Sense and Sens'ability: Semantic Data Modelling for Sensor Networks

Payam Barnaghi, Stefan Meissner, Mirko Presser, and Klaus Moessner
Centre for Communication Systems Research (CCSR)
University of Surrey, GU2 7XH, United Kingdom
{p.barnaghi,s.meissner,m.presser,k.moessner}@surrey.ac.uk
http://www.ee.surrey.ac.uk/ccsr

Abstract: Sensor networks are used in various applications in several domains for measuring and determining physical phenomena and natural events. Sensors enable machines to capture and observe characteristics of physical objects and features of natural incidents. Sensor networks generate immense amount of data which requires advanced analytical processing and interpretation by machines. Most of the current efforts on sensor networks are focused on networking and service development for various applications, but less on processing the emerging data. Sensor data in a real world application will be an integration of various data obtained from different sensors such as temperature, pressure, and humidity. Processing and interpretation of huge amounts of heterogeneous sensor data and utilising a coherent structure for this data is an important aspect of a scalable and interoperable sensor network architecture. This paper describes a semantic model for heterogeneous sensor data representation. We use common standards and logical description frameworks proposed by the semantic Web community to create a sensor data description model. The work describes a sensor data ontology which is created according to the Sensor Web Enablement (SWE) and SensorML data component models. We describe how the semantic relationship and operational constraints are deployed in a uniform structure to describe the heterogeneous sensor data.

Keywords: Sensor Networks, Ontologies, Knowledge Modelling, SensorML

1. Introduction

The current Web is a document centric platform for exchanging data amongst the users. The Internet and Web applications in recent years have seen tremendous growth in facilitating data exchange for different applications and purposes. The current networks, however, are limited in sensing and measuring the physical world phenomena and employing them for observing and controlling real world incidents. Sensor networks provide a potential for Internet applications to acquire contextual data and observe and measure physical events. This will lead to construct a platform for the future Internet which is aware of physical world incidents and enables new service types that remove the strict boundary between virtual and physical world. To achieve this, data collected from different types and levels of sensors and sensor networks will be used in future applications. Machines will need to collect and understand the data provided by the various types of sensors and networks. Enabling machine interpretability and reasoning requires a common approach to organise and structure the data. This paper provides an ontology based approach to the structuring of data obtained from different types of sensors.

The next section describes background studies and the semantic Web technologies. Section 3 discusses some of the current efforts on applying these technologies to sensor networks. In section 4, we describe a novel approach based on the semantic Web technologies and using a universal language to provide semantic data modelling for sensor networks. Section 5 provides an evaluation of the work and section 6 concludes the paper and discusses the future work.

2. Background

The Open Geospatial Consortium (OGC)[1] has recently established a group which is called Sensor Web Enablement (SWE). This group is responsible to specify interoperability interfaces and meta-data encodings for integration of heterogeneous sensor data [1]. The main specifications defined by the group are described in the following.

- *Observations & Measurements (O&M)* which define standard models and XML Schema for encoding real-time and archived observations and measurements of sensor data.

- *Sensor Model Language (SensorML)* is a standard model to describe sensor systems and processes associated with sensor observations in an XML-based structure. The information provided by SensorML can be used for sensor discovery, describing sensor data, and specifying sensor observations.

- *Transducer Model Language (TransducerML or TML)* provides a conceptual model to describe transducers and to support real-time data to and from sensor systems, sensors and actuators.

- *Sensor Observations Service (SOS)* is a standard Web service interface for requesting, filtering, and retrieving observations and sensor system information.

- *Sensor Planning Service (SPS)* is a standard Web service interface that acts as an intermediary between a client and a sensor collection management environment.

- *Sensor Alert Service (SAS)* is another standard Web service interface that enables publishing and subscribing to alerts from sensors.

- *Web Notification Services (WNS)* enables asynchronous delivery of messages or alerts from SAS and SPS Web services and other elements of service workflows.

The models and interfaces provided by SWE define a standard framework to deal with sensor data in heterogeneous sensor network applications. SensorML provides an extensive description model for various attributes of sensor data [2]. Its primary representation is defined in XML schema form. Although XML provides a remarkable solution for heterogeneous data representation, there are significant limitations in semantic interoperability and describing the semantics and relationships between different data element using XML representations [3].

---

[1] http://www.opengeospatial.org/

2.1 — The Semantic Web Technologies

Semantic Web is an extension to the current Web in which the meaningful relationships between resources is represented in machine processable formats [4]. The main idea in the semantic Web is to provide well defined and machine accessible representation of the resources and their relationships rather than simple links as they are offered by the link structure on the current Web (i.e. *href* links in HTML). Ontologies are utilised by the semantic Web applications to offer conceptualised representation of domains and to specify meaningful relationships between the resources. Ontologies provide common and shared understandings of different domains. The World Wide Web Consortium (W3C)has defined different standards for representing the semantic Web data in machine accessible and processable formats.

The primary technologies for the semantic Web include the Resource Description Framework (RDF)[2], RDF Schema[3], and the Web Ontology Language (OWL)[4]. OWL is based on description logic and facilitates construction of ontolgies for different domains. The OWL data can be accessed by software agents for reasoning and inferencing purposes and to enable systems to derive additional knowledge from the represented data. There are common query languages such as SPARQL[5] available for the OWL data. There are also widely used software systems such as Jena [5] and Sesame [6] to deploy and manage the constructed ontologies. The OWL representation of data enables expression of semantics and meaningful relationships between resources and amongst different attributes of complex data.

2.2 — Sensor Data Modelling

Russomanno *et al* [7] discuss a broad sensor ontology which is called OntoSensor. OntoSensor primarily adapts parts of SensorML descriptions and uses extensions to the IEEE Suggested Upper Merged Ontology (SUMO)[6] to describe sensor information and capabilities. The ontology is developed to support sensor information system applications in dynamic sensor selection, reasoning and querying various types of sensor. OntoSensor relies on deep knowledge models and provides extensive information about different aspects of the sensor nodes and devices. The ontology is represented in OWL format and the authors have discussed the advantages of the proposed approach compared to SensorML and XML based solutions. The main enhancement is providing self-descriptiv meta-data for the transducer elements and embedded semantics in the descriptions which could be utilised in various sensor discovery and reasoning applications. Although OntoSensor illustrates a semantic approach to sensor description and provides an extensive knowledge model, there is no distinctive data description model to facilitate interoperable data representation for sensors observation and measurement data. A universal sensor observation and measurement data model in collaboration with a sensor specification model create a semantic sensor network architecture. The

---

[2] http://www.w3.org/RDF/

[3] http://www.w3.org/TR/rdf-schema/

[4] http://www.w3.org/TR/owl-features/

[5] http://www.w3.org/TR/rdf-sparql-query/

[6] http://www.ontologyportal.org/

semantic sensor network will utilise semantic Web technologies and reasoning mechanisms to interpret sensor data from physical devices that perform observations and measurements. This will support building automated sensor information processing mechanisims to extract additional knowledge from real-time or archived sensor data.

Ontology-based description of a service oriented sensor network is discussed in [8]. The SWE and Geography Markup Language (GML)[7] classes and properties in collaboration with SensorML, Suggested Upper Ontology (SUMO) and OntoSensor are used to develop an ontology for sensor service description. The ontology consists of three main components *ServiceProperty*, *LocationProperty*, and *PhysicalProperty*. ServiceProperty explains what a service does and properties in the other two components describe the contextual and physical characteristics of the sensor nodes in a wireless sensor network architecture. The ontology is represented in OWL form and some initial consistency checking and query results are provided to evaluate the validity of the proposed solution. The system, however, does not specify how complex sensor data will be described and interpreted in a sensor network application. The proposed framework concentrates on building a sensor description ontology for sensor discovery and description of sensor meta-data in a heterogeneous environment. Although sensor device and service description will contribute to build more autonomous sensor networks, providing an interoperable data description model would be also an essential requirement in an architecture for semantically enabled sensor networks.

A high level design for a universal ontology which consists of extension plug-in ontologies, sensor data ontology and sensor hierarchy ontology is described in [9]. The extension plug-in ontologies enable the developers to integrate domain specific ontologies into the main ontology. This describes the sensor network capabilities and provides relations between the domain concepts and the sensor functionalities. The sensor hierarchy ontology is a knowledge model for the sensors and actuators and other physical devices in the network. It describes the features and capabilities of the elements and contains meta-data related to devices such as measurement range, accuracy and calibration. The sensor data ontology describes the dynamic observational data for transducers. The ontology model describes the contextual data with respect to the spatio-temporal attributes. However the illustrated model does not specify the details of sensor data specification and relationships between various types of complex sensor data. The taxonomy provided for the sensor hierarchy ontology specifies a set of primary numerical attributes for common types of sensors. In a practical scenario, sensor data will include more complex data types and there will be a requirement for a universal structure to define the sensor data and emerging semantics.

Seth and Hanson [10] discuss the idea of a semantic sensor Web framework to provide enhanced meanings to sensor data and to create situation awareness for the sensor networks. The semantics of sensor nodes is described within space and time dimensions, and it also includes thematic data. The spatial meta-data provides sensor location and data information in terms of a geographical reference system, location reference, or named locations. The

---

[7] http://www.opengeospatial.org/standards/gml

main assumption is that although the sensor's location might be changing, its location can be determined relative to the moving object. The temporal meta-data refers to the time interval duration whose sensor data has been captured. Thematic meta-data provides descriptive information about the sensor node which can be derived by sensor data analysis, and utilising tagging and textual descriptions [11]. The sensor Web facilitates interoperable architecture for sensor networks and enables the application to process and interpret the contextual, observation and measurement data obtained from a sensor in a heterogeneous environment. The authors describe different scenarios for applying the semantic Web technologies and ontologies to the sensor networks. One of the main issues in the semantic sensor Web architecture is employing a unified data model which supports universal interoperability and semantic description for sensor data. The latter will enable construction of content and context aware sensor network applications.

Henson *et al* [12] describe a prototype application for the sensor Web by using annotated video data. The dataset contains Youtube videos annotated with SensorML and XLINK[8] models with reference to a time ontology. The authors discuss how utilising the semantic leads to retrieve videos by specifying temporal concepts such as "*within*", "*contains*", or "*overlaps*" during a time interval query submission. The proposed application demonstrates the main benefits of adding semantics to the sensor network and sensor data. The authors use keyword tagging and meta-data description to provide references to temporal concepts and domain ontologies. An extension to this idea could be seen as providing a universal meta-data structure with a broaden scope to accommodate various sensor data types.

3. Data Components Modelling

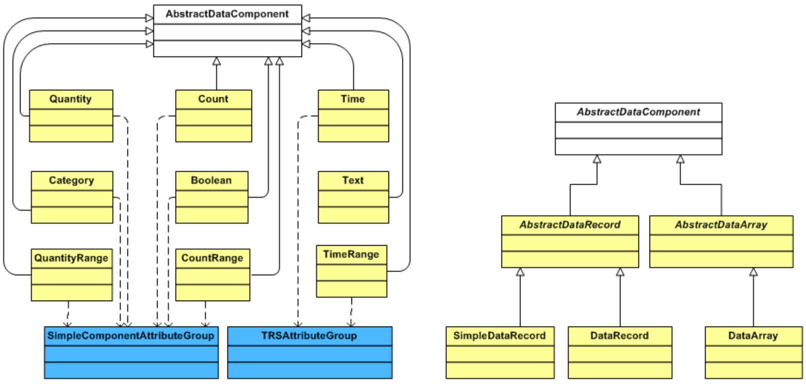
The SWE common namespace defines several value types and data types for sensor measurement and observation data. The data types fall into the following main categories [2].

- Primitive data types, which complement the data types defined in GML.
- General purpose aggregate data types such as records, arrays, vectors, and matrices.
- Aggregate data types with special semantics such as curve, and time aggregates.
- Standard encoding to include semantics, quality indications and constraints to primitive and aggregate types.
- Specialized components to support semantic definitions
- A notation for the description of XML and non-XML array encoding.

The data types are represented in XML encoded form; however it is also possible to use other alternative encodings for the data. The primitive data types describe the scalar values such as *Quantity*, *Count*, *Boolean*, *Category*, and *Time*. These data types provide primitives to define the sensor data. Figure 1(a) shows a model of the simple data types in SWE namespace. A data component describes an object whose values can be defined as a set of simple data types. The simple data types contain properties that describe different attributes required for sensor data. The data types can be grouped together to construct an aggregate object [2].

---

[8] http://www.w3.org/TR/xlink/



(a)

(b)

Figure 1: (a) The simple data types (b) The generic data aggregates

The generic aggregate components are defined as *RecordTypes* and *ArrayTypes*. There are also derived aggregates such as *DataRecord*, S*impleDataRecord*, *DataArray*, *Vector*, *ConditionalValue*, and *Curve*. Figure 1(b) shows a UML model for the generic data aggregation models in SensorML which is based on SWE namespace.

Position data is also an essential part of the sensor data types as many of sensor network applications utilise position data for context and information processing. The method to describe the position information in SWE namespace not only focuses on position information structuring itself, it also describes the dynamic information that are related to location and orientation. The position information in addition includes velocity, acceleration, angular velocity, and angular acceleration in combination with a time tag. Readers may refer to [2] for more information on SensorML data types and SWE common namespace.

Although XML provides a flexible method to represent the data, it does not provide a full potential for the machines to acquire and interpret the emerging semantics from data. The concepts in XML data are represented in the form of broader and narrower concepts and other types of meaningful relationship between data resources are not explicitly defined. Extending the XML descriptions to ontological primitives and using more descriptive representation structures enables advance analysis and enhanced data processing for heterogeneous sensor network applications. The next section describes an ontology-based representation for sensor data. The proposed model includes complex data types according to the SWE namespace definitions. It also describes the meaningful relationships between the data components and the attributes in a OWL representation form.

4. Sensor Data Ontology

Most of the current work on providing semantic data for sensor network is focused on using semantic description for sensor nodes and elements which support advance analytics and situation and context awareness in sensor networks. Creating a semantic sensor data model for sensor data related to measurements and observations is another important aspect in designing highly scalable and advanced heterogeneous sensor network applications. This will also support creating advanced data mining and knowledge extraction methods for real-time or archived sensor data. We propose a framework for a semantic data description model which provides interoperability and facilitates deriving additional knowledge from real-time and/or stored sensor data. The semantic data model in collaboration with a semantic sensor network architecture will support designing smart applications using sensor networks.

We have used Protégé[9] an opensource ontology editor and knowledge acquisition system developed at the University of Stanford. The Protégé editor is used to design the class and property structure of the proposed semantic data model and also to define the constraint rules for the associations and attributes. We have also imported a part of the NASA's SWEET ontology [10] for measurement units. The ontology is serialised in OWL form which can be deployed in common ontology management software such as Sesame and Jena. Figure 2 shows a fragment of our semantic sensor data ontology which is called SensorData Ontology.

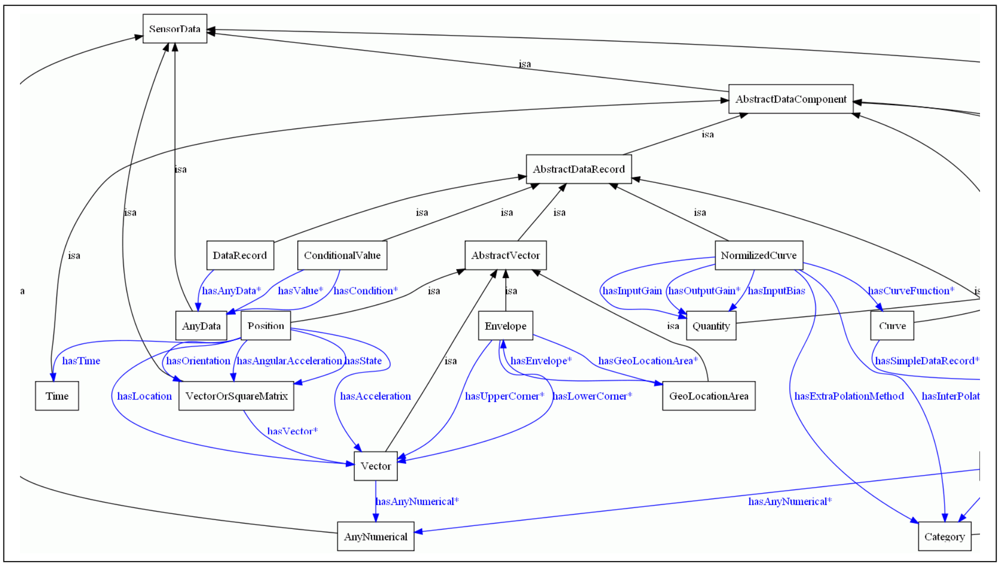


Figure 2: A snippet of the SensorData ontology

At the first look, it seems the represented data in OWL form adds some complexity to the data representation structure and there would be extra information that needs to be transmitted from the sensor nodes. Considering the fact that sensors nodes have limited process and memory capabilities, the data representation could appear as a bottleneck to the design. To address this issue, we assume each sensor node will utilise a gateway or a similar solution to wrap the observation data in the particular data type which is measured by the sensor without requiring to be aware of the whole ontology structure. This means, that essentially

---

[9] http://protege.stanford.edu/

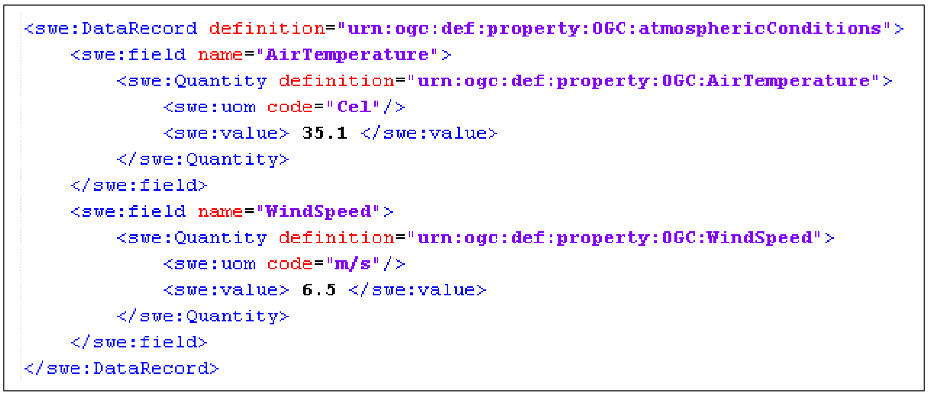
[10] http://sweet.jpl.nasa.gov/ontology/

the measurement and observation data from a sensor node will be in a format which complies with the SensorData ontology.

The data analysis and using ontology-based reasoning to extract additional knowledge from the data will only occur in processing nodes which have more powerful processing capabilities. The major cost of using the proposed method will be some extension in volume of the transmitted data from the sensor node.

5. Evaluation and Discussion

To evaluate the proposed approach, we focus on expressibility and scalability of the representations for different types of sensor observation and measurement data. Using a composite data as an example, we demonstrate how the representations differ by employing pure XML serialisation as those suggested by SensorML and the proposed semantic model. We also illustrate the RDFa[11] annotation of XML data which supports both legacy and the semantic data models [10]. Another important issues is the size increment for the data representation when the semantic model is utilised.



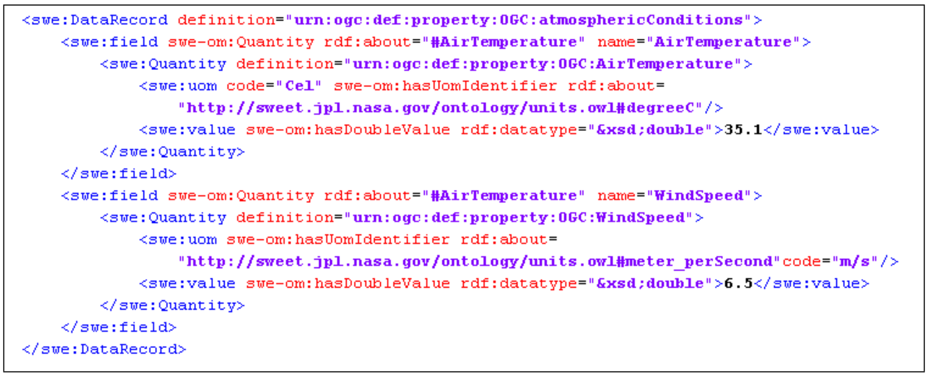
```
<swe:DataRecord definition="urn:ogc:def:property:OGC:atmosphericConditions">
    <swe:field name="AirTemperature">
        <swe:Quantity definition="urn:ogc:def:property:OGC:AirTemperature">
            <swe:uom code="Cel"/>
            <swe:value> 35.1 </swe:value>
        </swe:Quantity>
    </swe:field>
    <swe:field name="WindSpeed">
        <swe:Quantity definition="urn:ogc:def:property:OGC:WindSpeed">
            <swe:uom code="m/s"/>
            <swe:value> 6.5 </swe:value>
        </swe:Quantity>
    </swe:field>
</swe:DataRecord>
```

Figure 3: A sample sensor data in plain XML

The amount of data to be sent to the network is essential for the power consumption of the sensors. The main assumption is, transmitting more data requires more energy to transmit. Figure 3 shows a data snippet created in plain XML, and Figure 4 demonstrates the same data record expression using RDFa annotations. Using RDFa enables the representations to be compatible with the legacy data models such as SensorML and at the same time semantic meta-data can be added to the main structure. Figure 5 describes the sample data record according to the semantic model (represented in OWL form). The measurement data is not directly embedded into the data record. Rather the record is related to two quantities measuring two different phenomena of the physical world. All the used concepts are defined

---

[11] http://www.w3.org/TR/xhtml-rdfa-primer/



```xml
<swe:DataRecord definition="urn:ogc:def:property:OGC:atmosphericConditions">
    <swe:field swe-om:Quantity rdf:about="#AirTemperature" name="AirTemperature">
        <swe:Quantity definition="urn:ogc:def:property:OGC:AirTemperature">
            <swe:uom code="Cel" swe-om:hasUomIdentifier rdf:about=
                "http://sweet.jpl.nasa.gov/ontology/units.owl#degreeC"/>
            <swe:value swe-om:hasDoubleValue rdf:datatype="&xsd;double">35.1</swe:value>
        </swe:Quantity>
    </swe:field>
    <swe:field swe-om:Quantity rdf:about="#AirTemperature" name="WindSpeed">
        <swe:Quantity definition="urn:ogc:def:property:OGC:WindSpeed">
            <swe:uom swe-om:hasUomIdentifier rdf:about=
                "http://sweet.jpl.nasa.gov/ontology/units.owl#meter_perSecond"code="m/s"/>
            <swe:value swe-om:hasDoubleValue rdf:datatype="&xsd;double">6.5</swe:value>
        </swe:Quantity>
    </swe:field>
</swe:DataRecord>
```

Figure 4: A sample sensor data description in XML + RDFa annotations

externally and are therefore referenced in the descriptions. The overhead in OWL is significant compared to the pure XML specification. Sensor devices are typically constrained by the transmission power and processing capabilities.

```xml
<swe-om:Quantity rdf:ID="Quantity_AirTemperature">
    <swe-om:hasUomIdentifier rdf:resource=
        "http://sweet.jpl.nasa.gov/ontology/units.owl#degreeC"/>
    <swe-om:hasDoubleValue rdf:datatype="http://www.w3.org/2001/XMLSchema#double"
    >35.1</swe-om:hasDoubleValue>
    <swe-om:hasName xml:lang="en">air temperature</swe-om:hasName>
    <swe-om:hasDefinition rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI"
    >urn:ogc:def:property:OGC:AirTemperature</swe-om:hasDefinition>
</swe-om:Quantity>
<swe-om:Quantity rdf:ID="Quantity_WindSpeed">
    <swe-om:hasDefinition rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI"
    >urn:ogc:def:property:OGC:Windspeed</swe-om:hasDefinition>
    <swe-om:hasName xml:lang="en">wind speed</swe-om:hasName>
    <swe-om:hasUomIdentifier rdf:resource=
        "http://sweet.jpl.nasa.gov/ontology/units.owl#meter_perSecond"/>
    <swe-om:hasDoubleValue rdf:datatype="http://www.w3.org/2001/XMLSchema#double"
    >6.5</swe-om:hasDoubleValue>
</swe-om:Quantity>
<swe-om:DataRecord rdf:ID="DataRecord_AtmosphericConditions">
    <swe-om:hasField rdf:resource="#Quantity_AirTemperature"/>
    <swe-om:hasField rdf:resource="#Quantity_WindSpeed"/>
    <swe-om:hasDefinition rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI"
    >urn:ogc:def:property:OGC:atmosphericConditions</swe-om:hasDefinition>
</swe-om:DataRecord>
```

Figure 5: An example of the sensor data description in OWL

Figure 6 illustrates data size increment utilising the semantic data model compared to the legacy data. The machine interpretable representation needs more then the double amount of data to be transmitted to the sensor network. This would lead to an increase of sensor nodes' power consumption. most of the overhead consists of self explanatory meta-data that helps the receiver of the information to interpret the data.

In many real world applications, the power saving requirement of the sensor nodes has a

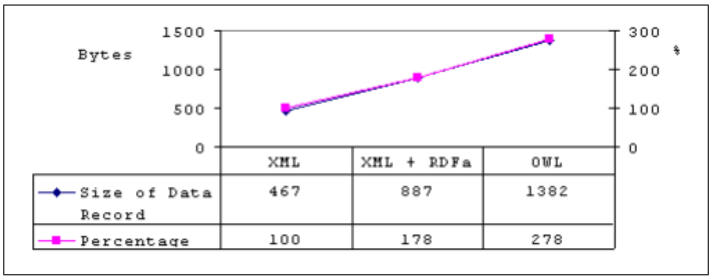


Figure 6: Evaluation of data size in the example

higher priority than the provisioning of a machine interpretable format delivered by the sensors itself. Increasing the power consumption means cutting the lifetime of a battery powered sensor. Such a trade-off between lifetime and machine interpretable data is very critical and needs to be addressed using other components in the sensor network architecture. A potential solution is using gateways between sensor node and a sensor network. In this case, the gateway receives binary data from sensor node which is optimised for the power and processing efficiency. The gateway then applies the meta-data template to the data to construct the semantic representation of sensor observation and measurement data. This enables the sensor nodes to operate in optimum mode and the gateway components will be responsible to construct the semantic representations of the sensor measurement and observation data.

6. Conclusions

The current data exchange for sensor networks relies on syntactic models which do not provide machine interpretable meanings to the data. A semantic model will not only enable a more interoperable structure for sensor data, it will also enable machines to process and interpret the emerging semantics to create more intelligent sensor networks. In this work we have analysed the current state of the art on sensor data modelling and employing the Semantic Web technologies. As a combination we propose a sensor data ontology that provides a complete semantic data modelling framework. The major drawback of introducing sensor networks (which are traditionally designed to be of low complexity) to semantic data modelling is the addition of meta-data that needs to be exchanged alongside the measured data. There are however several deployment and operational mechanisms that can keep this added complexity at bay. As already mentioned in section 4, processing of the meta-data, i.e. adding the semantics, can be achieved once the data has left the low complexity part of a sensor network, for instance the gateway or sink node can provide the additional processing; hence only keeping for instance binary XML formats on the sensor network side. The meta data annotation will be assigned to a designated gateway which receives the raw data and wraps the value with annotations taken from a template (i.e. semantic model). The annotated data can then be transmitted to the information subscribers.

The future work will focus on the evaluation of the impact of adding meta-data to the measured data on the sensor side and using binary XML to keep the sensor network side lightweight. In addition, all other processing to integrate the sensor data into the semantic data model will be outsourced to the sink or gateway. The context modelling will be also another step in developing automated mechanisms for resources discovery, composition, and utilisation in a semantic-enabled sensor network architecture.

References

[1] Botts, M. and Percivall, G., and Reed, C. and Davidson, J., OGC Sensor Web Enablement: Overview and High Level Architecture, The Open Geospatial Consortium.

[2] Sensor Model Language (SensorML), The Open Geospatial Consortium, `http://www.opengeospatial.org/standards/sensorml/`.

[3] G. Antoniou and F. van Harmelen, *A Semantic Web Primer (Cooperative Information Systems).* The MIT Press, April 2004.

[4] T. Berners-Lee, J. Hendler, and O. Lassila, “The semantic web,” *Scientific American*, vol. 284, pp. 28–37, May 2001.

[5] Jena: A Semantic Web Framework for Java, `http://jena.sourceforge.net/`.

[6] Sesame: RDF Schema Querying and Storage, `http://www.openrdf.org/`.

[7] D. Russomanno, C. Kothari, and O. Thomas, “Sensor ontologies: from shallow to deep models,” *System Theory, 2005. SSST ’05. Proceedings of the Thirty-Seventh Southeastern Symposium on*, pp. 107–112, March 2005.

[8] J.-H. Kim, H. Kwon, D.-H. Kim, H.-Y. Kwak, and S.-J. Lee, “Building a service-oriented ontology for wireless sensor networks,” *Computer and Information Science, 2008. ICIS 08. Seventh IEEE/ACIS International Conference on*, pp. 649–654, May 2008.

[9] M. Eid, R. Liscano, and A. El Saddik, “A universal ontology for sensor networks data,” *Computational Intelligence for Measurement Systems and Applications, 2007. CIMSA 2007. IEEE International Conference on*, pp. 59–62, June 2007.

[10] A. Sheth, C. Henson, and S. Sahoo, “Semantic sensor web,” *Internet Computing, IEEE*, vol. 12, pp. 78–83, July-Aug. 2008.

[11] A. Sheth and M. Perry, “Traveling the semantic web through space, time, and theme,” *IEEE Internet Computing*, vol. 12, no. 2, pp. 81–86, 2008.

[12] C. Henson, A. Sheth, P. Jain, and T. Rapoch, “video on the semantic sensor web,” *W3C Video on the Web Workshop*, 2007. `http://www.w3.org/2007/08/video/papers.html`.