Enhancing Discovery Capabilities of SDIs with Semantic Queries

Mattia Santoro (1), Paolo Mazzetti (1), Cristiano Fugazza (2), Stefano Nativi (1), and Massimo Craglia (2)

(1) CNR - National Research Council of Italy, IMAA {santoro@imaa.cnr.it,paolo.mazzetti,stefano.nativi@cnr.it}, (2) European Commission-Joint Research Center, Ispra {cristiano.fugazza,massimo.craglia@jrc.ec.europa.eu}

In recent years, the World Wide Web has undergone several important changes in terms of available applications, architectures and related technologies. Among these, the emergence of spatial data infrastructures (SDIs) introduced new challenges for the discovery of geospatial data. As a consequence, the need for a more effective resource sharing through the Web raised awareness on the role of semantics in annotation and discovery. This resulted in several efforts that aim at making semantics explicit in resource metadata in order to enable machine-to-machine applications on top of the Web architecture. Moreover, the lack of semantic characterisation of resources in state-of-the-art metadata stimulates the creation of semantics-aware discovery applications that can be transparently applied to existing catalogue services. From a scientific point of view, semantics is necessary to document clearly the theoretical and methodological assumptions underpinning applications in different scientific domains, and develop cross-domain ontologies to facilitate interdisciplinary dialogue and understanding. From the technological point of view, it is necessary to provide a set of special interoperability arrangement in order to develop flexible frameworks that allow to enhance present geo-information sharing solutions with semantics capabilities leveraging the Semantic Web technology. Our focus here is on enhancing discovery capabilities of current Spatial Data Infrastructures (SDIs) with semantics-enabled queries. To achieve this, several approaches are possible:

- Provider-based;

- User-based;

- Third-party.

In this presentation, we discuss a "third-party discovery augmentation approach". The concept behind this approach is to enhance discovery capabilities of SDIs by developing new components that leverage on existing systems and resources in order to automatically enrich available geospatial resource description with semantic meta-information.

This presentation focuses on the Discovery Augmentation Component (DAC), prototyped in the EC-funded Euro-GEOSS (A European approach to GEOSS) project in collaboration with the GENESIS project. This component was specifically developed to address the three thematic areas covered by the EuroGEOSS project: biodiversity, forestry and drought.

The EuroGEOSS DAC federates both multilingual controlled vocabularies providing semantics (i.e. SKOS repositories) and ISO-compliant geospatial catalogue services. The DAC can be queried using common geospatial constraints (i.e. what, where, when, etc.). Two different augmented discovery styles are supported: a) automatic query expansion; b) user-assisted query expansion.

In the first case, the main discovery steps are:

1) The query string (the "what" constraint) is matched against the multilingual textual representations of terms, retrieving language-neutral identifiers for them (e.g. URIs).

2) The set of terms is "expanded" by looking-up related terms in the set of federated semantic services (e.g. controlled vocabularies, gazetteers, etc.).

3) The terms are then translated back to a customisable set of languages and the corresponding geospatial queries are submitted to the federated catalogue services;

4) The DAC performs a "smart" aggregation of query results and sends them back to the client.

In the user-assisted case, the second step becomes:

- The user can browse the federated semantic services and select the set of terms that best reflect her query.

A Graphical User Interface (GUI) has also been developed for testing and interacting with the DAC.
The entire brokering framework has been deployed in the context of EuroGEOSS infrastructure and is being used in two distinct GEOSS AIP-3 use scenarios: The "e-Habitat Use Scenario" for the Biodiversity and Climate Change topic and the "Comprehensive Drought Index Use Scenario" for Water/Drought topic.