Semantic Mediation of Environmental Observation Datasets through Sensor Observation Services

Título	Semantic Mediation of Environmental Observation Datasets through Sensor Observation Services
Autor/a	Manuel Antonio Regueiro Seoane
Directores	José Ramón Ríos Viqueira, José Ángel Taboada González
Tipo	Tese doutoral
Data de lectura	02/02/2017
Lugar de lectura	Universidade de Santiago de Compostela
Doutorado	Doutorado europeo

Abstract A large volume of environmental observation data is being generated as a result of the observation of many properties at the Earth surface. And it is estimated that this amount increases considerably in the future. In parallel, there exists a clear interest in accessing data from different data providers related to the same property, in order to solve concrete problems. If we restrict to the area of Galicia, we find regional, national and international organizations that manage intersecting sets of meteorological stations. All of these organizations have huge amounts of available data, sometimes overlapped, but with a heterogeneous storage and access, hampered their integration in an automatic way. Based on such fact, there is also an increasing interest in publishing the above data through open interfaces in the scope of SDIs. There have been important advances in the definition of open standards of the Open Geospatial Consortium (OGC) that enable interoperable access to sensor data. Among the proposed interfaces, the Sensor Observation Service (SOS) is having an important impact in the development of current environmental information systems. This service enables standardized access to collections of observation data generated by different processes, which are in most cases physical sensors. Observations inside a SOS are organized in Offerings. Each observation of each offering has a value of an Observed Property for a given time instant. Besides, the observation references the domain specific entity to which the property applies, called Feature of Interest and the Process used to obtain the value, commonly a physical sensor. The Observations and Measurements (O&M) specification provides a data model for those observation collections. We have realized that currently there is no available solution to provide integrated access to various data sources through a SOS interface. This problem shows up two main facets. On the one hand, the heterogeneity among different data sources has to be solved. On the other hand, another problem is the existence of semantic conflicts. The most direct solution would be given by an ad-hoc implementation on the client side. The main drawbacks would be the lack of generality and the need to implement complex clients for specialized users. From the server side there are two clear alternatives for data integration. The first solution would be a data warehouse approach. This solution has to support both entity-based and array-based data models. Therefore, Extract Transform and Load (ETL) processes are needed for each data source, where the heterogeneity and semantic conflicts have to be solved. All the current SOS implementations follow this philosophy; however they are restricted to observations generated by in-situ devices in relational technologies. The second alternative would be a semantic mediation approach. In this case, the queries through SOS have to be transformed in suitable queries to each data source. Mediator/Wrapper architectures are used in this case, where an adapter for the model and format is developed for each data source, and a mediator is developed to resolve semantic conflicts. There is currently no SOS implementation of this type. To solve the problems introduced in the preceding paragraphs, the main goal of this thesis is to design and develop a semantic data mediation framework to access any kind of environmental observation dataset, including both relational data sources and multidimensional arrays. The whole proposed solution will use a Mediator/Wrapper approach. The mediator will use semantic technologies to solve conflicts. Generic wrappers for spatial databases and for multidimensional array data sources accessed through NetCDF Subset interface will reduce the development cost.

LIGAZÓNS

🔗 Teseo

DESCARGAS

BIB
TEX Referencia BibTex

⬇ Descargar versión completa