Semantic Sensor Web

Amit Sheth, Cory Henson,
and Satya S. Sahoo • *Kno.e.sis Center, Wright State University*

In March 2008, heavy rainstorms across the Midwestern region of the US caused many rivers to breach their banks. Residents of Valley Park, a small town along the Meramec River, Missouri, had to decide whether to rely on a newly constructed levee or abandon their homes for higher ground.[1] Although the levee held, many chose the latter option and fled their homes; it was a chaotic situation that might have been avoided through access to better situational knowledge regarding the current water pressure and the levee's structural integrity. Had pressure sensors been embedded in the levee, they might have provided accurate real-time information that let residents make informed decisions about the safety of the levee, their homes, and themselves. This scenario demonstrates the increasingly critical role of sensors that collect and distribute observations of our world in our everyday lives.

In recent years, sensors have been increasingly adopted by a diverse array of disciplines, such as meteorology for weather forecasting and wildfire detection (www.met.utah.edu/mesowest/), civic planning for traffic management (www.buckeyetraffic.org/), satellite imaging for earth and space observation (http://vast.uah.edu/), medical sciences for patient care using biometric sensors (www.liebertonline.com/doi/abs/10.1089/109350703322682531), and homeland security for radiation and biochemical detection at ports (www.msnbc.msn.com/id/8092280). Sensors are thus distributed across the globe, leading to an avalanche of data about our environment. The rapid development and deployment of sensor technology involves many different types of sensors, both remote and in situ, with diverse capabilities such as range, modality, and maneuverability. Today, it's possible to use sensor networks to detect and identify a multitude of observations, from simple phenomena to complex events and situations. The lack of integration and communication between these networks, however, often isolates important data streams and intensifies the existing problem of too much data and not enough knowledge.

With a view to addressing this problem, we discuss a semantic sensor Web (SSW) in which sensor data is annotated with semantic metadata to increase interoperability as well as provide contextual information essential for situational knowledge. In particular, this involves annotating sensor data with spatial, temporal, and thematic semantic metadata.

Background

The SSW approach presented here leverages current standardization efforts of the Open Geospatial Consortium (OGC; www.opengeospatial.org) and Semantic Web Activity of the World Wide Web Consortium (W3C; www.w3.org/2001/sw/) to provide enhanced descriptions and meaning to sensor data. We'll review relevant components of these next. Also relevant but outside the scope of this article is the semantic community Sensor Standards Harmonization Working Group (http://semanticommunity.wik.is/Sensor_Standards_and_Data_Harmonization) which takes user perspective.

OGC Sensor Web Enablement

The sensor Web is a special type of Web-centric information infrastructure for collecting, modeling, storing, retrieving, sharing, manipulating, analyzing, and visualizing information about sensors and sensor observations of phenomena.[2] The OGC, an international consortium of industry, academic, and government organi-

 1089-7801/08/$25.00 © 2008 IEEE

zations tasked with developing open geospatial standards, describes the sensor Web as "Web-accessible sensor networks and archived sensor data that can be discovered and accessed using standard protocols and application program interfaces."[3] The sensor Web has vast significance for applications using sensor technologies to attain actionable situation awareness. Lack of standardization, however, is the primary barrier to realizing a progressive sensor Web.

The OGC recently established Sensor Web Enablement (SWE) to address this aim by developing a suite of specifications related to sensors, sensor data models, and sensor Web services that will enable sensors to be accessible and controllable via the Web.[3] The core suite of language and service interface specifications includes the following:

- *Observations and Measurements* (O&M). These are standard models and XML schema for encoding archived and real-time observations and measurements from a sensor.
- *Sensor Model Language* (SML). These are standard models and XML schema for describing sensors systems and processes; they provide information needed for discovering sensors, locating sensor observations, processing low-level sensor observations, and listing taskable properties.
- *Transducer Model Language* (TML). These are standard models and XML schema for describing transducers and supporting real-time streaming of data to and from sensor systems.
- *Sensor Observation Service* (SOS). This is the standard Web service interface for requesting, filtering, and retrieving observations and sensor system information. It's also the intermediary between a client and an observation repository or near real-time sensor channel.
- *Sensor Planning Service* (SPS). This is the standard Web service interface for requesting user-driven acquisitions and observations. It's also intermediary between a client and a sensor collection management environment.
- *Sensor Alert Service* (SAS). This is the standard Web service interface for publishing and subscribing to alerts from sensors.
- *Web Notification Services* (WNS). This is the standard Web service interface for asynchronous delivery of messages or alerts from SAS and SPS Web services and other elements of service workflows.[3]

W3C Semantic Web

The Semantic Web, as envisioned by Tim Berners-Lee and described by the W3C Semantic Web Activity, is an evolving extension of the World Wide Web in which the semantics, or meaning, of information on the Web is formally defined. Formal definitions are captured in ontologies, making it possible for machines to interpret and relate data content more effectively. The principal technologies of the Semantic Web include the Resource Description Framework (RDF) data representation model and the ontology representation languages RDF Schema and Web Ontology Language (OWL).

Semantics of Sensors: Within Space, Time, and Theme

Sensors encoding of observed phenomena are by nature opaque (often in binary or proprietary formats); therefore, metadata play an essential role in managing sensor data. A semantically rich sensor network would provide spatial, temporal, and thematic information essential for discovering and analyzing sensor data.[4]

Spatial metadata provide information regarding the sensor location and data, in terms of either a geographical reference system, local reference, or named location (see Figure 1). Local reference is especially useful when a sensor is attached to a moving object such as a car or airplane. Although the sensor's location is constantly changing, its location can be statically determined relative to the moving object. In addition, data from remote sensors, such as video and images from cameras and satellites, require complex spatial models to represent the field of view being monitored, which is distinct from the sensor's location.

The semantic sensor Web enables interoperability and advanced analytics for situation awareness and other advanced applications from heterogeneous sensors.

Temporal metadata provides information regarding the time instant or interval when the sensor data is captured. Thematic metadata describe a real-world state from sensor observations, such as objects or events. Every discipline contains unique domain-specific information, such as concepts describing weather phenomena, structural integrity values of buildings, and biomedical events representing a patient's health status.

Thematic metadata can be created or derived by several means, such as sensor data analysis, extraction of textual descriptions, or social tagging.

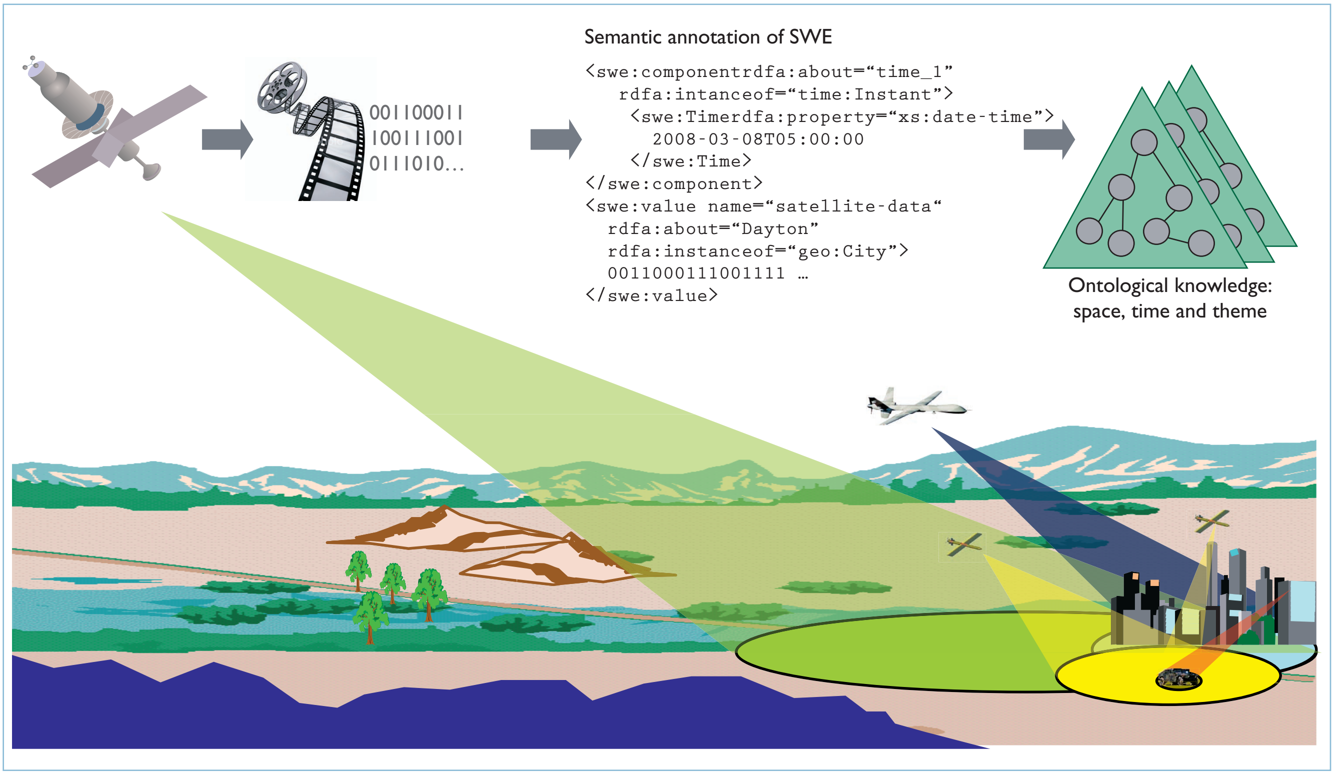

Figure 1. Progression from natural phenomena to raw sensor data to semantic annotation of Sensor Web Enablement to ontological knowledge of space, time, and theme.

Whereas the languages provided by the OGC SWE provide annotations for simple spatial and temporal concepts such as *spatial coordinate* and *time stamp*, more abstract concepts, such as *spatial region*, *temporal interval*, or any domain-specific thematic entity, would benefit from an ontological representation's expressiveness. Consider, for example, the semantics of a query about weather information at a particular time and place. The type of weather condition being sought could be a simple phenomenon, such as a single temperature reading, or a complex one, such as a tsunami. The location type within the query could be a single coordinate location, a spatial region within a bounding-box, or a named location such as a park or school. The semantics of the time interval specified by the query could be about weather conditions that fall *within* the time interval, *contain* the time interval, or *overlap* with the time interval. The type of metadata necessary to answer the queries listed requires knowledge of the situation the sensors observe. Such knowledge can be represented in ontologies and used to annotate and reason over sensor data to answer complex queries.

Next, let's look at the SSW and how it integrates the semantic metadata within the sensors domain.

Semantic Sensor Web

The SSW is a framework for providing enhanced meaning for sensor observations so as to enable situation awareness. It enhances meaning by adding semantic annotations to existing standard sensor languages of the SWE. These annotations provide more meaningful descriptions and enhanced access to sensor data than SWE alone, and they act as a linking mechanism to bridge the gap between the primarily syntactic XML-based metadata standards of the SWE and the RDF/OWL-based metadata standards of the Semantic Web. In association with semantic annotation, ontologies and rules play an important role in SSW for interoperability, analysis, and reasoning over heterogeneous multimodal sensor data.

Semantic Annotation

Many languages can be used for annotating sensor data, such as RDFa, XLink, and SAWSDL (Semantic Annotations for WSDL and XML Schema). Here, we describe the use of RDFa, a W3C proposed standard (www.w3.org/2006/07/SWD/RDFa/) and a markup language that enables the layering of RDF information on any XHTML or XML document. RDFa provides a set of attributes that can represent semantic metadata within an XML language from which we can extract RDF triples using a simple mapping. The core subset of RDFa attributes (http://en.wikipedia.org/wiki/RDFa) include

- *about* — a URI extracted as the

subject of an RDF triple that specifies the resource the metadata is about;
- *rel* and *rev* — extracted as the object property (predicate) of an RDF triple, these URI's specify a relationship or reverse-relationship with another resource;
- *href, src* and *resource* — extracted as the object of an RDF triple, this URI specifies the partner resource;
- *property* — extracted as the datatype property (predicate) of an RDF triple, this URI specifies a property for the content of an element; and
- *instanceof* — extracted as the object property "rdf:type" coupled with an RDF triple's object, this optional attribute specifies the RDF type of the subject (the resource that the metadata is about).

The following example shows a timestamp encoded in O&M and semantically annotated with RDFa. The timestamp's semantic annotation describes an instance of *time: Instant* (here, *time* is the namespace for an OWL-Time ontology):

```
<swe:component rdfa:
about="time_1" rdfa:
instanceof="time:Instant">
  <swe:Time rdfa:property=
  "xs:date-time">2008-
  0308T05:00:00</swe:Time>
</swe:component>
```

This example generates two RDF triples. The first, *time_1 rdf:type time:Instant*, describes *time_1* as an instance of *time:Instant* (subject is *time_1*, predicate is *rdf:type*, object is *time:Instant*). The second, *time_1 xs: date-time "2008-03-08T05:00:00,"* describes a data-type property of *time_1* specifying the time as a literal value (subject is *time_1*, predicate is *xs:date-time*, object is *"2008-03-08T05:00:00"*). This example illustrates the simple mechanics of embedding semantics in an XML document using RDFa. Semantically annotating SWE languages enables software applications to "understand" and reason over sensor data consistently, coherently, and accurately.

Ontologies

An ontology is a formal representation of a domain, composed of concepts and named relationships. At a broad level, we can classify ontologies along the three types of semantics associated with sensor data — spatial, temporal, and thematic — in addition to ontological models representing the sensor domain.

Several ongoing initiatives are helping to build relevant ontologies within various communities, such as the US National Institute of Standards and Technology (www.nist.gov/), the W3C, and the OGC. NIST has initiated a project titled "Sensor Standards Harmonization" to develop a common sensor ontology based on the existing standards within the sensor domain, including IEEE 1451, ANSI N42.42, the Chemical, Biological, Radiological, and Nuclear (CBRN) Data Model, and the OGC SWE languages. Several efforts are also underway to design an expressive geospatial ontology, including the W3C Geospatial Incubator Group (www.w3.org/2005/Incubator/geo/) and the Geographic Markup Language Ontology (http://loki.cae.drexel.edu/~wbs/ontology/ogc-gml.htm) of the OGC. OWL-Time (www.w3.org/TR/owl-time/), a W3C-recommended ontology based on temporal calculus, provides descriptions of temporal concepts such as *instant* and *interval*, which supports defining interval queries such as *within*, *contains*, and *overlaps*. Domain-specific ontologies that model various sensor-related fields such as weather and oceanography (www.oostethys.org/) are also necessary to provide semantic descriptions of thematic entities. We envision a registry of domain-specific ontologies for the SSW, similar to the Open Biomedical Ontologies at the National Center for Biomedical Ontologies (www.bioontology.org/). Figure 2 shows a subset of concepts and their relations from a suite of ontologies in SSW, modeling the weather domain.

Rule-Based Reasoning

To derive additional knowledge from semantically annotated sensor data, it's necessary to define and use rules. Rule languages and rule-processing systems are evolving. To demonstrate rules application

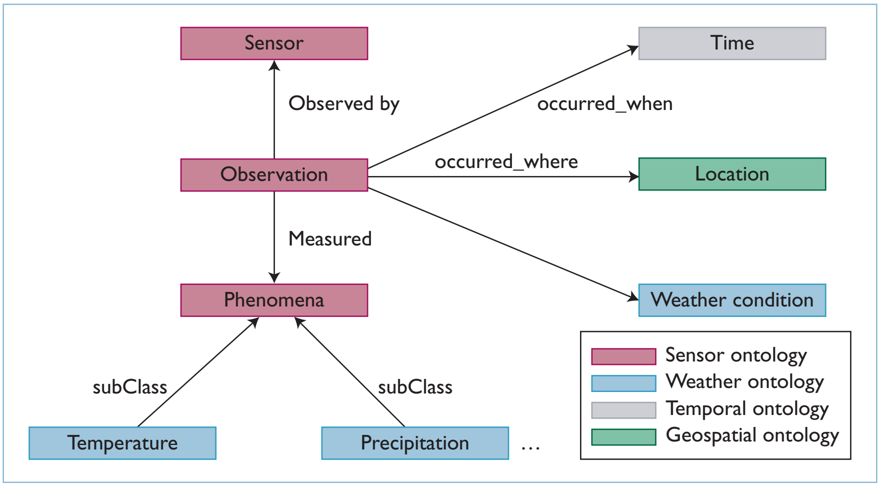

Figure 2. Subset of important concepts and relations in SSW, represented with a suite of ontologies. Specifically, a temporal, geospatial, sensor, and weather ontology.

Figure 3. A semantic mashup with ability to show videos that capture events from police cruiser cameras with user queries on spatial region and time interval.

and rule-based reasoning, we currently use Semantic Web Rule Language (SRWL)-based rules defined over OWL ontologies to deduce new ontological assertions from known instances. The W3C has proposed SWRL (www.w3.org/SWRL) as a standard rule language in the Semantic Web; it's based on OWL and uses the *antecedent → consequent* structure to define rules. Its primary advantage is that it seamlessly incorporates rules into an OWL ontology schema while providing enhanced application-specific expressivity.

The following sketch provides an example of rule usage in SSW: if a group of sensors explicitly provides information regarding temperature and precipitation, then, using these rules, we can specify possible road conditions. The following rule states that if the temperature is less than 32 degrees Fahrenheit and it's raining, then the roads are *potentially icy*.

```
Rule: Potentially Icy (with
freezing temperature and rain)
Observation(?obs) &
measured(?obs, ?precip) &
Rain(?precip) &
measured(?obs, ?temp) &
Temperature(?temp) &
temperature_value(?temp,
?tval) &
lessThanOrEqual(?tval, 32) &
unit_of_measurement(?temp,
Fahrenheit)
→ described(?obs, Potentially_
Icy)
```

SSW Application

As a proof of concept, we've implemented two prototype applications.

The first involves YouTube videos encoded in SensorML and semantically annotated with concepts from an OWL-Time ontology.[5] All videos in the prototype originate from Ohio State Patrol in-dash cameras that contain temporal information within the video frames. The temporal metadata is extracted using an open source optical character recognition (OCR) engine called Tesseract (http://code.google.com/p/tesseract-ocr/). Using this semantic metadata, we can retrieve videos by using semantic temporal concepts such as *within*, *contains*, or *overlaps* when querying with an interval of time. We can position the videos retrieved from a query onto a Google Map and play them from within an information window. Figure 3 shows a screenshot of the interface for this SSW prototype application.

The second prototype is an SOS, as specified by the SWE, which uses the SSW framework to enable complex queries over weather data. We refer to this type of service as a Semantic Sensor Observation Service (S-SOS). As described earlier, SOS is a service for requesting, filtering, and retrieving observations and sensor system information. SOS acts as an intermediary between a client and an observation repository or near real-time sensor channel. Our application implements an S-SOS weather service that uses weather readings available at BuckeyeTraffic.org, a Web site maintained by the Ohio Department of Transportation. BuckeyeTraffic provides road and weather observations from more than 200 sensors deployed along Ohio interstate highways. Our application collects and uses data including temperatures of the air, surface, subsurface, and dew point, as well as wind speed, wind direction, and precipitation. We collected and stored such data for one month at 10-second reading intervals. We then converted the data to O&M and SML representation formats and semantically annotated these documents with spatial, temporal, and weather ontological concepts. Figure 4 shows the overall architecture.

By leveraging SSW semantic annotations, we can fluently execute complex queries over simple weather readings. For example, let's revisit the example query from the previous section asking for weather information at a particular time and place. More specifically, suppose the query requests information about freezing or blizzard conditions. The freezing query requires only a temperature sensor and a rule specifying that any temperature less than 32 degrees Fahrenheit constitutes a freezing condition. The blizzard query, on the other hand, requires three sensor types — temperature, wind, and precipitation. In fact, we can describe the blizzard condition as a composition of several simple (single sensor) conditions including the freezing condition, high-winds condition, and snowing condition. Complex queries of this type require the situational awareness enabled by semantic annotation and reasoning over sensor data.

```
Rule: Blizzard Condition (with
freezing temperature, high
winds, and snow)
Observation(?obs) &
described(?obs, ?weather) &
FreezingCondition(?weather) &
HighWindCondition(?weather) &
SnowCondition(?weather) &
→BlizzardCondition(?weather)
```

By incorporating OGC and W3C standardization efforts into a SSW, we can provide an environment for enhanced query and reasoning within the sensor domain. We see great potential for the SSW in many different domains, including weather forecasting, oceanography, biometrics, video on the Web,[5] and EventWeb.[6]

In 1999, Neil Gross expressed a vision of the future in which sensors were ubiquitous and engrained in the fabric of our environment:

> In the next century, planet earth will don an electronic skin. It will use the Internet as a scaffold to support and transmit its sensations. This skin is already being stitched together. It consists of millions of embedded electronic measuring devices: thermostats, pressure gauges, pollution detectors, cameras, microphones, glucose sensors, EKGs, electroencephalographs. These will probe and monitor cities and endangered species, the atmosphere, our ships, highways and fleets of trucks, our conversations, our bodies — even our dreams.[2]

We share this vision and wish to provide meaning to this new world.

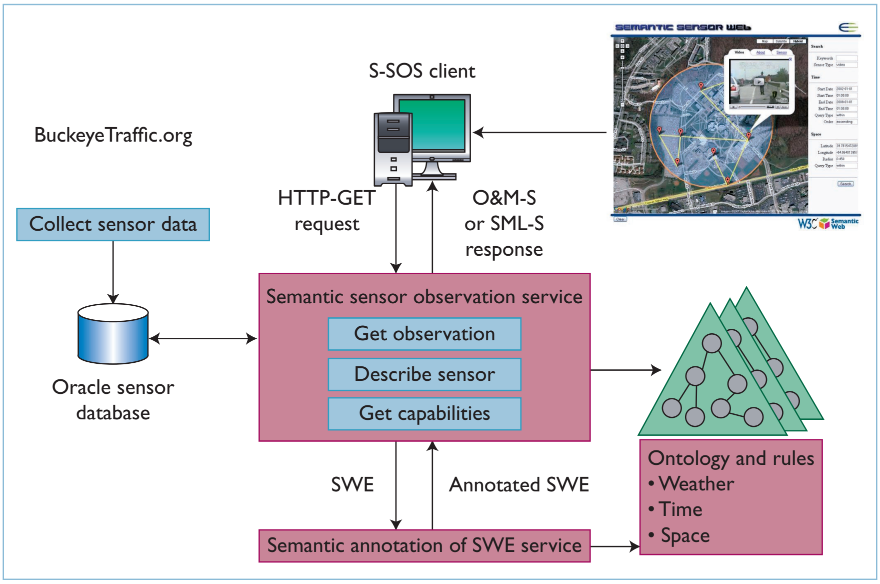

Figure 4. Semantic Sensor Observation Service (S-SOS) Architecture for a prototype application using SSW technologies to manage weather sensor observations.

References

1. J. Salter, “Heavy Rains Close Missouri Roads, Renew Flood Concerns,” *USA Today*, 10 Apr. 2008; www.usatoday.com/weather/storms/2008-04-10-missouri-flooding_N.htm.
2. Neil Gross, “The Earth Will Don an Electronic Skin,” *BusinessWeek*, Aug. 1999; www.businessweek.com/1999/99_35/b3644024.htm.
3. Mike Botts et al., “OGC Sensor Web Enablement: Overview and High Level Architecture (OGC 07-165),” Open Geospatial Consortium white paper, 28 Dec. 2007.
4. A. Sheth and M. Perry, “Traveling the Semantic Web through Space, Time, and Theme,” *IEEE Internet Computing*, vol. 12, no. 2, 2008, pp. 81–86.
5. C. Henson et al., “Video on the Semantic Sensor Web,” W3C Video on the Web Workshop, 2007; www.w3.org/2007/08/video/papers.html.
6. J. Ramesh, “EventWeb: Developing a Human-Centered Computing System,” *Computer*, vol. 41, no. 2, 2008, pp. 42–50.

Amit Sheth is the LexisNexis Ohio Eminent Scholar and director of Kno.e.sis Center at Wright State University. His research interests include the Semantic Web and its applications, semantics services and middleware, and computing for human experience. Sheth has a PhD in computer and information science from Ohio State University. He's an IEEE fellow. Contact him at amit.sheth@wright.edu; http://knoesis.wright.edu/amit.

Cory Henson is a PhD student at the Kno.e.sis Center at Wright State University. His research interests include knowledge representation and ontology modeling in order to reason over sensor data and manage situational awareness. Henson has a BS in computer science from the University of Georgia. Contact him at cory@knoesis.org; http://knoesis.wright.edu/researchers/cory.

Satya S. Sahoo is a PhD student at the Kno.e.sis Center at Wright State University. His research interests include semantic provenance, knowledge representation, and information integration in biomedical and sensor Web domains. Sahoo has a masters in computer application from Goa University, India. Contact him at satya@knoesis.org; http://knoesis.wright.edu/researchers/satya.