

# Integración semántica de datos de observación mediante servicios SOS

Manuel A. Regueiro<sup>1</sup>, José R.R. Viqueira<sup>1</sup>, Christoph Stasch<sup>2</sup>, and José M. Cotos<sup>1</sup>

<sup>1</sup> Computer Graphics and Data Engineering Group (COGRADE)  
Centro de Investigación en Tecnologías da Información (CITIUS)  
Universidade de Santiago de Compostela (USC)  
Santiago de Compostela, Spain  
{manuelantonio.regueiro, jrr.viqueira, manel.cotos}@usc.es  
<sup>2</sup> 52°North Initiative for Geospatial Open Source Software GmbH  
Martin-Luther-King-Weg 24, 48155 Muenster, Germany  
c.stasch@52north.org

**Abstract.** En este documento se describe el trabajo en marcha de diseño e implementación de un servicio de acceso a datos de observación medioambiental SOS (Sensor Observation Service) que permite la integración semántica de fuentes de datos accesible a través de este estándar.

**Keywords:** OGC, Sensor Web, Observación Medioambiental, Ontologías, Integración Semántica

## 1 Introducción

La cantidad de datos de observación medioambiental que se están generando hoy en día en distintos ámbitos de aplicación es enorme, y la estimación de su crecimiento futuro a corto plazo es también muy considerable. A este gran volumen de datos hay que sumar la alta heterogeneidad de los mismos, propiciada por las gran variedad de tipos de dispositivos y sistemas de sensorización utilizados. El acceso a toda esta información desde componentes de toma de decisión está en general dificultado por dos grandes barreras. La primera es una barrera sintáctica relacionada con distintos formatos de representación e interfaces de acceso. Esta barrera se salva con la adopción de estándares como los propuestos por el Open Geospatial Consortium (OGC) en su iniciativa Sensor Web Enablement (SWE) <sup>3</sup>. En concreto, de gran importancia para este trabajo es la interfaz de servicio Sensor Observation Service (SOS) <sup>4</sup>. En la actualidad existen varias implementaciones de servicios SOS que permiten la integración de datos de sensores. La segunda es una barrera semántica y para superarla es necesario el establecimiento de relaciones entre los conceptos presentes en las fuentes de datos a integrar utilizando como marco vocabularios reconocidos de

<sup>3</sup> <http://www.opengeospatial.org/ogc/markets-technologies/swe>

<sup>4</sup> <http://www.opengeospatial.org/standards/sos>

referencia. La ontología SSN (Semantic Sensor Network), propuesta por el W3C proporciona un marco de referencia que incluye los conceptos presentes en el modelo Observations and Measurements (O&M) utilizado en la interfaz SOS. Complementariamente, la ontología SWEET (Semantic Web for Earth and Environmental Terminology) propuesta por la NASA proporciona un marco de referencia genérico para el dominio medioambiental.

El objetivo principal del presente trabajo es el desarrollo de un framework que facilite la integración semántica de datos de sensorización medioambiental de todo tipo, incluyendo datos vectoriales y raster generados por sensores que miden in-situ y de forma remota. El framework proporcionará una interfaz compatible con el estándar SOS y utilizará las ontologías SSN y SWEET como marco para que los expertos del dominio de aplicación puedan configurar la integración a nivel semántico de los datos.

Las secciones restantes proporcionan una síntesis de los trabajos relacionados más importantes (Sección 2), una descripción breve de la solución propuesta (Sección 3) y líneas de trabajo futuro (Sección 4).

## 2 Trabajo relacionado

El uso de estándares OGC-SWE en procesos de observación y monitorización medioambiental ha sido descrito en el ámbito de diversos proyectos [6], y la importancia de disponer de tecnologías de mediación de datos ha sido justificada en trabajos anteriores [9]. En [4], se muestra la evolución de los estándares SWE y se presenta la integración semántica de datos como un reto a abordar dentro del área de la web semántica de los sensores (*Semantic Sensor Web*) [10].

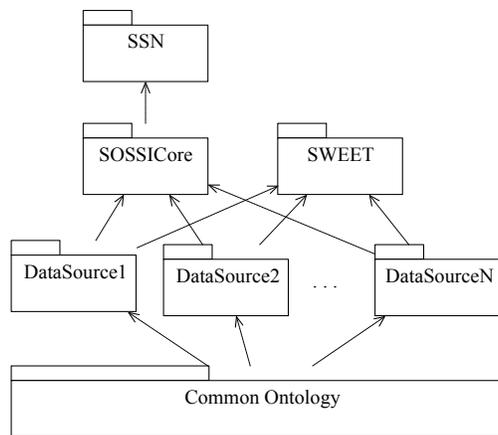
Respecto a la integración de datos científicos, en [2, 8] se propone una solución de mediación de datos basada en modelos. De forma resumida, las fuentes de datos a integrar deben proporcionar tanto los datos como modelos conceptuales que aporten semántica. Usando estos modelos y conocimiento adicional del dominio de aplicación (*glue knowledge*), el ingeniero de mediación define las vistas de integración de datos. Este conocimiento adicional se representa en forma de distintas relaciones entre conceptos de los modelos.

Si nos centramos en datos de sensores, en [3] se define un modelo conceptual de datos que permite añadir semántica de observación a conjuntos de datos, habilitando la anotación semántica de los mismos. La construcción de ontologías a partir de datos de sensorización medioambiental se aborda en [1] y la automatización del registro de sensores en entornos SWE mediante tecnologías semánticas se trata en [5].

Finalmente, en [7] se describe el diseño e implementación de un SOS semántico, en el que las peticiones SOS se transforman en expresiones SPARQL, cuyo RDF resultante es transformando a una codificación XML de O&M anotada semánticamente. En ninguno de los trabajos anteriores se aborda el problema de la integración semántica de datos de sensores a través de SOS.

### 3 Solución propuesta

El modelo O&M utilizado por el estándar SOS para representar y codificar los datos de observación define los siguientes conceptos fundamentales. De forma esquemática, cada *Observación* está compuesta por un valor observado, un instante de tiempo en el que el valor fue observado, una referencia a una *Propiedad* de una *Entidad* (*Feature of Interest*, FOI) observada y una referencia al *Proceso* de observación, generalmente un sensor, utilizado para generar el valor. Las observaciones en un SOS se organizan en *Ofertas* de manera que la consulta de observaciones dentro de una *Oferta* tenga poca probabilidad e no devolver datos. Los usuarios a través de SOS pueden consultar las observaciones de cada *Oferta*.



**Fig. 1.** Arquitectura de integración semántica de datos.

La solución de integración semántica de datos diseñada sigue una aproximación basada en la mediación basada en modelos [2, 8]. Como puede verse en la Fig.1, cada una de las fuentes de datos es anotada semánticamente con una ontología. Esta ontología utiliza como referencia los conceptos de O&M definidos en *SOSSICore* y los conceptos de modelado medioambiental definidos en *SWEET*. *SOSSICore* completa con conceptos SWE necesarios los ya proporcionados por *SSN*. Un ejemplo de estos conceptos necesarios es el concepto de *Oferta* que se define en *SOSSICore* como subclase del concepto *Observación* de *SSN*. Las *Ofertas* específicas de cada fuente de datos se definen como subclases de *Oferta* especificando las restricciones necesarias con *Procesos*, *Propiedades* y *FOIs*. En la *Common Ontology* se definen:

- Un conjunto de relaciones entre conceptos de las distintas ontologías de las fuentes de datos (Domain Maps en [2, 8]). Con estas relaciones se puede indicar por ejemplo que la Propiedad “Temperatura” de la fuente de datos

- “<http://www.meteogalicia.es/meteo>” es equivalente semánticamente a la propiedad “TemperaturaAire” de la fuente de datos “<http://www.meteogalicia.es/oceano>”.
- Nuevos conceptos que sirven para agrupar conceptos de varias fuentes de datos. Por ejemplo, un nuevo concepto “Temperatura” del que pueden ser definidos como subclasses los conceptos “TemperaturaAire” y “TemperaturaAgua” de dos fuentes distintas.
- Un conjunto de *Ofertas* globales definidas como vistas sobre la nueva visión integrada de las fuentes de datos.

## 4 Trabajo futuro

Las líneas de trabajo futuro relacionadas con este trabajo se centran principalmente en dos aspectos. El primero es la implementación de un prototipo que sirva como prueba de concepto. El segundo es la automatización de construcción de adaptadores que permitan publicar datos almacenados en bases de datos relacionales y archivos NetCDF a través de la interfaz SOS.

## References

1. Adams, B., Janowicz, K.: Constructing geo-ontologies by reification of observation data. In: Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. pp. 309–318. ACM, New York, NY, USA (2011)
2. Bowers, S., Lin, K., Ludascher, B.: On integrating scientific resources through semantic registration. In: Scientific and Statistical Database Management, 2004. Proceedings. 16th International Conference on. pp. 349–352 (June 2004)
3. Bowers, S., Madin, J., Schildhauer, M.: A conceptual modeling framework for expressing observational data semantics. In: Li, Q., Spaccapietra, S., Yu, E., Olivé, A. (eds.) Conceptual Modeling - ER 2008, Lecture Notes in Computer Science, vol. 5231, pp. 41–54. Springer Berlin Heidelberg (2008)
4. Bröring, A., Echterhoff, J., Jirka, S., Simonis, I., Everding, T., Stasch, C., Liang, S., Lemmens, R.: New generation sensor web enablement. *Sensors* 11(3), 2652–2699 (2011)
5. Bröring, A., Maué, P., Janowicz, K., Nüst, D., Malewski, C.: Semantically-enabled sensor plug & play for the sensor web. *Sensors* 11(8), 7568–7605 (2011)
6. Conover, H., Berthiau, G., Botts, M., Goodman, H.M., Li, X., Lu, Y., Maskey, M., Regner, K., Zavodsky, B.: Using sensor web protocols for environmental data acquisition and management. *Ecological Informatics* 5(1), 32 – 41 (2010)
7. Henson, C., Pschorr, J., Sheth, A., Thirunarayan, K.: Semsos: Semantic sensor observation service. In: Collaborative Technologies and Systems, 2009. CTS '09. International Symposium on. pp. 44–53 (May 2009)
8. Ludascher, B., Gupta, A., Martone, M.: Model-based mediation with domain maps. In: Data Engineering, 2001. Proceedings. 17th International Conference on. pp. 81–90 (2001)
9. Regueiro, M.A., Viqueira, J.R., Taboada, J.A., Cotos, J.M.: Virtual integration of sensor observation data. *Computers & Geosciences* (2015), to appear.
10. Sheth, A., Henson, C., Sahoo, S.: Semantic sensor web. *Internet Computing, IEEE* 12(4), 78–83 (July 2008)