WEB ONTOLOGY SERVICE, A KEY COMPONENT OF A SPATIAL DATA INFRASTRUCTURE

J. Lacasta[1], P.R. Muro-Medrano[1], F.J. Zarazaga-Soria[1], J. Nogueras-Iso[2]
[1]Computer Science and Systems Engineering Department, University of Zaragoza, Spain
{jlacasta, prmuro, javy}@unizar.es
[2] Institute for Environment and Sustainability, Joint Research Centre, European Commission
javier.nogueras@jrc.it

ABSTRACT

In the context of a Spatial Data Infrastructure ontologies are used to denote a formally represented knowledge that is mainly used to improve data sharing and information retrieval. This work describes a web service called Web Ontology Service, based on the OGC Web Service Architecture specification, whose purpose is to facilitate the management and use of ontologies in a Spatial Data Infrastructure. Focusing on the objective of enhancing the classification of resources and the ulterior information retrieval, this work analyzes the potential benefits that this proposed service may provide to the main components that form part of a Spatial Data Infrastructure.

KEYWORDS: Interoperability, Ontology, Spatial Data Infrastructures, Web Technologies

INTRODUCTION

The term ontology is used in information systems and knowledge representation to denote a knowledge model, which represents a particular domain of interest. A body of formally represented knowledge is based on a conceptualization: the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them. And an ontology provides "an explicit formal specification of a shared conceptualization" [Gruber, 1992], i.e. it facilitates a formal notation interpretable by machines that enables a shared and common understanding of a domain.

As far as Geographic Information (GI) and Spatial Data Infrastructures (SDI) are concerned, the research community[1] is aware of the potential benefits of using ontologies as a knowledge representation mechanism. For instance, [Nogueras-Iso et al, 2005] identifies three main areas for the application of ontologies in an SDI. First of all, they can be used for data sharing and systems development. Ontologies help to define the meaning of features contained in geo-spatial data and they can provide a "common basis" for semantic mapping, e.g. to find the similarity between two features that represent the same object but have been defined using different languages. In the same direction, ISO/TC211 proposed the standardization item 19126 [ISO, 2004] to create a data dictionary defining the features and attributes that may be of use to the wider international community. Other works like [Fonseca et al, 2000] even propose the creation of software components from diverse ontologies as a way to share knowledge and data. As a second application area, ontologies facilitate the classification of resources and information retrieval. Metadata (“data about data”) describe unambiguously information resources to enhance information retrieval, but this improvement depends greatly on the quality of metadata content.

[1] See research groups at the Geosemantics Interest Group (*http://www.geosemantics.org*)

One way to enforce the quality of these metadata is the use of a selected terminology for some metadata fields in the form of lexical ontologies, allowing not only to describe the contents but also to reason about them. And thirdly, ontologies also enable the management of metadata schemas. The structure of metadata schemas can be considered as ontologies, where metadata records are instances of these ontologies. These ontologies may be used to profile the metadata needs of a specific geospatial resource and its relationships with the metadata of other related geospatial resources; or to provide interoperability across metadata schemas where transformations of metadata between two different standards could be resolved by systems that observe the commonalities of the two ontologies and automatically detect the metadata element mappings.

Focusing on the abovementioned second area for the application of ontologies (classification of resources and information retrieval), national and international organizations have defined standards [ISO, 2003a; ISO, 2003b; ISO, 2005; FGDC, 1998] that establish the structure of descriptions (metadata) for geographic information, services or locations in a gazetteer. In this context, terms of controlled vocabularies (controlled lists, taxonomies, thesauri) are frequently recommended to harmonize the data and metadata of an SDI and improve the quality of query results (queries on homogeneous sets of descriptions produce better quality results than queries on heterogeneous sets that followed different cataloguing criteria). However, despite the advantages derived from the use of a controlled vocabulary, certain problems of the ambiguity inherent to the language persist. This ambiguity is mainly caused by the different semantic relations between the concepts of a language such as polysemy, homonymy, meronymy, hypernymy or hyponymy. These semantic relations are especially problematic when SDI users try to search data from several sources (and different cataloguing criteria) and their queries do not contain the same terms as the ones used in metadata, queries may be even expressed in a different language from the one used for metadata. Therefore, it becomes crucial to count on lexical ontologies that are able to deal with this ambiguity problems and inter-relate distinct controlled vocabularies.

Traditionally, the first approach in the information community to manage lexical ontologies has been to create different ad-hoc web services that provide their users with the access to a particular ontology. Some examples of this kind of service are the General Multilingual Environmental Thesaurus (GEMET, http://www.eionet.eu.int/GEMET), the FAO Agriculture Vocabulary (AGROVOC, http://www.fao.org/agrovoc/) or the Alexandria Digital Library Feature Type Thesaurus (http://www.alexandria.ucsb.edu/gazetteer/FeatureTypes/ver070302/). The Canadian Geospatial Data Infrastructure project (http://www.geoconnections.org/) advanced in 1999 that an SDI would need a centralised ontology service with the objective of providing a mechanism to maintain thesauri of terms when the number of ontologies to manage would increase. In 2004 they published a prototype of a web service, the Multilingual Geospatial Ontology (M3GO), with some limitation in the relations that it could manage and the ways to identify the ontologies. Another example in the modelling of ontologies and the specification of services is the Simple Organization of Knowledge System (SKOS) project [Miles and Brikley, 2005] that forms part of the W3C Semantic Web Activity. This project has proposed a model to represent lexical ontologies in RDF and has published a prototype of a web service to provide access to their ontologies, whose interface enables basically the retrieval of terms and some types of relations among these terms. This prototype service could be also considered as a centralised service but it does not exploit yet the use of the ontology metadata descriptions proposed in the SKOS model. In summary, although it has been identified that the use of a centralised service is an advance to facilitate the access and management of ontologies in complex information infrastructures, the lack of standardisation in access interfaces and exchange formats has limited its benefits. SKOS intends to unify the interchange format for lexical ontologies but for the ontology service there is no consensus for his interface and functionality. Besides, one of the main drawbacks of current interfaces is that they not provide proper discovery services for ontologies. Although it could be interesting to discover the more appropriate ontology for a specific geographic area or application domain, the present functionality only facilitates the access to an ontology by means of an agreed name.

The objective of this work is to propose and the describe the architecture of new centralized ontology service, called Web Ontology Service (WOS), which enables the uniform management of ontologies (including discovery services) and gives ontology-based support for the components of an SDI. One of the main features of this service is its full integration with the rest of components of a typical SDI, following and extending the standard interfaces used in the geospatial community. The rest of this paper is structured as follows. Next section describes the architecture of WOS and its integration within OGC Web Service Architecture [Lieberman, 2003]. Then, the potential uses of WOS in an SDI are presented. WOS has been initially designed to improve the classification of resources and improve the information retrieval, but with the objective to expand it to the two other described uses in a next step. Finally, this work ends with a conclusions and future lines section.

ARCHITECTURE OF WOS

The Open Geospatial Consortium (OGC) is a non-profit, international, voluntary consensus standards organization that is leading the development of standards for geospatial and location based services. [Whiteside, 2005] specifies the Application Programming Interface (API) that each OGC Web service of an SDI should conform to. In this architecture, every component that forms part of an SDI inherits from a general service whose unique operation is *getCapabilities*. This operation provides a description of the service, its operations, parameters and data types. It is used for the clients to identify if a service provides the needed functionality and how to access to it. Although, OGC has developed numerous specifications for SDI web services, they have not created yet a specification for a service used to manage ontologies. To fulfil this gap, our WOS component has been designed as a component in compliance with the general architecture of OGC with the objective of integrating it with the rest of the OGC services in an SDI.

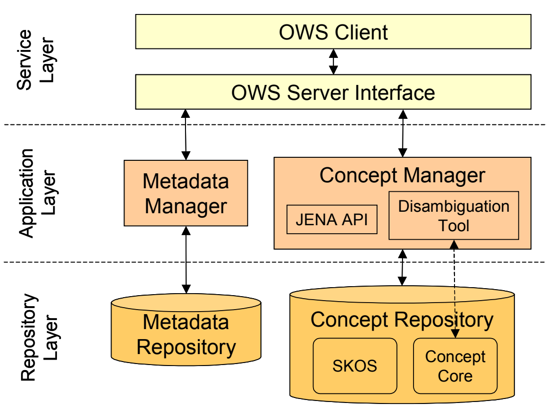

Figure 1: WOS Architecture

Figure 1 shows the architecture of WOS. It is composed of three layers: the repository layer which stores the information of the service, the ontology concepts, their metadata and the concept core used for disambiguation; the application layer which provides access to the concepts of the ontologies, their metadata and is able to disambiguate terms to allow conceptual search; and the service layer that provides access interfaces to the clients.

In the repository layer, we have selected SKOS model for the storage and exchange of ontologies. SKOS is a RDF based model that has been created specifically to manage lexical ontologies for the W3C Semantic Web project. In WOS, the access to the RDF documents storing ontologies is provided in the application layer by Jena (http://jena.sourceforge.net/). Jena is a library that allows the manipulation of RDF documents; it can store the RDFs in text files or in a relational database. One important advantage of using Jena is that it provides an open model that can be extended with specialised modules to access to other ways of storage, as the Jena-Sesame adapter (http://sjadapter.sourceforge.net/) that provides access to Sesame (http://www.openrdf.org/) databases.

Element name	Element Description
TITLE	A name given to the ontology
ALTERNATIVE TITLE	Refinement of title. Any form of the title used as a substitute or alternative (acronym) to the formal title of the ontology
SUBJECT	The topic of the content of the ontology.
CREATOR	A class of entity may be determined by the creator or the publisher or by a third party.
PUBLISHER	Name of the entity that publishes the ontology.
CONTRIBUTOR	An entity responsible for making contributions to the content of the ontology.
DESCRIPTION	An account of the content of the ontology.
DATE	A date associated with an event in the life cycle of the resource, in our system it will contain the publication date of the ontology
TYPE	Used to recognize the type of resource of the associated ontology. In WOS metadata it contains the term ONTOLOGY
FORMAT	Formats in witch is available the ontology for interchange. In our system it will contain SKOS
IDENTIFIER	An unambiguous reference to the ontology within our system
SOURCE	A reference to a resource from which the present ontology is derived. The reference is a unique identifier of the source resource
LANGUAGE	The languages in which the labels of the ontology are. It uses values of the two-letter code of ISO639
RELATION	This element contains the metadata identifiers value of ontologies with similar thematic
COVERAGE	If the ontology terms refer to geographic places, this element will contain the area covered. It will contain a place name obtained from an ontology of places
RIGHTS	Information about rights held in and over the resource (for access and use)
METADATA LANGUAGE	The language of the metadata describing the ontology. It uses values of ISO639.
METADATA IDENTIFIER	The unique identifier that identifies the metadata of the ontology. It is needed for the search system to distinguish between the different metadata it contains.

Table 1: Dublin Core application profile for the description of ontologies

Another important aspect in the repository layer is the description of ontologies. WOS considers metadata of ontologies as fundamental information to be facilitated to the clients. These metadata are stored, in the WOS architecture, in the metadata repository and managed by the metadata manager. The reason for this metadata-driven interface is that centralising the storage of ontologies in our system is not enough to exploit them as full as possible. Ontologies must be described and classified to facilitate the selection of the ontology that fits better the user needs, allowing searches not only by the agreed name given to the ontology but also by the application domain or the associated geographical area. For that purpose, a metadata profile based on Dublin Core [ISO, 2003b] has been created. The reason to use Dublin Core as basis of this profile has been because of its extensive use in the metadata community. It provides a simple way to describe a resource using very general metadata terms, and additionally, it is easy to extend, providing a way to define different application profiles for different kind of resources. Table 1 shows the definition and domains of the elements in this application profile, which as a special feature adds two new elements (metadata language and metadata identifier) to identify appropriately the metadata records describing the ontologies. The purpose of the metadata is not only to provide a human user with information to identify the nature of the ontologies, they can also be used in machine-to-machine communication to perform complex searches to identify ontologies useful for a specific task, as ontologies that cover a restricted geographical area or with a restricted thematic.

Apart from the metadata manager and the Jena API, a novel feature of the application layer of WOS is the integration of a disambiguation tool that enables the interrelation of ontologies with respect to an upper-level ontology, the concept core displayed in Figure 1. This disambiguation method is based on the method presented in [Nogueras-Iso et al, 2005], an unsupervised disambiguation method that uses the hierarchical structure of ontologies as the disambiguation context and applies a heuristic voting algorithm to decide the most adequate sense from the WordNet upper-level ontology [Miller, 1990]. This method is integrated in the WOS architecture to centralise the disambiguation functionality of lexical ontologies and with the objective to extend it for disambiguation of ontology terms in multiple languages, using a multilingual upper-level ontology (e.g., EuroWordnet [Vossen, 1998]).

HTTP_WOS_Interface

createOntology(ontMetadata : Array)
deleteOntology(ontName : String)
importOntology(ontContent : String, ontMetadata : Array)
exportOntology(ontURI : String) : String
getCapabilities(request : OWSGetCapabilities) : OWSServiceMetadata
getOntologyMetadata(ontURI : String) : Array
setOntologyMetadata(ontURI : String, ontMetadata : Array)
getRelatedConcepts(ontURI : String, conceptURI : String, relation : String) : Vector
query(ontURI : String, queryType : String, queryText : String) : ThResponse

Figure 2: WOS API

The API of WOS displayed in Figure 2 has been designed using as basis the ADL Thesaurus protocol [Janée et al, 2003] but it has been extended to give support for several ontologies in multiple languages and to provide compliance with the OGC Web Service Architecture (WSA). The compliance with the OGC WSA is addressed by implementing the *getCapabilities* method that provides the description of the service and its content and whose structure is described in detail in [Whiteside, 2005]. The rest of methods of the API can be classifies in two categories: query and administration. With respect to the query category methods, the *query* and *getRelatedConcepts* methods allow the client navigating by the relations between concepts and searching concepts by label in different languages. The *query* method uses the disambiguation tool of the architecture to expand the results returned providing equivalent terms from the same or different ontologies. As regards the administration category methods, they allow creating a new ontology given its metadata, modifying its metadata, deleting it, and importing or exporting it in SKOS format. Update methods for concepts and relations have not been included in this server interface because the intention of WOS service is to consider each ontology as a whole, managing their changes as different versions of the whole ontology.

USES OF WOS IN AN SDI

The WOS provides users with an enhanced vocabulary related with the theme the user is interested in. The objective is not only to provide the terms in different languages but also their definitions, synonyms, and narrower-broader or related terms from the same or other ontology. A more advanced functionality of the service would be to provide not only simple concepts from a lexical ontology but also complex ones, e.g. spatial reference systems with their parameters about datum, ellipsoid and projection or entities that describe an organization. These complex elements can be very useful to simplify the creation of the content and to provide the user with a detailed description about the element that is using as parameter of a service request.

However, the availability of ontology and terminology alone is not enough. Thus, this section describes how the functionality provided by WOS can be integrated with the rest of components of an SDI and contribute for an improved use of these components. Next subsections describe these benefits in three main categories: facilitate content harmonization, facilitate the construction of service requests, and facilitate the harmonization of content.

Facilitate content creation

The creation of the content of the services of an SDI is a long-term, high-cost process that increases with the number of resources to manage. Therefore, for large systems, it is important to facilitate as much as possible the content creation, providing tools that can help to reduce the time consumed by this task. Moreover, given that human typing errors in the metadata creation can imply not finding a resource using a search system, the control of content quality is even more important. Being homogeneous in the selection of the terms used to describe a resource is another important issue. If two resources have the same characteristics, they should be described with the same terms. Otherwise, if the resources have not been homogeneously described, a query system will only return a subset of the records that it should return. The use of controlled vocabulary for the most relevant elements of metadata can help to reduce the time of creation, the number and impact of human errors and increase the homogeneity. To reuse these vocabularies in different services is essential to manage them uniformly. WOS stores the lexical ontologies used in the SDI and provides the rest of services with the tools to create their contents. Given that each ontology not only contains isolated terms but provide translations to different languages, textual definitions and relations as synonyms, broader, narrower, related terms and similar concepts form other ontologies, WOS supplies content creators with the necessary information not only to avoid typing mistakes but also to help in the selection of the most adequate term to describe a resource. The lexical ontologies served by WOS also help to improve homogeneity in resource descriptions providing, when there are alternatives, the preferred terms to use and restricting in that way the number of terms used to describe the resources and facilitating their posterior location in a query system.

The uses of WOS for content creation in an SDI are multiple. First, every service in an OGC infrastructure has to implement the method *getCapabilities* whose structure is described in [Whiteside, 2005]. Such method provides a description of the service, its methods and parameters, and gives to the client the knowledge needed to access the service content. The returned object is an instance of the abstract *ServiceMetadata* class, allowing in that way that each service can expand the information returned to fulfil their own needs but maintaining a common core. This model contains some elements that have been derived from other standards such as ISO19115 [ISO, 2003a] or GML [Cox et al, 2003]. To create these descriptions, WOS can provide values (in the required language) for some of the elements such as the *keywords* describing the service, the possible *data types* for the parameters of the operations, or even the values for the *ServiceContact* section, which indicates who is the responsible of the service and whose values could be taken from an ontology of entities. Secondly, the OGC catalogue service specification [Nebert and Whiteside, 2004] recommends the use of ISO19119 [ISO, 2005] for the description of services and ISO19115 [ISO, 2003a] for geographic information, but other models like CSDGM [FGDC, 1998] or Dublin Core [ISO, 2003b] can be used too. All these standards define a big number of metadata elements, and many of them must or may contain terms from a controlled vocabulary. Some examples in ISO19115 would be the *descriptive keywords*, the *topic category*, the *distribution format*, the *spatial representation type*, the *reference system* or the different *citations* to related sources. As already mentioned, values for these elements can be facilitated by WOS to the metadata edition tool, reducing in that way the cost of the metadata creation and improving its quality and homogeneity. Thirdly, [Atkinson and Fitzke, 2002] indicates that the model that should follow the content of a Gazetteer is the ISO19112 [ISO, 2003c] standard. This model also contains some elements whose values can be provided by WOS, e.g. the *coordinate system* or the *location type*. And finally, [Vretanos, 2005] indicates that content of Web Feature Service (WFS) or its variants (e.g., Geocoder [Margoulies, 2001]) should follow the GML model. This model contains terms as the location keywords or some of the metadata describing each feature terms that can be provided by WOS.

Facilitate service requests

In many situations, human users who have to access a service find that they do not have a clear understanding about which values they should introduce in a query system or which term they should select from a controlled list. In order to facilitate the access to the services of an SDI, the WOS can provide two types of functionality.

First of all, it can provide selected terminology for the construction of queries in a search service, i.e. providing additional information about the concepts, such as definitions, synonyms or related terms that facilitate the establishment of restrictions on the characteristics (metadata descriptors) of the desired resources. For instance, in a thematic portal about environment the possible keywords that the user should introduce to obtain results are restricted to environmental related terms. This information can be shown in the form of a thesaurus or a code-list, and then, the user can browse the relations of the ontology and select the appropriate term. Given that the ontologies stored in WOS can be multilingual, this system provides a general way for locating resources to persons from different countries, being especially useful for international systems.

The second related functionality that WOS can give to the clients is to provide values and/or descriptions of the parameters used to invoke the services of an SDI. Every service in the OGC architecture implements the method *getCapabilities* that returns a description of the service, its operations and parameters. But it does not give information of the meaning of each one. For example, when one makes a query to an OGC Catalogue Service, one of the parameters to include is the format in which the results will be returned. The formats supported can be obtained thanks to the *getCapabilities* method, but that information may be not enough for the user. If the possible values are XML or HTML, some inexperienced users might not be aware of the characteristics of each format, not knowing which of them to select. Here, given that WOS can contain an ontology of formats, it can give to the clients the complete name of all the possible formats (instead of their acronyms) and their description, thus facilitating the user the task to decide which one to use. In the same way, it is easy to find other examples of this kind of applicability of WOS functionality. For instance, as regards the OGC catalogue service specification, the values of other parameters of the *query* operation (e.g., *resource type* or *result type)* or the *describeRecord* operation (e.g., *type name* and *schema language)* can be described by the WOS. And this is not the only service specification. In the *getFeatureInfo* operation of the Web Map Service (WMS) [Beaujardiere, 2004] parameters like *coordinate reference systems* or *return format* require a complex encoding of possible values. Additionally, other services of an SDI such as the Web Coverage Service (WCS) [Evans, 2003], the Gazetteer or the WFS also contain parameters in their methods whose possible values could be described with the appropriate ontologies served by the WOS.

Facilitate content harmonization

Despite recommending the use of selected terminology to harmonize data and descriptions, content creators from different organizations and application domains will apply their own criteria for the classification of resources, thus generating very diverse terminology, even for the description of similar resources. Moreover, typical users of catalogue services making free text searches will not probably use the same terms as the ones used by creators. All these issues produce an increase of heterogeneity that can affect the performance of information retrieval systems producing low quality results (only a subset of the relevant resources return). In order to reduce this problem, WOS provides a disambiguation tool as a way to harmonize content. This tool uses a concept-based language independent common core (the synset concepts of WordNet [Miller, 1990]) to disambiguate and interrelate the terms of the different ontologies, or even to disambiguate the terms used in user queries.

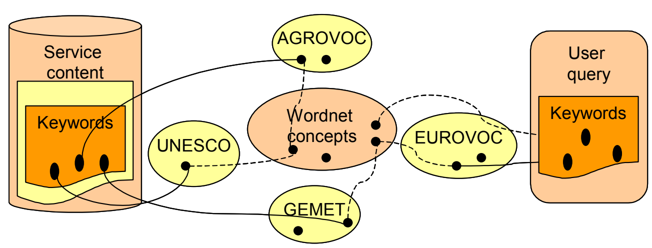

Figure 3: Disambiguation model

This disambiguation technique facilitates the creation of cross-language concept-based retrieval system where both user queries and service content are indexed using a language-independent concept core. That is to say, as depicted in Figure 3, this type of system, described in [Nogueras-Iso et al, 2004], would be able to match the query terms with the semantically equivalent terms in the queried service (e.g., a catalogue or a gazetteer). With this kind of system it is possible to reduce or even eliminate the ambiguity problems caused by semantic relations such as polysemy, homonymy, meronymy, hypernymy or hyponymy. As stated before, the content accessed by the SDI services can be quite heterogeneous. Not only the metadata terms used in a catalogue but also the location types managed by a Gazetteer or a Geocoder may introduce some degree of ambiguity, which may be caused, for instance, by synonym terms (e.g., hospitals, clinics or health centres). Furthermore, the performance of the implementation of other SDI services like Geoparsers [Lansing, 2001] depends absolutely in the ability to disambiguate the context to geoparse and detect the surrounding words that identify a location name (e.g., city-village or river-brook). In summary, this disambiguation functionality of the WOS can facilitate the implementation of other SDI services in order to enhance their query recall (retrieval of relevant content items) and liability.

CONCLUSIONS

This work has presented a Web Ontology Service compliant with the OGC Web Service Architecture specification and whose purpose is to facilitate the management and use of ontologies in a Spatial Data Infrastructure. Designed as a centralized service the architecture of this service aims at reducing the cost of creation of a new ontology, facilitating reusability and avoiding duplicities and inconsistencies. In addition, focusing on the objective of the classification of resources and the improvement information retrieval, this work has analyzed the potential benefits that this service may provide to the main components that form part of a Spatial Data Infrastructure. First of all, the Web Ontology Service can be used to facilitate the creation of multilingual content as it gives access to the concepts of the ontologies, properties, definitions and relations between concepts and other ontologies. Secondly, WOS can also facilitate the construction of service requests since it provides a restricted vocabulary to establish query restrictions for a search system. Additionally, it may give information about the possible values of the parameters to invoke the different SDI services, facilitating this way the labour of understanding which values should be used. And thirdly, WOS can facilitate content harmonization by interrelating ontologies and user query terms with respect to a common and language-independent concept core.

As a continuation of this work, it is planned to submit the specification of this Web Ontology Service as a new OGC Web Service specification that could be integrated in the future with the rest of Web Service specifications already issued by the Open Geospatial Consortium, which at least will provide the required feedback to improve, if necessary, the functionality offered by this service. Additionally, it is expected to explore new possibilities for the applicability of the WOS in the areas of data sharing and reuse, and on the area of metadata schema representation.

ACKNOWLEDGEMENTS

This work has been partially supported by the Spanish Ministry of Education and Science through the project TIC2003-09365-C02-01 from the National Plan for Scientific Research, Development and Technology Innovation. The work of J. Lacasta has been partially supported by a grant (ref. B139/2003) from the Aragon Government.

BIBLIOGRAPHY

Atkinson, R. and Fitzke, J. (eds) (2002). Gazetteer Service Profile of the Web Feature Service Implementation Specification, v0.9. OGC 02-076r3. Open Geospatial Consortium.

Beaujardiere, J. (ed) (2004). Web Map Service, v1.3. OGC 04-024. Open Geospatial Consortium.

Cox, S., Daisey, P., Lake, R., Portele, C. and Whiteside, A. (eds) (2003). OpenGIS Geography Markup Language (GML) Implementation Specification, v3.0. OGC 02-23r4. Open Geospatial Consortium.

Evans, J. (ed) (2003). Web Coverage Service (WCS), v1.0. OGC 03-065r6. Open Geospatial Consortium.

FGDC (1998). Content standard for digital geospatial metadata, Document FGDC-STD-001-1998. Federal Geographic Data Committee (FGDC), Metadata ad hoc Working Group.

Fonseca, F.T., Egenhofer, M.J., Davis, Jr. C.A., and Borges, K.A.V. (2000). Ontologies and knowledge sharing in urban GIS. Computers, Environment and Urban Systems, vol. 24, pp. 251-271.

Gruber, T. (1992). A translation approach to portable ontology specifications. Technical Report KSL 92-71, Knowledge Systems Laboratory, Standford University, Stanford, CA.

ISO (2003a). Geographic information – Metadata, ISO19115:2003. International Organization for Standardization (ISO).

ISO (2003b). Information and documentation - The Dublin Core metadata element set. ISO 15836:2003. International Organization for Standardization (ISO).

ISO (2003c). Geographic information – Spatial referencing by geographic identifiers. ISO19112:2003. International Organization for Standardization (ISO).

ISO (2004). CD 19126 Geographic information - Profiles for feature data dictionary registers and feature catalogue registers, doc. nr. 1561. International Organization for Standardization (ISO), TC 211.

ISO (2005). Geographic information – Services, ISO19119:2005. International Organization for Standardization (ISO).

Janée, G., Ikeda, S. and Hill, L.L. (2003). The ADL Thesaurus Protocol. Alexandria Digital Library Project, http://www.alexandria.ucsb.edu/~gjanee/thesaurus/specification.html.

Lansing, J. (ed) (2001). Geoparser Service Specification, v0.7.1. OGC 01-035. Open Geospatial Consortium.

Margoulies, S. (ed) (2001). Geocoder Service Specification, v0.7.6. OGC 01-026r1. Open Geospatial Consortium.

Miles, A. and Brikley, D. (eds) (2005). SKOS Core Vocabulary Specification W3C Editor's Working Draft, http://www.w3.org/TR/2005/WD-swbp-skos-core-spec-20050510/.

Miller, G.A. (1990). Wordnet: An on-line lexical database. *Int. J. Lexicography,* 3.

Nebert, D. and Whiteside, A. (eds) (2004). OpenGIS Catalogue Services Specification, v2.0. OGC 04-021r2. Open Geospatial Consortium.

Nogueras-Iso, J., Lacasta, Bañares, J.A., Muro-Medrano, P.R. and Zarazaga-Soria, F.J. (2004). Exploiting disambiguated thesauri for information retrieval in metadata catalogs. *Lecture Notes in Artificial Intelligence*, vol. 3040, pp 322-333.

Nogueras-Iso, J., Zarazaga-Soria, F.J. and Muro-Medrano, P.R. (2005). *Geographic Information Metadata for Spatial Data Infrastructures - Resources, Interoperability and Information Retrieval*. Springer Verlag.

Vossen, P. (1998). Introduction to EuroWordNet. *Computers and the Humanities*, vol. 32, no. 2-3 (Special Issue on EuroWordNet), pp. 73-89.

Vretanos, P. (ed) (2005). Web Feature Service Implementation Specification, v1.1.0. OGC 04-094. Open Geospatial Consortium.

Lieberman, J. (ed) (2003). OpenGIS Web Services Architecture, v0.3. OGC 03-025, Open Geospatial Consortium.

Whiteside, A. (ed) (2005). OGC Web Service Common Specification, v1.0. OGC 05-008. Open Geospatial Consortium.

Web Ontology Service, a Key Component of a Spatial Data Infrastructure

Javier Lacasta, Pedro R. Muro-Medrano, F. Javier Zarazaga Soria, Javier Nogueras-Iso
Computer Science and Systems Engineering Department
University of Zaragoza, Zaragoza, Spain
Joint Research Centre - European Commission

The term ontology is used in information systems and knowledge representation to denote a knowledge model, which represents a particular domain of interest. A body of formally represented knowledge is based on a conceptualization: the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them. And an ontology provides "a explicit formal specification of a shared conceptualization" [Gruber, 1992], i.e. it facilitates a formal notation interpretable by machines that enables a shared and common understanding of a domain.

As far as Geographic Information (GI) and Spatial Data Infrastructures (SDI) are concerned, this research community[1] is also aware of the potential benefits of using ontologies as a knowledge representation mechanism, which facilitates knowledge sharing and reuse in interoperable environments. First of all, they are used for data sharing and systems development. Ontologies help to define the meaning of features contained in geo-spatial data and they can provide a "common basis" (as the model proposed in [ISO, 2004]), for semantic mapping (e.g. to find the similarity between two features that represent the same object but have been defined using different languages). Some works [Fonseca, 2000] even propose the creation of software components from diverse ontologies as a way to share knowledge and data. Secondly, they facilitate the classification of resources and information retrieval. Metadata ("data about data") enhance information retrieval because they intend to describe unambiguously information resources. But this improvement depends greatly on the quality of metadata content. One way to enforce the quality of metadata is the use of a selected terminology for some metadata fields in the form of lexical ontologies, allowing not only to describe the contents but also to reason about them. And thirdly, ontologies also enable the management of metadata schemas. The structure of metadata schemas can be considered as ontologies, where metadata records are the instances of those ontologies. Then, ontologies may be used to profile the metadata needs of a specific geospatial resource and its relationships with the metadata of other related geospatial resources; or to provide interoperability across metadata schemas where transformations of metadata between two different standards could be resolved by systems that observe the commonalities of the two ontologies and automatically detect the metadata element mappings.

Given the importance of ontologies in the SDI context, this paper proposes the inclusion in an SDI of a specific component called Web Ontology Service (WOS), which enables the management of ontologies and gives ontology-based support for the rest of components of an SDI. In particular, this WOS component has been especially designed to facilitate the second of the above mentioned uses of ontologies, i.e. the classification of resources and information retrieval.

National and international organizations have defined standards that establish the structure of data descriptions (metadata) [ISO, 2003a; ISO, 2003b; FGDC, 1998]. Those descriptions are not

[1] See research groups at the Geosemantics Interest Group (*http://www.geosemantics.org*)

only used to describe geographical information, but also to describe other elements that integrate the SDI such as services or locations in a gazetteer. In this context, terms from a controlled vocabulary (controlled lists, taxonomies, thesaurus) are frequently used to harmonize the data and metadata of a SDI, because queries on homogeneous sets of descriptions produce better quality results than queries on heterogeneous sets in which each record has been classified following different criteria. However, despite the advantages derived from the use of a controlled vocabulary, certain problems of the ambiguity inherent to the language persist. This ambiguity is mainly caused by the different semantic relations between the concepts of a language such as polysemy, homonymy, meronymy, hypernym or hyponymy. These semantic relations are especially problematic when SDI users try to search data from several sources (and different cataloguing criteria) and their queries do not contain the same terms as the ones used in metadata, queries may be even expressed in a different language from the one used for metadata. Therefore, it becomes crucial to count on lexical ontologies that are able to deal with this ambiguity problems and inter-relate distinct controlled vocabularies. The objective of WOS will be to manage in an appropriate way these lexical ontologies that improve the quality of metadata. It is essential to compile the knowledge and the experience of their creators and to manage them uniformly, reusing and improving them when necessary.

The WOS component has been designed as a component in compliance with the general architecture of the Open Geospatial Consortium (OGC). [Vretanos, 2003] specifies the Application Programming Interface (API) that each OGC Web service should conform to. Thus, as the WOS component complies with this API, it can be easily integrated with the rest of OGC web services. Figure 1 shows the architecture of WOS, which is composed of three layers: the service layer that provides the access to the clients; the application layer which provides access to the concepts of the ontologies, their metadata and it provides utilities of disambiguation to allow semantic search; and the repository layer which stores the information of the service.

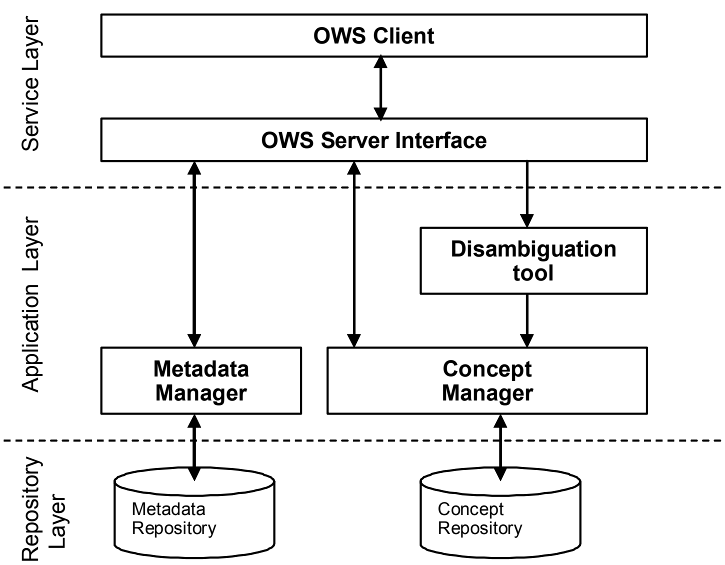

Figure 1: WOS Architecture

The API of WOS is displayed in Figure 2. It must be remarked that metadata plays an important role in this API and it is used as input parameters for most of the methods. The reason for this metadata-driven interface is that simply storing the ontologies in our system is not enough to allow its use in a proper way. Quite the opposite, ontologies must be described and classified in different languages to facilitate the selection of the ontology that fits better with the user needs. A Dublin Core metadata profile has been created used for this purpose. The methods offered within this API can be classified in the following categories:

Queries. First of all, the compliance with the OGC Web Services Architecture is addressed by implementing the getCapabilities method, provides the description of the service and its content. And secondly, the query methods allow the client navigating by the relations between concepts, searching concepts by label in different languages and using the disambiguation tool [Nogueras, 2004] to expand the results returned.

Administration. On one hand, the API provides methods for Ontology administration. It is possible to create a new ontology given its metadata in multiple languages, delete an ontology, modify the metadata describing an specific ontology, and exchange ontologies using the SKOS format [Miles, 2005]. And on the other hand, there are also methods for administration of concepts. It is possible to create, update and delete concepts, attributes and relations between concepts.

HTTP_WOS_Interface
createOntology(ontMetadata : Array) deleteOntology(ontName : String) importOntology(ontContent : String, ontMetadata : Array) exportOntology(ontURI : String) : String getCapabilities(request : OWSGetCapabilities) : OWSServiceMetadata getOntologyMetadata(ontURI : String) : Array setOntologyMetadata(ontURI : String, ontMetadata : Array) getRelatedConcepts(ontURI : String, conceptURI : String, relation : String) : Vector query(ontURI : String, queryType : String, queryText : String) : ThResponse

Figure 2: WOS API

Finally, it is worth mentioning the benefits the use of WOS will provide to rest of services of an SDI. The main areas where WOS will facilitate added-value functionality are the following:

Service description creation: Every OGC service has to implement the method getCapabilites, which provides a description of the service. To create this descriptions the WOS can provide concepts to use in the metadata in different languages from certain ontologies, related concepts as synonyms or narrower concepts, textual definitions to help the cataloguer to decide between similar concepts, and also it can suggest concepts from others ontologies.

Content creation: Services as geographic metadata catalogue, service catalogue [Nebert, 2004], gazetteer [Atkinson, 2001] or geocoder [Margoulies, 2001] store, between other elements, geographic data descriptions, services… The WOS can provide terms, relation and definitions in a similar way as in service description creation. It can also provide help to the data creator to describe the features in web map server [Beaujardiere, 2004], the feature description in a web feature service [Vretanos, 2002] or the coverages in a Web Coverage Service [Evans, 2003].

Query results improvement: The WOS can be used as the disambiguation base of a conceptual retrieval system for the metadata contained in the geographic metadata catalogue, service catalogue, gazetteer or geocoder. The advantage of using a conceptual retrieval system is that the user can use his own terms to define his query and the system, using the existent ontologies, is able to match this query with the metadata in the catalogue, although the terms were different. Other OGC services which could use the WOS in this same way would be the geoparser [Lansing, 2001] to disambiguate the context (e.g. city-village, river-brook) of the analyzed geographic terms, with the objective to identify with a higher liability the place which is being referred in the stored document, or the goeolocator in which the types of the stored elements can be named in different ways (hospitals, clinics, health centres…).

Content homogenization: The WOS can provide to the service catalogue the ability to eliminate inconsistencies between the descriptions stored in this service about the rest of the components of the SDI and the descriptions returned by the getCapabilities methods of those services.

BIBLIOGRAPHY

Atkinson, R., Gazetteer Service Draft Candidate Implementation Specification 0.84. OGC 01-036. Open Geospatial Consortium. March 01.
Beaujardiere, J., Web Map Service 1.3. OGC 04-024. Open Geospatial Consortium. August 2004.
Evans, J. Web Coverage Service 1.0. OGC 03-065r6. Open Geospatial Consortium. August 2003.
Federal Geographic Data Committee (FGDC). Content standard for digital geospatial metadata, Document FGDC-STD-001-1998. Metadata ad hoc Working Group. 1998.
Fonseca FT, Egenhofer MJ, Davis Jr. CA, Borges KAV. Ontologies and knowledge sharing in urban GIS. Computers, Environment and Urban Systems, 24:251-271. 2000
Gruber T. A translation approach to portable ontology specifications. Technical Report KSL 92-71, Knowledge Systems Laboratory, Standford University, Stanford, CA. 1992.
International Organization for Standardization (ISO). Geographic information – Metadata, ISO19115:2003. 2003a.
International Organization for Standardization (ISO). Information and documentation - The Dublin Core metadata element set. ISO 15836:2003. 2003b.
International Organization for Standardization (ISO). Geographic information - Profiles for feature data dictionary registers and feature catalogue registers. Doc. nr. 1561. ISO/TC 211. 2004.
Lansing, J., Geoparser Service Specification 0.71. OGC- 01-035. Open Geospatial Consortium. March 2001.
Margoulies, S., Geocoder Service Specification 0.76. OGC 01-026r1. Open Geospatial Consortium. March 2001.
Miles, A., Brikley, D., SKOS Core Vocabulary Specification (http://www.w3.org/2004/02/skos/core/spec/2005-03-24). W3C Editor's Working Draft. March 2005.
Nebert, D., Whiteside, A., Open Gis Catalogue Service Specification 2.0. OGC 04-021r2. Open Geospatial Consortium. May 2004.
Nogueras, J., Lacasta, J. Bañares, J., Muro, P., Zarazaga, J., Exploiting disambiguated thesauri for information retrieval in metadata catalogs. Lecture Notes in Artificial Intelligence Volume 3040. March 2004.
Vretanos, P., Web Feature Service Implementation Specification 1.0. OGC 02-058. Open Geospatial Consortium. May 2002
Vretanos, P., Open Gis Web Service Architecture 0.3. OGC 02-058. Open Geospatial Consortium. January 2003.

Institute for Environment and Sustainability

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Abstracts

ESDI: Setting the Framework

11th EC-GI&GIS Workshop

Alghero, Sardinia
29 June-1 July 2005

www.ec-gis.org/Workshops/11ecgis

INSPIRE · Infrastructure for Spatial Information in Europe

REGIONE AUTONOMA DELLA SARDEGNA

Table of Contents

ABSTRACTS FOR PARALLEL SESSIONS

SESSION: INSPIRE FACTS 1

INSPIRE - STATE OF PLAY STUDY: STATUS OF THE NATIONAL SPATIAL DATA INFRASTRUCTURES IN EUROPE 2
D. Vandenbroucke, K. Janssen, J. Van Orshoven

SDIGER: A CROSS-BORDER INTER-ADMINISTRATION SDI TO SUPPORT WFD INFORMATION ACCESS FOR ADOUR-GARONNE AND EBRO RIVER BASINS 5
M.A.Latre, F.J.Zarazaga-Soria, J.Nogueras-Iso, R. Béjar, P.R.Muro-Medrano

INSPIRE AND THE PSI DIRECTIVE: PUBLIC TASK VERSUS COMMERCIAL ACTIVITIES? 8
K. Janssen

INSPIRE AND E-GOVERNMENT 10
Eva Pauknerová

SESSION: TECHNICAL RESEARCH ISSUES 13

APPROACHES TO SOLVE SCHEMA HETEROGENEITY AT THE EUROPEAN LEVEL 14
Anders Friis-Christensen, Sven Schade, Stephen Peedell

RESEARCH ISSUES IN CONSTRUCTING GEOGRAPHIC ONTOLOGIES FOR ENVIRONMENTAL DATA DISCOVERY AND EXPLOITATION 17
G. G. Wilkinson and D. Cobham

WEB ONTOLOGY SERVICE, A KEY COMPONENT OF A SPATIAL DATA INFRASTRUCTURE 19
Javier Lacasta, Pedro R. Muro-Medrano, F. Javier Zarazaga Soria, Javier Nogueras-Iso

DEVELOPING AN SDI FOR TIME-VARIANT AND MULTI-LINGUAL INFORMATION DISSEMINATION AND DATA DISTRIBUTION 23
Nicole Ostländer, Sascha Tegtmeyer, Theodor Foerster

SESSION: THEMATIC SDI 27

DELIVERING GEOSCIENTIFIC INFORMATION AND PRODUCING NEW SERVICES BASED ON STANDARD PROTOCOLS 28
F. Robida, J.J.Serrano

WORLD METEOROLOGICAL ORGANISATION OPERATIONAL METEOROLOGY 29
G. H. Ross, A. Rubli, A. Broad

REFERENCE DATA IN THE INTERNET – IMPLEMENTATION OF SDI-SERVICES AS PART OF E-GOVERNMENT 30
Heinz Brüggemann, Jens Riecken

THE ENVIRONMENTAL INFORMATION SYSTEMS *UDK*, *GEIN*®, AND *PORTAL-U* AS PART OF THE NATIONAL GERMAN SDI 31
T. Vögele, M. Klenke, F. Kruse

EUROGEONAMES – INTEGRATION OF GEOGRAPHICAL NAMES DATA IN A EUROPEAN SPATIAL DATA INFRASTRUCTURE (ESDI) 34
P.-G. Zaccheddu, Dr. J. Sievers

SESSION: SDI TECHNICAL DEVELOPMENTS 37

DEVELOPMENT OF THE KNMI OPERATIONAL DATA CENTER (KODAC) 38
Wim Som de Cerff, Frans van der Wel, John van de Vegte, Ian van der Neut and Maarten van der Hoeven

ADAPTATION METHOD OF STRATIGRAPHY DATA TO INSPIRE STANDARDS 42
J. Chełmiński, M. Rossa

EUROPEAN SUSTAINABLE DEVELOPMENT RELATED POLICIES AND LEGISLATION, INSPIRE AND GEOSCIENTIFIC DATA 43
P. Christmann, K. Asch, Rafaelle Pignone, Iain Jackson, F. Robida, P. Ryghaug, R. Tomas, L. Persson

ISSUES OF MULTILINGUALITY IN CREATING A EUROPEAN SDI – THE PERSPECTIVE FOR SPATIAL DATA INTEROPERABILITY 47
Joanna Nowak, Javier Nogueras Iso, Stephen Peedell

SESSION: NATIONAL/REGIONAL SDI 1 49

ORGANIZATIONAL TOPICS FOR THE CREATION OF AN ESDI FRAMEWORK 50
Bas C. Kok,
NSDI CROATIA – THE ROADMAP 52
A. Wytzisk, A. Remke, Z. Bačić
IDEZAR: AN EXAMPLE OF USER NEEDS, TECHNOLOGICAL ASPECTS AND THE INSTITUTIONAL FRAMEWORK OF A LOCAL SDI 56
D. Portolés-Rodríguez, P. Álvarez, R.Béjar, P.R. Muro-Medrano
SIGMATER: A PROJECT TO CREATE AN INFRASTRUCTURE FOR EXCHANGING AND INTEGRATING REGIONAL CADASTRAL INFORMATION. 59
Giovanni Ciardi, Nicola Cracchi Bianchi, Luigi Zanella

SESSION: DATA QUALITY AND EXCHANGE 61

DATA QUALITY AND SCALE IN CONTEXT OF DATA HARMONISATION 62
Katalin Tóth, Vanda de Lima,
DATA EXCHANGE AND INTEROPERABILITY IN SUPPORT OF THE IMPLEMENTATION OF THE COMMON AGRICULTURE POLICY 65
Armin Burger, Paul Hasenohr
A STANDARDISED GEO-IDENTIFIER IN THE CONTEXT OF GEO-TRACEABILITY AND COMMON AGRICULTURAL POLICY 68
D. Buffet, R. Oger
A CENTRALIZED SPATIAL DATABASE FOR ACCESSING NATURA2000 DATA, OVERVIEW OF DESIGN AND CURRENT STATUS 70
Tomas De Leus, Petra Michiels, Jan De Belder, Danny Vandenbroucke

SESSION: NATIONAL/REGIONAL SDI 2 73

REBUILDING A SDI – THE PORTUGUESE EXPERIENCE 74
R. P. Julião
COORDINATION OF THE NATIONAL SDI IN GERMANY 75
Martin Lenk
THE GEOINFORMATION INFRASTRUCTURE IN THE CZECH REPUBLIC: THE KEY ROLE OF METADATA 78
B. Horakova, P.Kubicek, J.Horak,
ONE SCOTLAND – ONE GEOGRAPHY: A SMALL COUNTRY WITH BIG IDEAS 80
Cameron Easton
DIGITAL SOUTH-EAST EUROPE – A REGIONAL DISTRIBUTED GIS AND GEO-PORTAL 81
Ulrich Boes

SESSION: COMPONENTS AND STRUCTURES 85

OPEN SOURCE COMPONENTS TO BUILD A GEOPORTAL 86
M.A. Manso, M.A. Bernabé
REENGINEERING THE GEOPORTAL APPLYING HCI AND GEOVISUALIZATION DISCIPLINES 88
T. Aditya, MJ. Kraak
MULTI-SOURCE FRAMEWORK FOR SEAMLESSLY EXPLOITING AND LEVERAGING DISPARATE SPATIAL DATA CATALOGUES 91
Oscar Cantán , F. Javier Zarazaga-Soria, Javier Nogueras-Iso
A HUB & SPOKE MODEL FOR SPATIAL INFRASTRUCTURE, USING SPATIAL DATA WAREHOUSES 96
Eamon G. Walsh
IDENTIFYING INFRASTRUCTURE COMPONENTS – FUNDAMENTAL DATA SETS AND SERVICES 97
Morten Lind,