

Video on the Semantic Sensor Web

[Cory Henson](#), [Amit Sheth](#), Prateek Jain, Josh Pschorr, Terry Rapoch
[Kno.e.sis Center](#) and [daytaOhio](#)
[Wright State University](#), Dayton, OH 54324

*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.*

- Neil Gross, "The earth will don an electronic skin," BusinessWeek, August 1999

Introduction

Millions of sensors around the globe currently collect avalanches of data about our world. The rapid development and deployment of sensor technology is intensifying the existing problem of too much data and not enough knowledge. With a view to alleviating this glut, we propose that sensor data, especially video sensor data, can be annotated with semantic metadata to provide contextual information about videos on the Web. In particular, we present an approach to annotating video sensor data with spatial, temporal, and thematic semantic metadata. This technique builds on current standardization efforts within the W3C and Open Geospatial Consortium (OGC) and extends them with Semantic Web technologies to provide enhanced descriptions and access to video sensor data.

Background on the Sensor Web

The Sensor Web is a special type of web-centric information system for collecting, modeling, storing, retrieving, sharing, manipulating, analyzing, and visualizing information of sensors, sensor observations, and associated phenomena.¹ Lack of standardization is the primary barrier to the realization of a progressive Sensor Web. The OGC has recently established the Sensor Web Enablement Group in order to address this problem by developing a suite of specifications related to sensors, sensor data models, and sensor web services.² The core suite of specifications includes:

- *Observations & Measurements (O&M)* – Standard models and XML schema for encoding observations and measurements from a sensor, both archived and real time.
- *Sensor Model Language (SML)* – Standard models and XML schema for describing sensors systems and processes; provides information needed for discovery of sensors, location of sensor observations, processing of low-level sensor observations, and listing of taskable properties.

- *Transducer Model Language (TML)* – The conceptual model and XML schema for describing transducers and supporting real-time streaming of data to and from sensor systems.
- *Sensor Observations Service (SOS)* – Standard Web service interface for requesting, filtering, and retrieving observations and sensor system information. This is the intermediary between a client and an observation repository or near-real-time sensor channel.
- *Sensor Planning Service (SPS)* – Standard Web service interface for requesting user-driven acquisitions and observations. This is the intermediary between a client and a sensor collection management environment.
- *Sensor Alert Service (SAS)* – Standard Web service interface for publishing and subscribing to alerts from sensors.
- *Web Notification Services (WNS)* – Standard Web service interface for asynchronous delivery of messages or alerts from SAS and SPS Web services and other elements of service workflows.

Introducing the Semantic Sensor Web

Because of the opaque nature of observed phenomena encodings, metadata play an essential role in managing sensor data. A semantically rich Sensor Web would provide provenance-context and thematic information essential for discovery and retrieval of video sensor data. Provenance-context metadata include spatial and temporal information such as the spatial region and temporal interval contained within a video. Space and time metadata can be provided by associating each video with the spatial and temporal coordinates of the corresponding sensor at a particular location and time. Thematic metadata describe an interpretation of real-world state from sensor observations, such as objects or events. Theme can be provided by several means, such as video analysis, extraction of textual descriptions, or social tagging. The latter approach can be found on many contemporary video websites such as Picasa Web and YouTube. Whereas SML provides annotations for simple 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 the expressiveness of an ontological representation. Consider, for example, the semantics of a simple query asking for videos at a particular period of time. Does the user want videos that fall *within* the time interval, *contain* the time interval, or *overlap* the time interval? OWL-Time, a W3C-recommended ontology based on temporal calculus, could provide descriptions of such temporal concepts.³ In addition, the W3C has recently established a Geospatial Incubator Group (GeoXG) tasked with developing a spatial ontology.⁴ Domain-specific ontologies would be needed to provide semantic descriptions of thematic entities such as objects/events and their relations. A linking mechanism is needed to bridge the gap between the XML-based metadata standards of the Sensor Web Enablement Group and the RDF/OWL-based metadata standards of the semantic Web. Several solutions to the fundamental Web problem of linking metadata may be explored, including SA-REST and SAWSDL, which is a W3C recommendation.^{5,6} One such solution is XLink, XML Linking Language, a markup language that “allows elements to be inserted into XML documents in order to create and describe links between resources. XLink provides a

framework for creating both basic unidirectional links and more complex linking structures. It allows XML documents to:

- Assert linking relationships among more than two resources
- Associate metadata with a link
- Express links that reside in a location separate from the linked resources⁷⁷

Figure 1 shows a fragment of an SML document describing a video on the Web. Notice, in particular, the role XLink plays in providing a link from the SML document to an ontological resource. This example shows how two time-stamps annotated in SML are linked through model references to two instances of Instant in an OWL-Time ontology.

```
<swe:definition>
  <swe:dataComponents name="video-provenance">
    <swe:DataGroup>
      <swe:component name="start-time"
        xlink:type="simple"
        xlink:role="time:Instant"
        xlink:href="http://knoesis.wright.edu/projects/sensorweb/ontology/video-time.owl#instant_6HVMHPjDY2M_begin">
        <swe:Time definition="urn:ogc:def:phenomenon:time"
          referenceTime="1970-01-01T00:00:00Z"
          uom="urn:ogc:def:unit:date-time"/>
      </swe:component>
      <swe:component name="end-time"
        xlink:type="simple"
        xlink:role="time:Instant"
        xlink:href="http://knoesis.wright.edu/projects/sensorweb/ontology/video-time.owl#instant_6HVMHPjDY2M_end">
        <swe:Time definition="urn:ogc:def:phenomenon:time"
          referenceTime="1970-01-01T00:00:00Z"
          uom="urn:ogc:def:unit:date-time"/>
      </swe:component>
      <swe:component name="latitude">
        <swe:Quantity definition="urn:ogc:def:phenomenon:angle"
          uom="urn:ogc:def:unit:degree"/>
      </swe:component>
      <swe:component name="longitude">
        <swe:Quantity definition="urn:ogc:def:phenomenon:angle"
          uom="urn:ogc:def:unit:degree"/>
      </swe:component>
    </swe:DataGroup>
  </swe:dataComponents>
  <swe:encoding>
    <swe:AsciiBlock decimalSeparator=","
      tokenSeparator=" "
      tupleSeparator=" " />
  </swe:encoding>
</swe:definition>
<swe:value name="video-provenance">
  2002-11-10T015:31:00.00,2002-11-10T015:34:31.00,39.779535,-84.063821
</swe:value>
```

Figure 1. XLink from SML to OWL-Time

Prototyping the Semantic Sensor Web

We are currently developing a prototype application for the Semantic Sensor Web (see Figure 2 for a screenshot). The application dataset includes YouTube videos annotated with SML and XLink model references to an OWL-Time ontology. All videos used in the prototype originate from Ohio State Patrol in-dash cameras that contain temporal information within the video frames. The temporal metadata are extracted using an open source optical character recognition (OCR) engine called Tesseract.⁸ Figure 3 provides a brief overview of the extraction process. Utilizing these semantic metadata we are able to retrieve videos by semantic temporal concepts such as *within*, *contains*, or *overlaps* when querying with an interval of time. The videos retrieved from a query are positioned on a Google Map and playable from within an information window. We plan to generate model references from SML to geospatial ontology in order to provide more expressive spatial queries in the future. In addition, keyword tags gathered from YouTube will be used to provide thematic metadata and extended with model references to domain-specific ontologies.

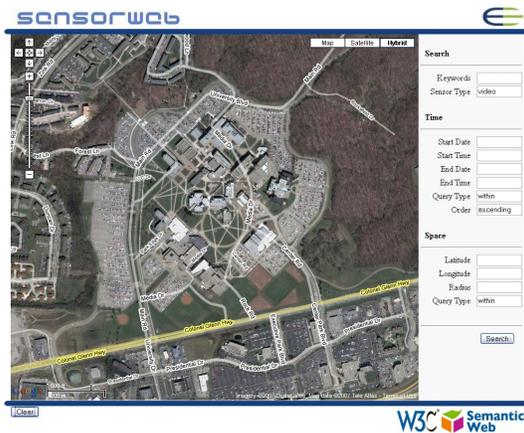


Figure 2. Prototype Application

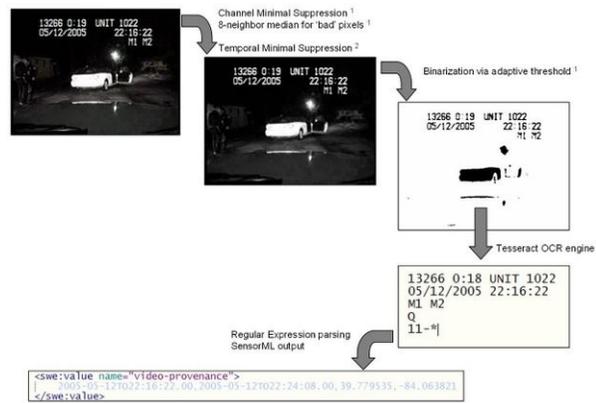


Figure 3. Extraction Process

Conclusion

By incorporating the standardization efforts of the OGC and W3C into a robust Semantic Sensor Web, we are able to provide an environment for the effective discovery and retrieval of video sensor data. Beyond video on the Web, this framework could play an important role in emergent applications, including video on mobile devices. Imagine a mobile device (perhaps a phone running Android OS) capable of capturing video and annotating it with semantic metadata. Immediately the video would be associated with provenance-context information such as time and place in addition to thematic information submitted by the user. Such metadata could then be used to query and retrieve similar videos or share this experience with the world. We believe that the expressivity and accessibility provided by such an integrated vision is essential to realizing video as a first-class citizen of the Web.

References

1. Neil Gross, The earth will don an electronic skin, BusinessWeek, August 1999.
2. Open Geospatial Consortium, Sensor Web Enablement WG, <http://www.opengeospatial.org/projects/groups/sensorweb>
3. W3C, Time Ontology in OWL, <http://www.w3.org/TR/owl-time>
4. W3C, Geospatial Incubator Group, <http://www.w3.org/2005/Incubator/geo>
5. Amit Sheth et al., [SA-REST: Semantically Interoperable and Easier-to-Use Services and Mashups](#), IEEE Internet Computing, November/December 2007 (Vol. 11, No. 6) pp. 91-94
6. W3C, [Semantic Annotations for WSDL and XML Schema](#)
7. W3C, XML Linking Language, <http://www.w3.org/TR/xlink>
8. Google Code, Tesseract, <http://code.google.com/p/tesseract-ocr/>