Publication and Discovery of Semantically Annotated Geospatial Web Services

Iker Larizgoitia[1], Arturo Beltran[2], Ioan Toma[1], Patrick Maué[3]

Abstract

Environmental information and services have become a crucial asset in the creation of decission support systems. Unfortunately, this information and services are not usually exposed in an interoperable and standard way, limiting their reusability and impact in the community. Publishing and discovering geospatial information and services on the Web is therefore an important challenge in order to create a breeding ground for collaboration and more sophisticated environmental platforms. Based on common standards defined by the Open Geospatial Consortium (OGC) as the starting point to ensure interoperability, we propose a discovery mechanism based on semantic annotations. OGC service descriptions are annotated with SAWSDL and linked to concepts in domain ontologies, following a common semantic service model. We seamlessly integrate the semantics in the standard OGC discovery infrastructure, extending the CSW service catalogues with semantic publication and discovery. Semantics queries can be created based on formal languages like WSML, significantly improving the precission of discovery. In this paper we present our approach, which provides a semantic infrastructure for publication and discovery of environmentally enabled web services.

1. Introduction

Collecting and sharing environmental information has become a top priority for many initiatives around the world [1] [2]. Environmental information can be presented in different fashions, from a simple sensor reading of the water level in a river, to a highly sophisticated rain prediction model. Unfortunately, it is usually provided through monolithic and isolated applications, with manual linkage to the real information and with little sign of reusability. In the search for the evolution towards a common framework for environmental information, not only data has to become available, but also a large number of models need to be exported as services in order to be reused, combined and exploited by the community. In this context, the OGC (Open Geospatial Consortium) [3] is focused on providing a set of standardized interfaces for Web Services linked to any kind of geospatially enabled information. These standards define common schemas and metadata descriptions that enable service providers to describe the capabilities of their services, emphasizing the environmental information involved. These descriptions can be exploited in order to discover potential services that might be needed for a certain scenario or decision making process.

In general, discovery of services based on user requirements has always been a challenge, and considering environmental or geospatial aspects of those services does not facilitate the task. The OGC defines a standard Web architecture for environmental service discovery (more precisely, for geospatially enabled resources), called CSW (Catalogue Service Web) [4]. CSW defines an abstract operation of the discovery process, tightly coupled to a set of standards for metadata descriptions, query capabilities and interaction with the discovery infrastructure. Nevertheless, the expressivity is limited, because it relies only on certain metadata properties referring to generic types, syntax and technical aspects of the service. This makes it difficult to discern the meaning of it, based on the actual information used or produced by the service.

[1] Semantic Technology Institute (STI), Innsbruck, Austria
[2] Institute of New Imaging Technologies (INIT), Universitat Jaume I (UJI), Castellón, Spain
[3] Institute for Geoinformatics (ifgi), University of Münster, Münster, Germany

Our contribution is therefore focused on providing a comprehensive infrastructure to improve the capabilities of discovery mechanisms for environmental services, built on top of standard OGC service representations. Considering that one of the main disadvantages of the service descriptions provided by OGC standards (e.g. SOS, WPS, WCS, WFS [5]) is its weak relation to real concepts, we enriched these Web service descriptions with semantic annotations (in SAWSDL [6]). We also combine the semantic annotations with a light-weight semantic procedure oriented service model (POSM [7]) and formal language to define conditions (WSML [8]). The purpose of this fusion is twofold. On the one hand, users can take advantage of this formal language to create more detailed queries describing the services they want from a semantic point of view. On the other hand, service descriptions are linked to a semantic knowledge model, where intelligent match-making algorithms can be effectively applied. We exploit a technique of logic systems called query containment [9], which defines formally based on the service description and the query from the user, if the service fulfils the conditions defined by the user. Our work is part of the infrastructure defined in the ENVISION project - ENVIronmental Services Infrastructure with Ontologies [10]. The objective of ENVISION is broader than just discovery; the aim is to develop a comprehensive platform to support non ICT-skilled users during the whole process of creating decision support systems based on environmental information and services. Through this platform we expect to boost the cooperation among different users/providers of environmental information and services by providing an infrastructure for composing, executing and discovering environmental services, as well as introducing reusable Web components that can be combined to create decision support portals.

The remainder of this paper tackles the problem of environmental service discovery. Section 2 presents the background concepts and related work in OGC service discovery and semantic annotations, while Section 3 and 4 focus on our approach to service discovery. Section 3 presents the architecture, semantic extensions and workflow of the discovery process. Section 4 describes the implementation of our approach and section 5 ends the paper with some conclusions and future work discussion.

2. Background and Related work

Discovery can be defined as the process of finding resources that fulfil certain conditions specified by a user. Usually hidden in this definition is the need for making the descriptions of the services available, the so-called publication phase, in which service descriptions are stored in a registry-like infrastructure. Based on this set of published services, the conditions to be met are provided and the appropriate matching mechanisms are applied in order to find adequate services. In the following subsections we analyse the current status of discovery mechanisms in the OGC world and we present what semantic annotations are, and why they can improve the discovery of environmental services.

2.1 Service discovery based on OGC standards

The OGC Catalogue Service specification [11] enables different entities to publish information about services (called metadata) and then establishes the procedures to query or retrieve this information. The discovery mechanism is formally defined by (1) the schema of the metadata (how to describe services), (2) the query language (how to define queries on the metadata) and (3) the interface to discovery (how to invoke the process). The OGC Catalogue Service defines these three aspects as follows:

1. In order to promote interoperability, OGC recommends the use of already defined metadata schemas provided by some standardization activities (e.g. ISO 19119 [12], ebXML [13]), which define schemas for geospatial information. Based on these common schemas, there is a predefined set of core properties, on which queries can be created (e.g. subject, title, abstract, format, identifier, bounding box, etc).

2. The specification also defines a Common Query Language (CQL) and provides a minimum set of data types and query operations that every OGC compliant catalogue must implement. CQL predefines support for some filter operations (e.g. boolean, text matching, temporal data types, geospatial operators, etc...) as well as providing a framework for possible extensions.
3. For the invocation interface, a set of abstract operations are defined, in order to support discovery. The publication is defined by a set of operations called transactions (insert, delete, and update) and the discovery covers different query operations based on services which are described as record structures. More details about these operations can be found in the specification [11].

Following these principles, queries can be formulated containing keywords (i.e. any arbitrary text) or spatial filters (i.e. using spatial operators). Even though mask parameters can be used, for a query to find a match in the registered services, the keywords must be exactly the same at a syntactic level. Even if the keywords are the same, it might be the case that they do not really have the same meaning. Moreover, we cannot use synonyms to search for similar things or even use a word with a broader meaning (for example "water") and expect to discover related resources (like "rivers").

Several implementations of the OGC CSW catalogue services can be found on the Web. We analysed some of the most popular implementations under the requirements of the ENVISION project. This analysis can be found here [14].

2.2 Semantic annotations for service discovery

Semantic annotations using ontologies offer a complement to define geospatial services, adding meaning to the service descriptions. As we have to extend already existing schemas (e.g. SOS, WPS, WCS) with semantic annotations, we rely on SAWSDL [9]. SAWSDL provides extension attributes (e.g. "*modelReference*"), which can be applied to various elements in the service descriptions. These attributes point to concepts in an arbitrary ontology having the semantic description of the underlying element. As we are describing services, we also need a semantic service model to be used within these annotations. Several semantic service models are already available (e.g. DAML-S [15], OWL-S, [16]). A recent light-weight alternative for enabling semantic descriptions of existing Web services is WSMO-Lite [17]. The Procedure-Oriented Service Model (POSM) defined as part of it has been adopted and used in our service descriptions. More information about the annotation process can be found here [18]. The result of this annotation process is OGC service descriptions semantically defined based on the POSM service model. These annotations can then be exploited as part of the discovery process. Users can define their conditions using a formal rule language (WSML [19]) and then applying a reasoner for the matching process will assure that the capabilities of a service match the query. Queries defined in WSML are more expressive and relate to the real meaning of the information or operation of the service.

3. Semantic OGC service discovery

One of the aims of our approach was to define the semantic discovery mechanism as an extension of the existing OGC CSW infrastructure so that it could be easily plugged-in to any already existing OGC CSW catalogue implementation. Based on the annotated service descriptions, we define semantic extensions to the query language and provide the necessary architecture to enable publication and discovery of semantically enhanced descriptions.

3.1 Semantic query operator

OGC CSW queries are based on a common query language (CQL), that defines temporal, spatial and keyword based queries, for which the implementations of the catalogues are already optimized. The use of semantic annotations enables the definition of more sophisticated queries using goals. A goal is defined as a WSML logical condition that states what is expected from the service in terms of the service description (e.g. pre-conditions or post-conditions, kind of information provided, type of service).

To maintain the compliance with CQL, we define an additional operator *matchesGoal*, the purpose of which is to check if a resource matches the goal provided in the query, combining this type of queries with the already existing query mechanism in OGC CSW Catalogues. In this way, we can take advantage of the already optimized spatial and temporal query mechanisms while adding support for complementary semantic queries.

3.2 System architecture

In order to support semantic discovery, we seamlessly extend the OGC CSW architecture with two additional components, the Semantic CSW service and the Semantic Reasoner.

The Semantic CSW service provides a wrapped implementation of the CSW interface, performing additional tasks to support semantics when the services are published or discovered. This is connected to an underlying OGC CSW implementation so that the functionality of storing metadata and geospatial query support is reused.

The Semantic Reasoner provides the infrastructure to store the semantic descriptions of the registered services that are going to be used for discovery. It also implements the discovery mechanism that matches the goals with the service descriptions. Using the service model defined in WSMO-Lite/POSM, preconditions and effects of a service can be specified using WSML logical expressions [19]. The semantic annotations, including these preconditions and effects can be seen as logical queries, thus solving the service discovery problem. This translates into solving the query containment problem, which determines whether the results of a query are contained within the results of another query [20].

As part of the ENVISION architecture more components are integrated in order to support discovery from the user point of view. The ENVISION architecture has many different components, but for discovery the Resource Module is the one involved in the discovery, as it enables the user to publish semantically annotated services.

3.2.1 Publication of semantically annotated OGC Services

Semantic OGC service descriptions are based on the Capabilities description of the services, a standard document that each service has to provide describing metadata information about itself.

However, OGC CSW catalogues cannot store those descriptions directly. This is due to the fact that the catalogue, as defined by the standard, has a broader use. It is aimed at storing not only services, but also any kind of geospatial resource. The information that is stored in the catalogue is called metadata and complies with different ISO standards (e.g. ISO/TS 19139:2007 and ISO 19119:2005).

The publication of services in the catalogue is done using these metadata descriptions, so in order to fully integrate our semantic approach, the ISO metadata has to reference the semantically annotated descriptions, while preserving the compatibility with the standard. This is achieved by adding an *onlineResource* element to the service description.

This metadata needs to be created for each service that needs to be registered in the catalogue. Different implementations provide means of translating *Capabilities* documents to ISO metadata, but it is based on the criteria of the implementer and there is no commonly accepted way of doing it. In addition, the semantic annotations would be lost in this translation because they are not mapped to any ISO property. This is a requirement for the publication, but not directly related to the process. Administrators have the liberty to define this metadata in the most convenient way for their services.

The ENVISION project provides a set of components that do the transformations automatically so that the end-user does not need to create this metadata, as it is automatically generated. These components are part of the Resource Management infrastructure, whose main purpose in regards to discovery is to allow users to publish semantically annotated services.

The process of publishing a semantically annotated OGC service is as follows:

1. The Resource Management infrastructure transforms the semantically annotated capabilities document of the service into the correspondent ISO metadata to be published.
2. The Resource Management sends a CSW Transaction Insert request with the ISO metadata to the Semantic CSW service.
3. The Semantic CSW service first redirects the transaction to an underlying OGC CSW implementation using the same message, acting as a proxy.
4. If the publication in the catalogue was successful, it will receive a UUID that identifies the newly added resource.
5. Next, the semantic annotations are extracted following the links in the metadata information. All the models and ontologies are collected and stored in the Semantic Reasonser's knowledge base, so that it can be accessed in the discovery phase for the match-making.
6. The UUID identifying the new record is returned as part of the standard CSW response.

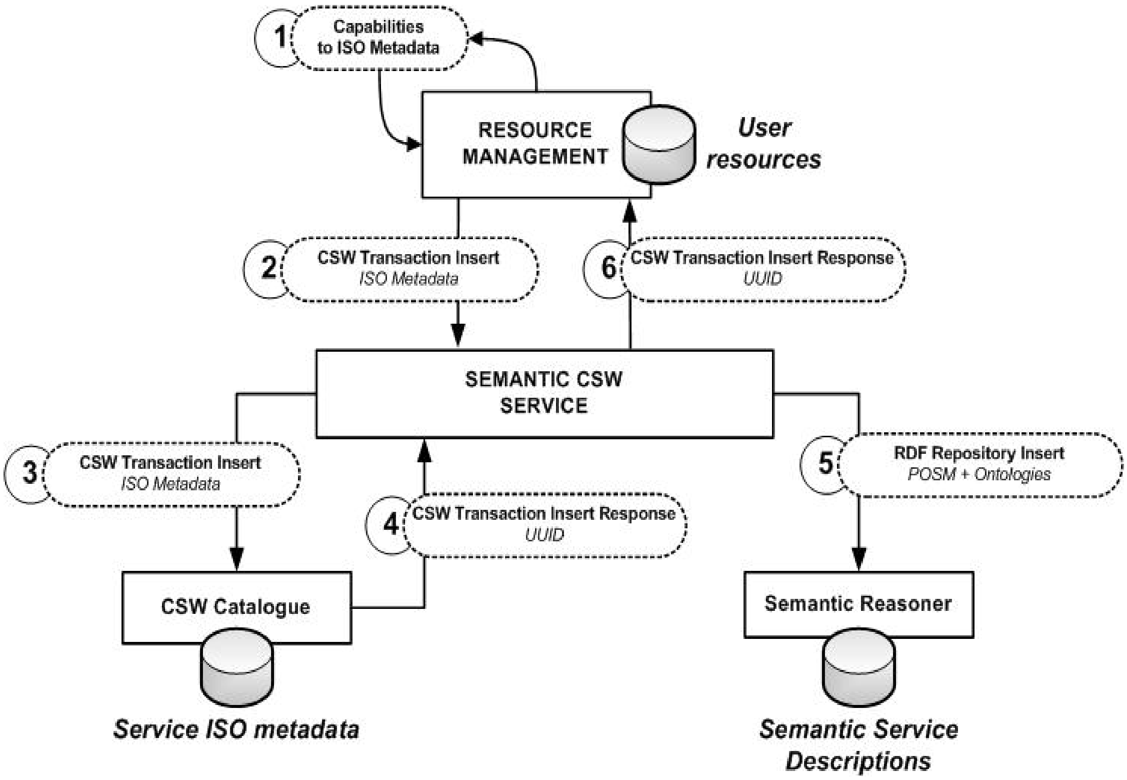

Figure 1
Publication workflow

3.2.2 Semantic discovery of OGC services

Following the standard CSW GetRecords operation the server can be queried about services. The semantic operation *matchesGoal* provides the necessary extension to include the semantic goals. How these goals are created based on the ontologies is not part of the discovery infrastructure. Different implementations of discovery user interfaces might provide varied alternatives.

In the ENVISION platform, we provide different Web components to make it easier to create this queries based on the ontology terms and graphs of conditions, but they are out of the scope of this paper. The workflow of the discovery process is as follows:

1. The Semantic CSW service receives a CSW GetRecords request with an OGC CQL Filter that contains the semantic operator *matchesGoal* and a goal definition in WSML.
2. The request is redirected to the underlying implementation of a standard OGC CSW catalogue.
3. These returned results, are the services that match the non-semantic part of the OGC CQL Filter.
4. With this set of services, the semantic properties have to be checked. The UUIDs are used to retrieve the ontologies and descriptions of the services from the RDF repository.
5. The set of ontologies and descriptions are handed to the Semantic Reasoner, as well as the goal that was included in the filter.
6. The Semantic Reasoner will check which services are a match, based on the objective goal and the semantic descriptions of the services, applying the query containment algorithm.
7. The identifiers of the matching services will be returned.
8. With these identifiers, the discovery service can already prepare the final response with only the matching services and send it back to the caller.

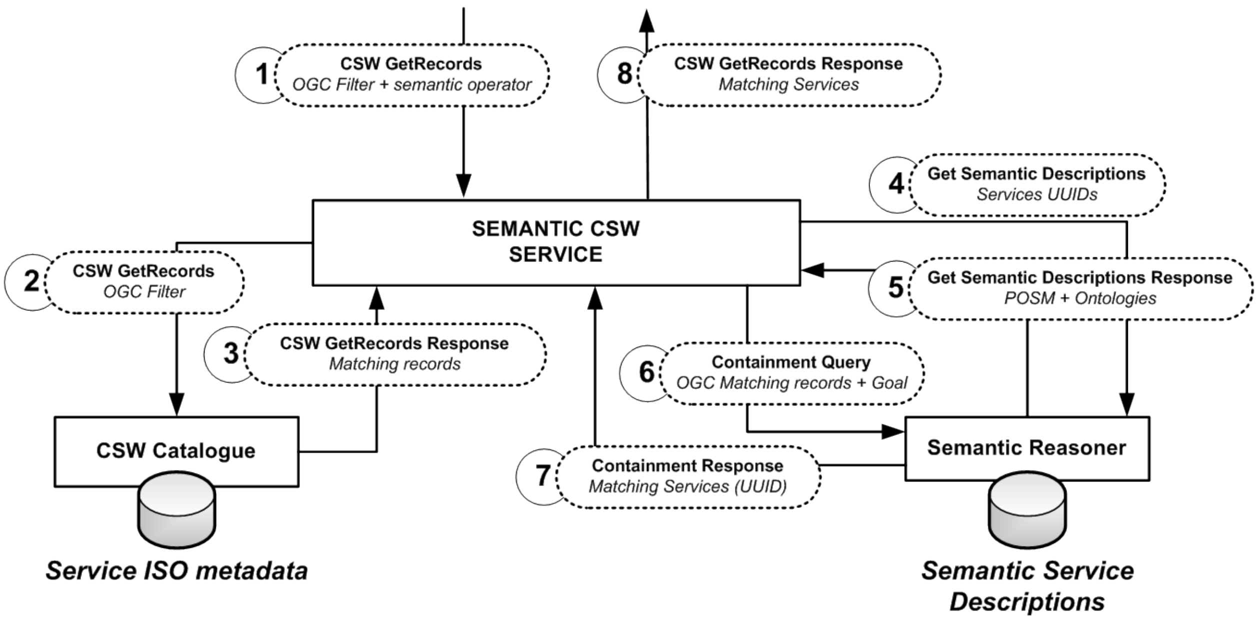

Figure 2
Semantic discovery workflow

4. Implementation

The semantic discovery infrastructure was developed in Java based on open source libraries and platforms. The Semantic CSW service is provided as a CSW HTTP binding implementation using Java Servlets. These can be deployed in any Java application container and linked to any implementation of the OGC CSW catalogue, adding the semantic capabilities. The Semantic Reasoner component is based on the Sesame Triple store (http://www.openrdf.org/) and it is linked to the IRIS Datalog Reasoner (http://www.iris-reasoner.org/) for the query containment evaluation.

The ENVISION platform integrates the discovery infrastructure providing a user interface based on GWT (https://developers.google.com/web-toolkit/) and Java portlets (http://www.jcp.org/en/jsr/detail?id=286). Both the discovery and the platform are provided under an open source license and can be downloaded at: http://kenai.com/projects/envision.

5. Conclusions and future work

In this paper, we have presented an approach to improve environmental and geospatially enabled service discovery based on semantically enhanced descriptions. We acknowledge that current environmental catalogue implementations provide a solid infrastructure for publishing and discovering environmental services based on metadata. However, metadata information is too generic, not flexible enough, language dependent and mostly based on syntactical aspects of the service. By adding semantics to the service descriptions (annotations), services can be discovered querying on the real meaning of the service operation and information. Basing our approach on OGC standards, we ensure interoperability and seamless integration with already existing solutions.

In the future, special attention should be given to a proper evaluation of our approach from the user perspective. The different formalisms presented as part of the discovery process, especially the semantic annotations and the definition of the goals for discovery, might be difficult to define for non-ICT users. As part of the ENVISION platform, we are delivering a full set of UI portlets to facilitate the creation of environmental and geospatial queries that exploit the semantic extensions. Different end-user workshops will be organized by our end-user partners, as part of the acceptance validation of the platform for the main use cases present in the project, such as landslide prediction, oil spill, and river floods detection.

6. Acknowledgment

This work has been partly funded by the European research project ENVISION (FP7 249170), and partly by the Universitat Jaume I predoctoral fellowship (PREDOC/2008/06) and by the Fundació Caixa Castelló-Bancaixa mobility grant (E-2011-12).

Bibliography

[1] European Commission, European Environment Agency, „SEIS - Shared Environmental Information System," [Online]. Available: http://ec.europa.eu/environment/seis.

[2] E. Comission, „INSPIRE - Infrastructure for Spatial Information in the European Community," [Online]. Available: http://inspire.jrc.ec.europa.eu/.

[3] „OGC - Open Geospatial Consortium," [Online]. Available: http://www.opengeospatial.org/.

[4] Open Geospatial Consortium, „OpenGIS Catalogue Service Implementation Specification," [Online]. Available: http://www.opengeospatial.org/standards/cat.

[5] Open Geospatial Consortium, „OGC Schema Repository Updates," [Online]. Available: http://www.opengeospatial.org/standards/ogc_schema_updates.

[6] W3C, „Semantic Annotations for WSDL and XML Schema," [Online]. Available: http://www.w3.org/TR/sawsdl/.

[7] WSMO working group, „POSM - Procedure Oriented Service Model," [Online]. Available: http://www.wsmo.org/ns/posm/0.1/.

[8] WSML Working Group, „WSML - Web Service Modelling Language," [Online]. Available: http://www.wsmo.org/wsml/.

[9] U. Leser. Query Planning in Mediator Based Information Systems. 2000, Technische Universität Berlin.

[10] ENVISION Consortium, „ENVISION - Environmental Services Infrastructre with Ontologies," [Online]. Available: http://www.envision-project.eu/.

[11] Nebert D, Whiteside A, et al. Open Geospatial Consortium Inc. OpenGIS®Catalogue Services Specification. 2007.

[12] Percivall, G. (2002). "ISO 19119 and OGC Service Architecture", Proceedings of FIG XXII International Congress, Washington D.C., USA.

[13] OASIS ebRIM, ebXML Registry Information Model Version 3.0, OASIS Standard (May 2005), available [online]: <http://www.oasis-open.org/committees/download.php/13591/docs.oasis-open.orgregrepv3.0specsregrep-rim-3.0-os.pdf>.

[14] Toma, I., Maué, P. et al. D5.1 Deployment of the open-source OGC Catalogue. ENVISION Deliverable D5.1, 2010

[15] M. H. Burstein, J. R. Hobbs, O. Lassila, D. L. Martin, D. V. McDermott, S. A. Mcllraith, S. Narayanan, M. Paolucci, T. R. Payne, and K. P. Sycara, "Daml-s: Web service description for the semantic web," in International Semantic Web Conference, 2002, pp. 348-363.

[16] "Owl-s: Semantic markup for web services," Available on Internet at: http://www.w3.org/Submission/OWL-S/, 2004.

[17] T. Vitvar, I. Kopecký, I. Viskova, and D. Fensel, "Wsmo-lite annotations for web services," in ESWC, 2008, pp. 674-689.

[18] Mladenić, D., Moraru, A. et al., D4.2 Model annotation components and guidelines. ENVISION Deliverable D4.2. 2011.

[19] de Bruijn, J., and Heymans, S. A semantic framework for language layering in WSML. In Proceedings of the 1st International Conference on Web Reasoning and Rule Systems (RR2007), pages 103-117, Innsbruck, Austria, June 7-8 2007. Springer

[20] R. Ramakrishnan, Y. Sagiv, J.D. Ullman, M.Y. Vardi. Proof-Tree Transformation Theorems and their Applications. in 8th ACM Symposium on Principles of Database Systems. 1989. Philadelphia.