interact with their resources in an integrated manner. As a result, users need to understand how to formulate their data needs as separate requests submitted to distinct systems. Furthermore, users are required to join the results returned by the systems. This typically involves considerable effort in reconciling the syntax and semantics of disjoint metadata and data sets.

We have discussed and demonstrated a set of technologies and approaches that support linking distributed data resources of distinct collaboratories so that users can formulate their data needs in a federated manner on distributed systems. Specifically, the research question explored if Semantic Technologies can increase the interoperability of metadata about marine observatories and of their observational data. The three research objectives aimed at (1) the semantic representation of *metadata* about observatories and attached sensing devices utilized in marine monitoring; (2) the semantic representation of *observational data* collected from observatories; and (3) the *linking* of metadata and data about these heterogeneous resources.

Much remains to be done. As argued earlier, we only scratch the surface of what may be possible. Concretely, it would be useful to implement the approach sketched here in solid test cases, e.g. in at least one collaboratory and two observatories. We have presented the approach discussed here at the 3rd ENVRIweek in Prague, November 14-18, 2016. While most people are unfamiliar, even unimpressed, by the approach, there certainly are some who understand the potential and are committed to bringing Semantic Technologies into application in marine and environmental monitoring.

We are thus cautiously optimistic that future data resources in marine and environmental monitoring may be developed so that they appear to scientists, and users more generally, as (ideally) one globally integrated distributed database.

Undoubtedly, we will need to solve technology challenges along the way. However, it seems that, today, the technology readiness is higher than the readiness of the community to adopt them. Thus, we need a push by the few to convince the many about the potential of these technologies and the need for investment in this direction.

Bibliography

Abadi, D. J., Ahmad, Y., Balazinska, M., Çetintemel, U., Cherniack, M., Hwang, J.-H., Lindner, W., Maskey, A. S., Rasin, A., Ryvkina, E., Tatbul, N., Xing, Y., and Zdonik, S. (2005). The Design of the Borealis Stream Processing Engine. In *Second Biennial Conference on Innovative Data Systems Research (CIDR 2005)*, pages 277–289, Asilomar, CA.

Abecker, A., Brauer, T., Magoutas, B., Mentzas, G., Papageorgiou, N., and Quenzer, M. (2014). A Sensor and Semantic Data Warehouse for Integrated Water Resource Management. In Gómez, J. M., Sonnenschein, M., Vogel, U., Winter, A., Rapp, B., and Giesen, N., editors, *Proceedings of the 28th Conference on Environmental Informatics - Informatics for Environmental Protection, Sustainable Development and Risk Management (EnviroInfo 2014)*, pages 517–524, Oldenburg, Germany. BIS-Verlag, Oldenburg.

Ahmedi, L., Jajaga, E., and Ahmedi, F. (2013). An Ontology Framework for Water Quality Management. In Corcho, O., Henson, C., and Barnaghi, P., editors, *Proceedings of the 6th International Workshop on Semantic Sensor Networks*, volume 1063, pages 35–50, Sydney, Australia. CEUR-WS.

Albertoni, R. and Isaac, A. (2016). Data on the Web Best Practices: Data Quality Vocabulary. Editor's draft, W3C.

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., and Ives, Z. (2007). DBpedia: A Nucleus for a Web of Open Data. In Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., and Cudré-Mauroux, P., editors, *The Semantic Web*, volume 4825 of *Lecture Notes in Computer Science*, pages 722–735. Springer Berlin Heidelberg.

Babu, S. and Widom, J. (2001). Continuous Queries over Data Streams. *SIGMOD Rec.*, 30(3):109–120.

Balazinska, M., Deshpande, A., Franklin, M., Gibbons, P., Gray, J., Nath, S., Hansen, M., Liebhold, M., Szalay, A., and Tao, V. (2007). Data Management in the Worldwide Sensor Web. *Pervasive Computing, IEEE*, 6(2):30–40.

Barbieri, D., Braga, D., Ceri, S., Valle, E. D., and Grossniklaus, M. (2010). Stream Reasoning: Where We Got So Far. In *Proceedings of the 4th International Workshop on New Forms of Reasoning for the Semantic Web: Scalable & Dynamic*, pages 1–7, Heraklion, Crete, Greece.

Barbieri, D. F., Braga, D., Ceri, S., Della Valle, E., and Grossniklaus, M. (2009). C-SPARQL: SPARQL for Continuous Querying. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 1061–1062, New York, NY, USA. ACM.

Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F., and Stein, L. A. (2004). OWL Web Ontology Language Reference. Recommendation, W3C.

Beckett, D. (2004). RDF/XML Syntax Specification (Revised). Recommendation, W3C.

Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H. F., and Secret, A. (1994). The World-Wide Web. *Commun. ACM*, 37(8):76–82.

Berners-Lee, T., Fielding, R., and Masinter, L. (2005). Uniform Resource Identifier (URI): Generic Syntax. Request for Comments 3986, IETF.

Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5):29–37.

Bolles, A., Grawunder, M., and Jacobi, J. (2008). Streaming SPARQL - Extending SPARQL to Process Data Streams. In Bechhofer, S., Hauswirth, M., Hoffmann, J., and Koubarakis, M., editors, *The Semantic Web: Research and Applications*, volume 5021 of *Lecture Notes in Computer Science*, pages 448–462. Springer Berlin Heidelberg.

Borst, W. N. (1997). *Construction of Engineering Ontologies for Knowledge Sharing and Reuse.* PhD thesis, Centre for Telematics and Information Technology, University of Twente, Enschede, The Netherlands.

Botts, M., Percivall, G., Reed, C., and Davidson, J. (2007). OGC© Sensor Web Enablement: Overview And High Level Architecture. OGC White Paper OGC 07-165, Open Geospatial Consortium Inc.

Botts, M., Percivall, G., Reed, C., and Davidson, J. (2008). OGC© Sensor Web Enablement: Overview and High Level Architecture. In Nittel, S., Labrinidis, A., and Stefanidis, A., editors, *GeoSensor Networks*, volume 4540 of *Lecture Notes in Computer Science*, pages 175–190. Springer Berlin Heidelberg.

Botts, M. and Robin, A. (2007). OpenGIS Sensor Model Language (SensorML) Implementation Specification. OpenGIS Implementation Specification OGC 07-000, Open Geospatial Consortium Inc.

Bray, T., Paoli, J., and Sperberg-McQueen, C. M. (1998). Extensible Markup Language (XML) 1.0. Recommendation, W3C.

Brickley, D. and Guha, R. (2004). RDF Vocabulary Description Language 1.0: RDF Schema. Recommendation, W3C.

Brickley, D. and Guha, R. (2014). RDF Schema 1.1. Recommendation, W3C.

Bröring, A., Maué, P., Malewski, C., and Janowicz, K. (2012). Semantic mediation on the Sensor Web. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 2910–2913. Institute of Electrical & Electronics Engineers (IEEE).

Calbimonte, J.-P., Corcho, O., and Gray, A. J. (2010). Enabling Ontology-Based Access to Streaming Data Sources. In Patel-Schneider, P. F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J. Z., Horrocks, I., and Glimm, B., editors, *The Semantic Web – ISWC 2010*, volume 6496 of *Lecture Notes in Computer Science*, pages 96–111. Springer Berlin Heidelberg.

Calbimonte, J.-P., Yan, Z., Jeung, H., Corcho, O., and Aberer, K. (2012). Deriving Semantic Sensor Metadata from Raw Measurements. In Henson, C., Taylor, K., and Corcho, O., editors, *Proceedings of the 5th International Workshop on Semantic Sensor Networks*, volume 904, pages 33–48, Boston, Massachusetts, USA. CEUR-WS.

Carothers, G. and Seaborne, A. (2014). RDF 1.1 N-Triples: A line-based syntax for an RDF graph. Recommendation, W3C.

Chamberlin, D. D. and Boyce, R. F. (1974). SEQUEL: A Structured English Query Language. In *Proceedings of the 1974 ACM SIGFIDET (Now SIGMOD) Workshop on Data Description, Access and Control*, SIGFIDET '74, pages 249–264, New York, NY, USA. ACM.

Chandrasekaran, S., Cooper, O., Deshpande, A., Franklin, M. J., Hellerstein, J. M., Hong, W., Krishnamurthy, S., Madden, S., Raman, V., Reiss, F., and Shah, M. A. (2003). TelegraphCQ: Continuous Dataflow Processing for an Uncertain World. In *Proceedings of the 1st Biennial Conference on Innovative Data Systems Research*, Asilomar, CA, USA.

Chang, K., Yau, N., Hansen, M., and Estrin, D. (2006). SensorBase.org – A Centralized Repository to Slog Sensor Network Data. In *Proceedings of the International Conference on Distributed Networks (DCOSS)*, San Francisco, USA.

Chaves, F., Mossgraber, J., Schenk, M., and Bügel, U. (2013). Semantic Registries for Heterogeneous Sensor Networks: Bridging the Semantic Gap for Collaborative Crises Management. In *Database and Expert Systems Applications (DEXA), 2013 24th International Workshop on*, pages 118–122. Institute of Electrical & Electronics Engineers (IEEE).

Chen, Y., Martin, P., Schentz, H., Magagna, B., Zhao, Z., Hardisty, A., Preece, A., and Atkinson, M. (2013). D3.4 ENVRI Reference Model V1.1. Technical Report RDTI-RI-283465, Cardiff University.

Compton, M., Barnaghi, P., Bermudez, L., García-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., Huang, V., Janowicz, K., Kelsey, W. D., Phuoc, D. L., Lefort, L., Leggieri, M., Neuhaus, H., Nikolov, A., Page, K., Passant, A., Sheth, A., and Taylor, K. (2012). The SSN ontology of the W3C semantic sensor network incubator group. *Web Semantics: Science, Services and Agents on the World Wide Web*, 17(0):25–32.

Compton, M., Henson, C., Lefort, L., Neuhaus, H., and Sheth, A. (2009). A Survey of the Semantic Specification of Sensors. In Taylor, K., Ayyagari, A., and Roure, D. D., editors, *Proceedings of the 2nd International Workshop on Semantic Sensor Networks*, volume 522, pages 17–32, Washington DC, USA. CEUR-WS.

Cox, S. (2011). Observations and Measurements - XML Implementation. OGC Implementation OGC 10-025r1, Open Geospatial Consortium Inc.

Cox, S. and Little, C. (2016). Time Ontology in OWL. Working draft, OGC and W3C.

Cyganiak, R., Reynolds, D., and Tennison, J. (2014a). The RDF Data Cube Vocabulary. Recommendation, W3C.

Cyganiak, R., Wood, D., and Lanthaler, M. (2014b). RDF 1.1 Concepts and Abstract Syntax. Recommendation, W3C.

de Mello, G., Graef, F., Stocker, M., Schindler, U., Dasler, R., McEntyre, J., and Dallmeier-Tiessen, S. (2016). Demonstration of Services to Integrate ORCIDs into Data Records and Database Systems. Technical report, Zenodo.

Decker, S., Melnik, S., van Harmelen, F., Fensel, D., Klein, M., Broekstra, J., Erdmann, M., and Horrocks, I. (2000). The Semantic Web: The Roles of XML and RDF. *Internet Computing, IEEE*, 4(5):63–73.

Della Valle, E., Ceri, S., van Harmelen, F., and Fensel, D. (2009). It's a Streaming World! Reasoning upon Rapidly Changing Information. *Intelligent Systems, IEEE*, 24(6):83–89.

Diepenbroek, M., Grobe, H., Reinke, M., Schindler, U., Schlitzer, R., Sieger, R., and Wefer, G. (2002). PANGAEAan information system for environmental sciences. *Computers & Geosciences*, 28(10):1201–1210. Shareware and freeware in the Geosciences II. A special issue in honour of John Butler.

Dow, A. K., Dow, E. M., Fitzsimmons, T. D., and Materise, M. M. (2015). Harnessing the Environmental Data Flood: A Comparative Analysis of Hydrologic, Oceanographic, and Meteorological Informatics Platforms. *Bull. Amer. Meteor. Soc.*, 96(5):725–736.

Egenhofer, M. J. (2002). Toward the Semantic Geospatial Web. In *Proceedings of the 10th ACM International Symposium on Advances in Geographic Information Systems*, GIS '02, pages 1–4, New York, NY, USA. ACM.

Fernández, J. D., Llaves, A., and Corcho, O. (2014). Efficient RDF Interchange (ERI) Format for RDF Data Streams. In Mika, P., Tudorache, T., Bernstein, A., Welty, C., Knoblock, C., Vrandečić, D., Groth, P., Noy, N., Janowicz, K., and Goble, C., editors, *The Semantic Web – ISWC 2014*, volume 8797 of *Lecture Notes in Computer Science*, pages 244–259. Springer International Publishing.

Genesereth, M. R. and Nilsson, N. J. (1987). *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Gil, Y. and Miles, S. (2013). PROV Model Primer. Working Group Note, W3C.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220.

Guarino, N. and Giaretta, P. (1995). Ontologies and Knowledge Bases: Towards a Terminological Clarification. In Mars, N., editor, *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*, pages 25–32, Amsterdam. IOS Press.

Guarino, N., Oberle, D., and Staab, S. (2009). What Is an Ontology? In Staab, S. and Studer, R., editors, *Handbook on Ontologies*, International Handbooks on Information Systems, pages 1–17. Springer Berlin Heidelberg.

Hakala, J. (2010). Persistent identifiers – an overview. Technical report, The National Library of Finland.

Harris, S. and Seaborne, A. (2013). SPARQL 1.1 Query Language. Recommendation, W3C.

Heflin, J. and Hendler, J. (2001). A Portrait of the Semantic Web in Action. *Intelligent Systems, IEEE*, 16(2):54–59.

Hendler, J. (2001). Agents and the Semantic Web. *Intelligent Systems, IEEE*, 16(2):30–37.

Henson, C. A., Pschorr, J. K., Sheth, A. P., and Thirunarayan, K. (2009). SemSOS: Semantic Sensor Observation Service. In *Proceedings of the 2009 International Symposium on Collaborative Technologies and Systems (CTS 2009)*, Baltimore, MD.

Hill, D. J., Liu, Y., Marini, L., Kooper, R., Rodriguez, A., Futrelle, J., Minsker, B. S., Myers, J., and McLaren, T. (2011). A virtual sensor system for user-generated, real-time environmental data products. *Environmental Modelling & Software*, 26(12):1710–1724.

Hobbs, J. R. and Pan, F. (2006). Time Ontology in OWL. Working Draft, W3C.

Hodgson, R., Keller, P. J., Hodges, J., and Spivak, J. (2014). QUDT – Quantities, Units, Dimensions and Data Types Ontologies. Technical report, TopQuadrant, Inc.

Horsburgh, J. S., Tarboton, D. G., Maidment, D. R., and Zaslavsky, I. (2011). Components of an environmental observatory information system. *Computers & Geosciences*, 37(2):207–218.

Huang, V. and Javed, M. (2008). Semantic Sensor Information Description and Processing. In *Sensor Technologies and Applications, 2008. SENSORCOMM '08. Second International Conference on*, pages 456–461.

Klyne, G. and Carroll, J. J. (2004). Resource Description Framework (RDF): Concepts and Abstract Syntax. Recommendation, W3C.

Lassila, O. and Swick, R. R. (1999). Resource Description Framework (RDF) Model and Syntax Specification. Recommendation, W3C.

Le-Phuoc, D., Quoc, H. N. M., Parreira, J. X., and Hauswirth, M. (2011). The Linked Sensor Middleware—Connecting the real world and the Semantic Web. In *Proceedings of the Semantic Web Challenge at the 10th International Semantic Web Conference*, Bonn, Germany.

Lefort, L., Bobruk, J., Haller, A., Taylor, K., and Woolf, A. (2012). A Linked Sensor Data Cube for a 100 Year Homogenised Daily Temperature Dataset. In Henson, C., Taylor, K., and Corcho, O., editors, *Proceedings of the 5th International Workshop on Semantic Sensor Networks*, volume 904, pages 1–16, Boston, Massachusetts. CEUR-WS.

Lewis, M., Cameron, D., Xie, S., and Arpinar, I. B. (2006). ES3N: A Semantic Approach to Data Management in Sensor Networks. In *Proceedings of the 1th International Workshop on Semantic Sensor Networks 2006 (SSN 2006).*

Lindstrom, E., Gunn, J., Fischer, A., McCurdy, A., and Glover, L. (2012). A Framework for Ocean Observing. Task Team for an Integrated Framework for Sustained Ocean Observing IOC/INF-1284, UNESCO.

Madden, S. R., Franklin, M. J., Hellerstein, J. M., and Hong, W. (2005). TinyDB: an acquisitional query processing system for sensor networks. *ACM Trans. Database Syst.*, 30:122–173.

Marascu, A., Pompey, P., Bouillet, E., Wurst, M., Verscheure, O., Grund, M., and Cudre-Mauroux, P. (2014). TRISTAN: Real-time analytics on massive time series using sparse dictionary compression. In *Big Data (Big Data), 2014 IEEE International Conference on*, pages 291–300.

Margara, A., Urbani, J., van Harmelen, F., and Bal, H. (2014). Streaming the Web: Reasoning over Dynamic Data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 25:24–44.

Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A., and Schneider, L. (2002). WonderWeb Deliverable D17: The WonderWeb Library of Foundational Ontologies.

Moraru, A. and Mladenić, D. (2012). A Framework for Semantic Enrichment of Sensor Data. *Journal of Computing and Information Technology*, 20(3):167–173.

Motik, B., Patel-Schneider, P. F., and Parsia, B. (2012). OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax (Second Edition). Recommendation, W3C.

Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T., and Swartout, W. R. (1991). Enabling Technology for Knowledge Sharing. *AI Magazine*, 12(3):36–56.

Peroni, S., Shotton, D., and Vitali, F. (2012). The Live OWL Documentation Environment: A Tool for the Automatic Generation of Ontology Documentation. In ten Teije, A., Völker, J., Handschuh, S., Stuckenschmidt, H., d'Acquin, M., Nikolov, A., Aussenac-Gilles, N., and Hernandez, N., editors, *Knowledge Engineering and Knowledge Management: 18th International Conference, EKAW 2012*, pages 398–412, Galway City, Ireland. Springer Berlin Heidelberg.

Perry, M. and Herring, J. (2012). OGC GeoSPARQL - A Geographic Query Language for RDF Data. Technical Report OGC 11-052r4, Open Geospatial Consortium Inc.

Probst, F. (2006). Ontological Analysis of Observations and Measurements. In Raubal, M., Miller, H. J., Frank, A. U., and Goodchild, M. F., editors, *Geographic Information*

Science, volume 4197 of *Lecture Notes in Computer Science*, pages 304–320. Springer Berlin Heidelberg.

Prud'hommeaux, E. and Carothers, G. (2014). RDF 1.1 Turtle: Terse RDF Triple Language. Recommendation, W3C.

Prud'hommeaux, E. and Seaborne, A. (2008). SPARQL Query Language for RDF. Recommendation, W3C.

Randell, D., Cui, Z., and Cohn, A. (1992). A Spatial Logic based on Regions and Connections. In Nebel, B., Rich, C., and Swartout, W., editors, *Principles of Knowledge Representation and Reasoning*, pages 165–176, San Mateo, CA, USA. Morgan Kaufmann.

Rönkkö, M., Kauhanen, O., Stocker, M., Hytönen, H., Kotovirta, V., Juuso, E., and Kolehmainen, M. (2015). *Quality Control of Environmental Measurement Data with Quality Flagging*, pages 343–350. Springer International Publishing, Cham.

Sheth, A., Henson, C., and Sahoo, S. (2008). Semantic Sensor Web. *Internet Computing, IEEE*, 12(4):78–83.

Stocker, M., Shurpali, N., Taylor, K., Burba, G., Rönkkö, M., and Kolehmainen, M. (2015). Emrooz: A Scalable Database for SSN Observations. In Kyzirakos, K., Henson, C., Perry, M., Varanka, D., and Grütter, R., editors, *Joint Proceedings of the 1st Joint International Workshop on Semantic Sensor Networks and Terra Cognita (SSN-TC 2015) and the 4th International Workshop on Ordering and Reasoning (OrdRing 2015) co-located with the 14th International Semantic Web Conference (ISWC 2015)*, volume 1488, pages 1–12, Bethlehem, PA, USA. CEUR-WS.

Studer, R., Benjamins, V., and Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, 25(1–2):161–197.

Swartout, B., Patil, R., Knight, K., and Russ, T. (1996). Toward Distributed Use of Large-Scale Ontologies. In *Procedings of the 10th Knowledge Acquisition for Knowledge-Based Systems Workshop*, Banff, Canada.

Swartout, W. and Tate, A. (1999). Ontologies. *Intelligent Systems and their Applications, IEEE*, 14(1):18–19.

Taylor, P. (2014). OGC WaterML 2.0: Part 1 – Timeseries. OGC Implementation Standard - Corrigendum OGC 10-126r4, Open Geospatial Consortium Inc.

Wang, P., Zheng, J. G., Fu, L., Patton, E. W., Lebo, T., Ding, L., Liu, Q., Luciano, J. S., and McGuinness, D. L. (2011). A Semantic Portal for Next Generation Monitoring Systems. In Aroyo, L., Welty, C., Alani, H., Taylor, J., Bernstein, A., Kagal, L., Noy, N., and Blomqvist, E., editors, *The Semantic Web – ISWC 2011*, volume 7032 of *Lecture Notes in Computer Science*, pages 253–268. Springer Berlin Heidelberg.

Wang, W., Barnaghi, P., Cassar, G., Ganz, F., and Navaratnam, P. (2012). Semantic Sensor Service Networks. In *Sensors, 2012 IEEE*, pages 1–4. Institute of Electrical & Electronics Engineers (IEEE).

Wei, W. and Barnaghi, P. (2009). Semantic Annotation and Reasoning for Sensor Data. In Barnaghi, P., Moessner, K., Presser, M., and Meissner, S., editors, *Smart Sensing and Context*, volume 5741 of *Lecture Notes in Computer Science*, pages 66–76. Springer Berlin Heidelberg.

Wetz, P., Trinh, T.-D., Do, B.-L., Anjomshoaa, A., Kiesling, E., and Tjoa, A. M. (2014). Towards an Environmental Information System for Semantic Stream Data. In *Proceedings of the 28th International Conference on Informatics for Environmental Protection (EnviroInfo 2014)*, pages 637–644, Oldenburg, Germany. BIS-Verlag.

Wikipedia (2016). Acoustic Doppler current profiler.

Yao, Y. and Gehrke, J. (2002). The Cougar Approach to In-network Query Processing in Sensor Networks. *SIGMOD Rec.*, 31(3):9–18.