Semantic web application in the context of marine remote sensing data catalogues

Dominique Briand, SISMER, IFREMER (France), dominique.briand@ifremer.fr

Context

The constant improvement of technologies, their central role and systematic use in marine scientific works lead to data volume increase and new data management issues. While the data volume augmentation is balanced by data processing and storing capacities augmentation of infrastructures, the context of "open data", the diversity of users and data leads to new specific questions addressed by, among others, people working on centralized services such as Data Information and Access Service (DIAS) to make it easier to discover, search, access, upload and process the data sets online.

In a world of interconnected information systems *interoperability* is a key word, it allows distant and independent infrastructures to communicate without any effort. Beyond technical aspects for exchanging data (http protocols, API, web services, etc.), international standards are also used to rule the way data are described or, in other words, what the metadata contains and how to serialize it (e.g.: xml specific tags). For geographic information, and specifically imagery, the ISO 19115-2 standard is commonly used. In Europe, the INSPIRE directive defines obligatory fields and controlled vocabulary (theme) that users must adopt.

However, it occurs that describing the data sets using those standards can be tricky: sometimes the fields or theme doesn't fit the data, and sometimes a data fits in various theme depending on the user's point of view. Still, in order to facilitate data discovery based on the metadata records, the use of a list of keywords is also possible and turns out to be essential.

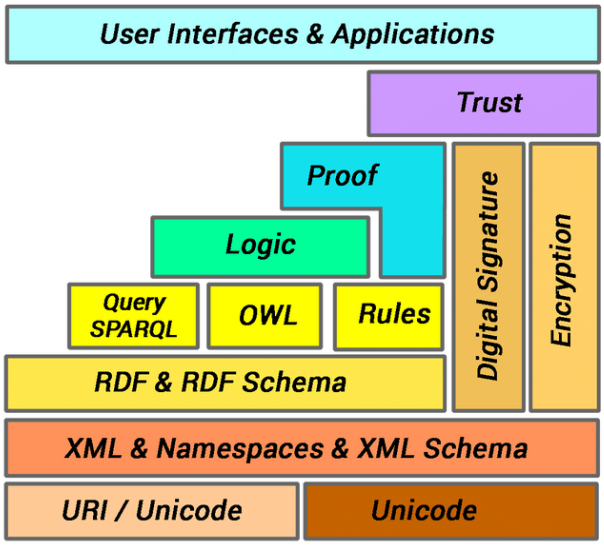

Figure 1: the semantic web technology stack

Controlled vocabularies and knowledge modelling tools from the semantic web (fig. 1), like thesaurus and ontologies, are a complementary approach to metadata standards in order to express better the large spectrum of final data applications and the different user's points of view. In addition to enjoy

more flexibility for describing the data sets, the use of technologies from the semantic web (also known as web 3.0 or web of linked data) provides new data searching opportunities. Indeed, the semantic web technologies enable software to better understand the content of the data and even make explicit links between them that were implicit.

Teams involved is this work

The SISMER (Scientific Information System for the Sea) is the service in IFREMER (French Research Institute for Exploitation of the Sea) in charge of the administration of numerous and heterogeneous databases and information systems. Among them, remote sensing data.

The CERSAT (Centre ERS d'Archivage et de Traitement) is the IFREMER multi-mission data centre for archiving, processing and validating data from space borne sensors (such as altimeters, scatterometers, radiometers, SAR,...) with a focus on sea surface parameters and air-sea interactions.

Application in the context of a multi-mission satellite centre

The principal challenge is to provide an efficient search tool that answers the needs of a large spectrum of users (that have different areas of interest). In other words, how to give each dataset the same chance to be retrieved? The cataloguing step is a key one as the search tool is based on the catalogue.

First a metadata application profile is defined, based on the ISO 19115-3 standard and the INSPIRE European recommendation.

Different initiatives for defining and using controlled vocabulary in remote sensing applications are used in order to define technical thesaurus. The Committee on Earth Observation Satellites (CEOS) database is the main one, it allows for every dataset to be described.

The entry point for the search tool is only the measured parameter. Even if the measured parameter is not recorded in the metadata application profile, the sensor is. An ontology is built in order to model the link between the different sensors and the different parameters.

An elastic search index is built, it combines information from the metadata application profile and the ontology. It allows for a quick response and also for users to dynamically display the more accurate facets in order to refine the results.

Perspectives

This work was done with the perspective of in-situ datasets integration, the same methodology will be deployed for cataloguing them in order to have a unique access for remote sensing and in-situ data.