Building the Semantic Web for Earth Observations

Martina Baučić
Faculty of civil engineering, architecture and geodesy
University of Split
Split, Croatia
martina.baucic@gradst.hr

Damir Medak
Faculty of geodesy
University of Zagreb
Zagreb, Croatia
damir.medak@geof.hr

***Abstract*—The growth in number of Earth sensors and increase in data volumes have raised a problem of observations integration, data analysis and reasoning over the integrated data. The two initiatives are building an interoperable environment for Earth observations: the Sensor Web Enablement and the Semantic Sensor Web. The standards for web services and observation encodings are resolving syntactic interoperability between sensors. The Semantic Web standards are enriching observations with description of data semantics and thus improving data integration. The paper demonstrates the building of Semantic Web for Earth observation data. It explains development of meteorological data ontology and provides an example of transforming meteorological data into Resource Description Framework data model. Although at the very beginning, the current implementations have proved the Semantic Web as an emerging technology for Earth observations integration and web computational modelling.**

Keywords—Earth observations; Sensor Web; Semantic Web; meteorology; ontology; RDF data model

I. INTRODUCTION

Key emerging trends in Earth observations include the growing number of sensors, growth in data volume, real-time processing, distribution via Web, crowdsourcing etc. There is a need for integration of Earth observation data coming from various sensors that will allow analysis and reasoning over integrated data. These trends can be found in reports such as the Report on future trends in geospatial information management by UN Committee of Experts on Global Geospatial Information Management [1].

The work presented here discusses current efforts in building the Semantic Web as an interoperable environment for Earth observations. The Open Geospatial Consortium (OGC) initiative called Sensor Web Enablement (SWE) has defined open standards for exploiting Web connected sensors. The standards include encodings for describing sensors and sensor observations, and interface definitions for web services. The syntactic interoperability is achieved by adoption of these standards, but semantics of observations remain ambiguous. The Semantic Sensor Web initiative by World Wide Web Consortium (W3C) extends SWE standards with spatial, temporal, and thematic description of observations by ontologies. There are three main reference ontologies for building Earth observations ontologies. W3C Semantic Sensor Network (SSN) ontology models sensor devices, systems and processes; W3C Time ontology models temporal concepts such as instants, intervals, durations etc.; OGC GeoSPARQL standard models geospatial objects and their topological and geometrical properties. To build domain ontology, such as ontology for meteorological data, one should define basic concepts in the domain and relations among them. To enable integration of data from various domains, domain ontology concepts should be linked to concepts in reference upper ontologies.

Implementation of the Semantic Web technologies for Earth observations is at the very beginning. There are projects such as European research project TELEIOS that builds Virtual Earth Observatories [2], or National Aeronautics and Space Administration project that builds Semantic Web for computational modeling of the impacts of changing climate [3]. Looking at scope of the projects and organizations that implement them, the Semantic Web will be emerging technology not only in integration of Earth observations but also in web computational modelling of geospatial and temporal data.

The remainder of this paper is organized as follows. Section 2 briefly describes the OGC Sensor Web Enablement initiative. Section 3 provides main information about the Semantic Sensor Web initiative by W3C. Section 4 explains the development of meteorological data ontology. It also provides an example of transforming meteorological data into Resource Description Framework (RDF) data model. Finally, we present the conclusions.

II. SENSOR WEB ENABLEMENT INITIATIVE

There are millions of sensors, in and around the Earth, collecting massive amounts of data. Sensors could be from a barometer at fixed location to hyper-spectral sensor on board of a satellite. Each sensor observes a certain condition (wind, pressure, etc.) in a particular place and time. This spatio-temporal information is stored on the sensor or directly sent to server, but having its own data format and software for processing, and its own semantics. Overwhelming number of observations must be processed and explained, and thus we need interoperability between the heterogeneous sensor data and applications.

The OGC SWE initiative is developing the global standards to enable discovery, exchange, processing of observations and controlling of sensor systems via the Web. The standards include encodings for describing sensors and observations, and interface definitions for Web services. The built and prototyped SWE standards include the following [4]:

- Observations & Measurements Schema
- Observations and Measurements XML encoding
- Sensor Model Language
- Sensor Observations Service
- Sensor Planning Service
- SWE Common Data Model
- SWE Services Common
- PUCK Protocol Standard
- Sensor Alert Service
- Web Notification Services

The SWE enables interoperability between sensors, models and decision support systems as shown on Fig.1. It is a middleware layer that provides description and discovery of sensor assets and capabilities, access to data, tasking of sensors, and subscription to alerts. The goal of SWE is a distributed sensing system in which information is globally shared and used by all networked clients. Some current SWE implementation efforts are listed in [4]. We will mention the organization 52North that provides a complete set of SWE services under General Public License.

However, while the syntactic interoperability is achieved by adoption of the SWE standards, the semantics of observations remain ambiguous. Also, the SWE standards do not provide a basis for reasoning that can ease development of advanced applications for discovery and retrieval of sensor data.

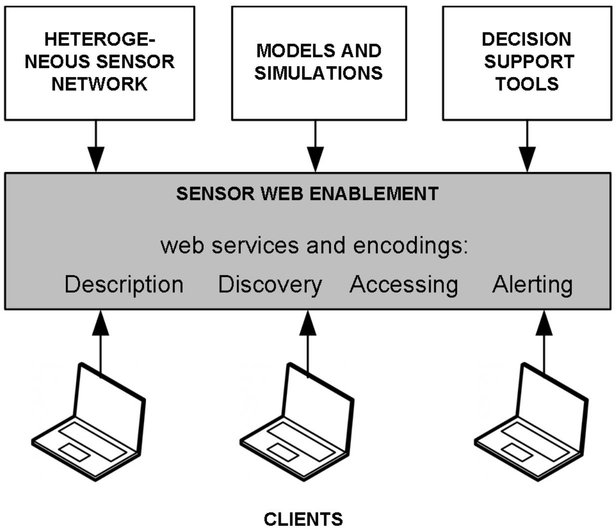

Fig. 1. SWE framework

III. Semantic Sensor Web initiative

The Semantic Web is an extension of the Web facilitating users to find, share, and combine information more easily. It is a vision of "Web of data" that can be readily interpreted by machines, instead of today "Web of documents" that can be read by people. Semantics, or meaning, of information on the Web is formally defined by ontologies. The Semantic Web stack builds on the W3C standards: Resource Description Framework (RDF), SPARQL Protocol and RDF Query Language (SPARQL), Ontology Web Language (OWL), Extensible Markup Language (XML), and Uniform Resource Identifier (URI). These technologies provide machine-readable descriptions of the content of Web documents and reasoning algorithms for automated information search.

To improve semantic interoperability and integration of sensor data, the SWE initiative is enriched with Semantic Web technologies. The Semantic Sensor Web initiative by W3C extends SWE standards with spatial, temporal, and thematic description of observations by ontologies. These ontologies allow integration, classification and reasoning over the sensors data and observations.

The Semantic Sensor Networks Community Group is developing Semantic Sensor Network (SSN) ontology which models sensor devices, systems, processes, and observations [5]. The SSN ontology is domain independent and it merges sensor-focused, observation-focused and system-focused views. It is aligned with the DOLCE Ultra Lite (DUL) upper ontology to facilitate reuse and interoperability. The SNN is a formal OWL Description Logic ontology available as single OWL file [6]. It consists of 41 classes and 39 properties. Fig. 2 shows a small part of SSN ontology with a central concept: *Sensor*, as the broadest concept of any entity capable of sensing. The nine SNN classes (shown on Fig. 2 as white ovals) are connected by properties. The property *subClassOf* (shown on Fig. 2 as arrow with no filled head) means: e.g. any member of *Device* class is a member of *System* class, and of *Physical object* class. The arrows with filled heads show various properties, but their names are omitted from Fig. 2 due to figure size limits. E.g. *observesOnly* is the property linking *Sensor* with *Property*. Some classes are linked with the upper ontology classes of DUL ontology (shown on Fig. 2 as grey ovals). E.g. *Property* class is subclass of *Quality* class.

There are concepts not described with the SSN: e.g. units of measurements, locations, features and property hierarchies. The idea is that knowledge engineers of particular domain include domain feature ontology, location and units ontology by linking them to SSN ontology. E.g. SSN ontology is combined with NASA SWEET (Semantic Web for Earth and Environmental Terminology) ontology modelling the Earth observed properties.

Although recently published, the SNN ontology is already being used in several projects. The examples and uses of the SSN ontology are given in [7]. Some of them are: SENSEI and SPITFIRE projects in the EU's Seventh Framework Programme; the projects of the Kno.e.sis Centre at the Wright State University; the projects of the 52North organization and the SemsorGrid4Env project. Linked Sensor Data and Linked Observation Data are projects of the Kno.e.sis Centre. The projects RDF datasets contain description of circa 20.000 weather stations and hurricane observations in the USA since 2002. The datasets are part of the Linked Open Data. These projects have shown that the use of SNN ontology is enabling integration of sensor data with other data and applications relying on Semantic Web technologies like RDF and SPARQL.

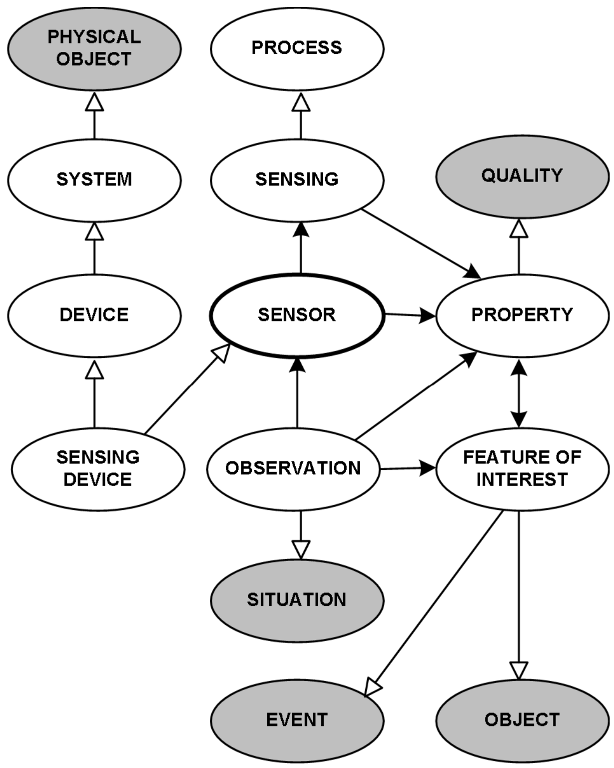

Fig. 2. Part of SSN ontology aligned with the DOLCE Ultra Lite ontology classes (colored in grey)

IV. DEVELOPMENT OF METEOROLOGICAL DATA ONTOLOGY AND RDF DATABASE

Ontology represents knowledge of a domain as a hierarchy of concepts (also called classes), their properties (also called attributes) and relationships. Ontology languages are used to construct ontologies. The current W3C standards are: OWL, a formal language based on description logics; RDF; and RDF Schema. Domain ontology represents concepts of a particular domain. Upper ontology represents concepts applicable across a range of domain ontologies (e.g. SUMO or DOLCE ontology). To enable integration of data from various domains, the domain ontology concepts should be linked to concepts in the reference upper ontologies. In addition to taxonomic hierarchies of classes and properties, the ontology can state axioms constraining the possible interpretations and describe the logical inferences that can be drawn from asserted data.

Several methodologies are guiding experts in the process of ontology building, but there are two main steps (Fig.3). In first step, an expert models the domain knowledge: define basic concepts and relations among them, and define axioms and rules for data interpretation and reasoning. The second step is to link the domain concepts with concepts in reference upper ontologies. One should consider the reuse of the already developed ontological resources.

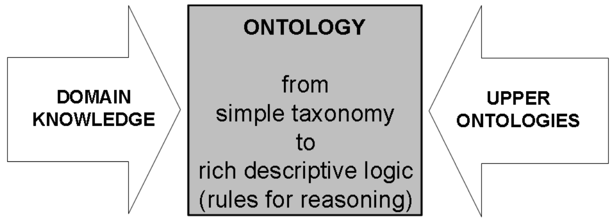

Fig. 3. Ontology building process

The Croatian Meteorological and Hydrological Service is publishing meteorological data in XML files. Fig. 4 shows an excerpt from the XML file. Each file contains 8 meteorological observations from 38 weather stations for a particular date and hour.

```
  <?xml version="1.0" encoding="UTF-8"?>
- <Hrvatska>
  - <DatumTermin>
      <Datum>24.05.2014</Datum>
      <Termin>18</Termin>
    </DatumTermin>
  + <Grad>
  - <Grad>
      <GradIme>Crikvenica</GradIme>
    - <Podatci>
        <Temp>22</Temp>
        <Vlaga>68</Vlaga>
        <Tlak>1014.9</Tlak>
        <TlakTend>-</TlakTend>
        <VjetarSmjer>E</VjetarSmjer>
        <VjetarBrzina>04</VjetarBrzina>
        <Vrijeme>-</Vrijeme>
        <VrijemeZnak>-</VrijemeZnak>
      </Podatci>
    </Grad>
  - <Grad>
      <GradIme>Gorice-Nova Gradiška</GradIme>
    - <Podatci>
        <Temp>25</Temp>
```

Fig. 4. Excerpt from XML with meteorological data

Our attempt aims to facilitate the use of meteorological data by adding semantic description and offering as RDF data.

We started with by searching existing ontological resources. We have considered the W3C SSN ontology, W3C Time ontology and OGC GeoSPARQL ontology as the reference ontologies. Fig. 5 shows the links between the main concepts in the three reference ontologies. *Observation* is a subclass of *Temporal entity*, and thus it has its beginning and end. *Observation* and *Feature of Interest* are subclasses of *Feature*, and thus they have their geometries. By linking SSN ontology to GeoSPARQL ontology, the sensor concepts may have complex descriptions of their geospatial characteristics such as types of geometry, coordinate reference systems and topological relations. The Geography Markup Language

(GML) and well-known text (WKT) standards are used for geospatial data encoding.

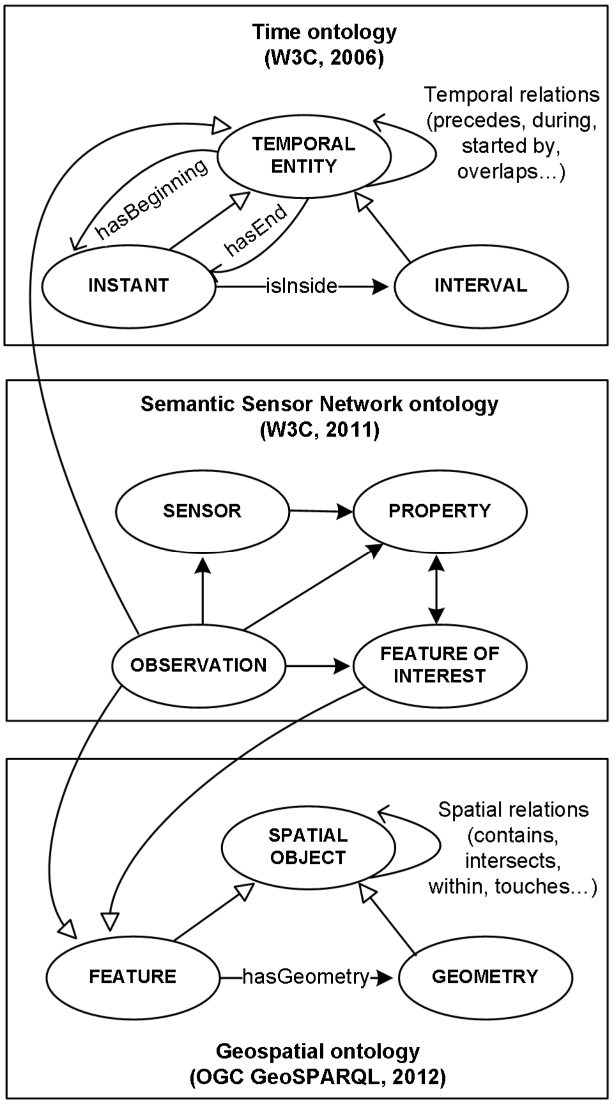

Fig. 5. Links between the main concepts of the three reference ontologies

By extending W3C SSN ontology, we have defined basic concepts and their relationships for the meteorological observations stored in XML file. Fig. 6 shows some meteorological classes, their relationships and links to W3C SSN, OGC GeoSPARQL and W3C Time ontology.

New defined meteorological classes and their relationships can be encoded in a TBox part of knowledge base. The TBox contains ontological schema describing terminology and data semantics. The definition of new class *TemperatureObservation* and its relationship with *Observation* class is written in OWL language with Turtle RDF serialization as follows:

dhmz:TemperatureSensorOutput rdf:type owl:Class.

dhmz:TemperatureObservation rdf:subclass ssn:Observation.

The prefixes *dhmz*, *rdf*, *owl* and *ssn* in the above statements are URI abbreviations (e.g. *rdf* is abbreviation of *www.w3.org/1999/02/22-rdf-syntax-ns#*). A URI provides a global identification for a Web resource.

A key feature of OWL is its ability to describe class by restricting the values allowed for certain properties. It allows us to make inferences about members of a class. The description of class *TemperatureSensorOutput* by restriction is as follows:

dhmz:TemperatureSensorOutput rdf:type owl:Class;

rdfs.subClassOf

[rdf:type owl:Restriction;

owl:onProperty ssn:hasValue;

owl:allValuesFrom dhmz:TemperatureValue].

The above statements will classify all instances as members of class *TemperatureSensorOutput* for which all values of the property *hasValue* come from class *TemperatureValue*.

Having classes and properties written in the TBox, we can encode meteorological observations in an Abox part of knowledge base. The ABox contains asserted instances. For example, air temperature of 22^{o}C, measured at the weather station Crikvenica at 18 o'clock on May 24, 2014. This observation is encoded as follows:

TemperatureObservation_Cr_2405201418 rdf:type dhmz:TemperatureObservation;

dhmz:observationResult dhmz:TemperatureSensorOutput_1234;

geo:hasGeometry dhmz:geo_WS_Crikvenica.

dhmz:TemperatureSensorOutput_1234 rdf:type dhmz:TemperatureSensorOutput;

ssn:hasValue dhmz:TemperatureValue_1;

dhmz:hasTime dhmz:TemperatureDateTime_1.

dhmz:TemperatureValue_1 ssn:hasQuantityValue "22"^^qudt:DegreeCelsius.

dhmz:TemperatureDateTime_1 time:inXSDDateTime "2014-05-24T18:00:00"^^xsd:dateTime.

In order to add geospatial location to the above observation, the weather station Crikvenica is defined as a point with coordinates and a coordinate reference system:

dhmz:geo_WS_Crikvenica rdf:type geo:Point;

geo:asWKT "<http://www.opengis.net/def/crs/EPSG/0/3765> POINT(35787.4 5005291.0)"^^geo:wktLiteral.

The TBox and ABox use the same RDF data model and the same OWL encoding language. The data and their description (semantics) are stored together and can be queried together by SPARQL. Moreover, sensors data from other sources can be converted to RDF, merged into one federated RDF database, and queried together with their semantics.

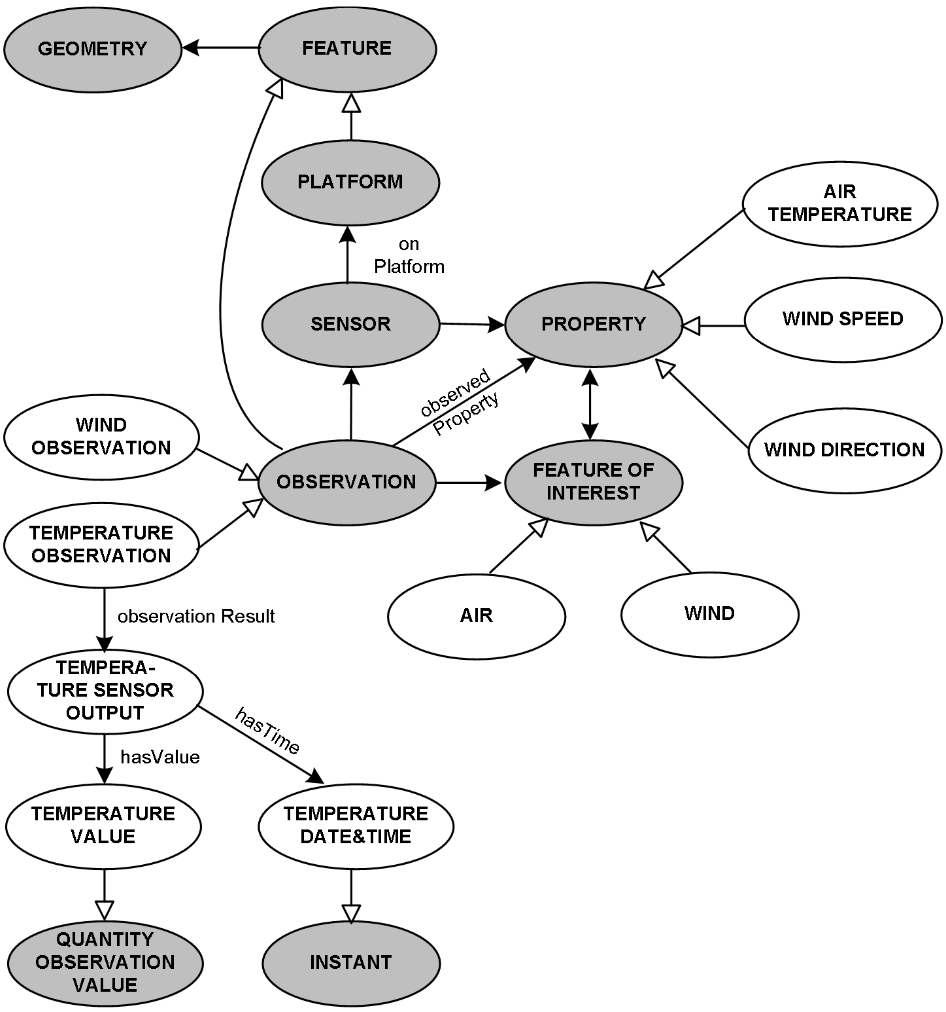

Fig. 6. Some meteorological classes (white ovals), their relationships and links with SSN, GeoSPARQL and Time ontology classes (grey ovals)

An example of SPARQL query over federated RDF database is presented bellow. The query shows which weather stations (labeled with *?ws*) are within which national parks (labeled with *?np*).

SELECT ?ws ?np

WHERE {

?ws rdf:type dhmz:Wather_station;

geo:hasGeometry ?geo_ws.

?np rdf:type hrnp:National_park;

geo:hasGeometry ?geo_np.

?geo_ws geo:sfWithin ?geo_np.

}

In the previous example, the federated RDF database is merged from two imaginary RDF databases: *dhmz* (could be the RDF database of The Croatian Meteorological and Hydrological Service) and *hrnp* (could be the RDF database of Croatian Registry of National Protection Areas). GeoSPARQL property *sfWithin* defines topological relations between weather stations and national parks. The example clearly demonstrates the power of RDF data model in integrating data which could be used for the integration of Earth observations.

V. CONCLUSIONS

Our attempt was to explain and demonstrate the building of Semantic Web for Earth observation data. The new technologies are emerging and able to integrate, process, and explain overwhelming number of observations. The current efforts encompass two initiatives: the OGC Sensor Web Enablement and the W3C Semantic Sensor Web. Recently developed standards are already successfully implemented throughout many projects and it seems the Semantic Web technologies will take a significant place in integration of sensor data and applications.

In this paper we have briefly described Semantic Web concepts and we have demonstrated a domain ontology development combining thematic, spatial and temporal ontologies. The meteorological data available in XML files published by the Croatian Meteorological and Hydrological Service is converted into RDF data model. Enriched with semantics, the meteorological data can effectively be used with data from other sources. The example of SPARQL query demonstrates the integration of data from two RDF databases and use of the GeoSPARQL topological relation property *stWithin*.

Instead of commonly used W3C Basic Geo Vocabulary standard which can only describe points with latitude, longitude, and altitude in the WGS84 coordinate reference system, we have used GeoSPARQL standard which provides complete semantics of geospatial objects and their spatial relations.

In our future work we will explore qualitative spatial reasoning over the integrated Earth observation data by building more complex OWL models.

REFERENCES

[1] UN Committee of Experts on Global Geospatial Information Management, Future Trends in Geospatial Information Management: The Five to Ten Year Vision, J. Carpenter and J. Snell, Eds. UN, July 2013.

[2] Project TELEIOS - Virtual Observatory Infrastructure for Earth Observation Data. (*http://www.earthobservatory.eu*)

[3] NASA Earth, Life and Semantic Web (ELSeWeb) project. (*https://earthdata.nasa.gov/our-community/community-data-system-programs/access-projects/earth-life-and-semantic-web-elseweb*)

[4] Open Geospatial Consortium, OGC® Sensor Web Enablement: Overview And High Level Architecture, C. Reed, M. Botts, G. Percivall, and J. Davidson, Eds. 2013.

[5] W3C Incubator Group, Semantic Sensor Network XG Final Report, L. Lefort, C. Henson, and K. Taylor, Eds. 2011.

[6] W3C Incubator Group, Semantic Sensor Network ontology OWL file, 2011. (*www.w3.org/2005/Incubator/ssn/ssnx/ssn.owl*)

[7] M. Compton, P. Barnaghi, L. Bermudez, et al., "The SSN ontology of the W3C semantic sensor network incubator group" in Journal of Web Semantics, vol. 17. Elsevier, 2012, pp 25-32.