ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

Semantic DESCaaS - Extending the Description as a Service Concept to Enable Semantic Annotations

Henry Michels and Marcell Roth
University of Münster (WWU) / Institute for Geoinformatics (ifgi)
Weseler Straße 253, D-48151
Münster, Germany
{henry.michels, marcell.roth}@uni-muenster.de

Arturo Beltran
Universitat Jaume I (UJI) / Institute of New Imaging Technologies (INIT)
Av. de Vicent Sos Baynat s/n, 12071
Castellón, Spain
arturo.beltran@uji.es

Abstract

Semantic interoperability covers the conflict-free and meaningful exchange of resources by improving the mutual understanding of participants in a communication process. Especially, the communication between humans via machines is fraught with misunderstandings including machine-machine communication processes. The focus is on an improved support of the human participants by enabling intelligent and independent behaviour of the machines. The realization of semantic interoperability inheres two main tasks in practice. On the one hand, there is a substantial need of unambiguous vocabularies corresponding to the purpose of a communication process. On the other hand, the suitable vocabularies have to be used by all participants. In this paper, we concentrate on the second issue. We present a well-tried strategy and recent technical solutions enabling the annotation of Web services with appropriate knowledge representations. We will draw the current limits of this approach with respect to certain kinds of resources and come up with a conceptual and partly technical solution to semantically enhance any type of resource.

Keywords: Semantic Annotation, Semantic Interoperability, Resource Description, Web service

1 Introduction

The exchange of resources[1] between humans is more and more undertaken by machines replacing former face to face processes. The evolution of the Web is one of the driving forces illustrating nicely the increasing use of machines for the exchange of information. This inheres human-machine, machine-machine, and machine-human communication. Successful communication processes require interoperability between the participating parties. Interoperability addresses the requirements of the members of a communication process twofold. (1) Syntactic interoperability refers to the capabilities of two or more systems to communicate and exchange resources between them and is mostly resolved by metadata and standards [9]. (2) Semantic interoperability is about the avoidance of semantic conflicts appearing during communication processes [8, 10]. With respect to the semiotic triangle [15], communication processes are never clean of misunderstandings. Above all, indirect communication is prone to misconceptions between the attendees [17].

Discovery is a vital part of resource exchanges and a good example for semantic conflicts in communication processes. Traditionally, Web search engines and catalogue services are used to retrieve desired information. Mostly, they are able to discover resources from different sources based on (free-text) keywords specified by the user. Nevertheless, advanced implementations include controlled vocabularies and spatio-temporal queries. Metadata descriptions are the key requirement to enable effective discovery in distributed environments. Usually, provided resources lack of sufficient descriptions, i.e. metadata. If metadata are in place, they are mostly offered as unstructured text or as semi-structured descriptions based on textual information. Only in some cases, metadata contain also values from code lists or controlled vocabularies or provide spatio-temporal information. Discovery via text matching methods limits the amount of appropriate search results due to the semantic conflicts, e.g. misunderstandings appearing in multilingual communication processes or semantic heterogeneities, between an information provider and an information consumer caused by the lack of semantic descriptions.

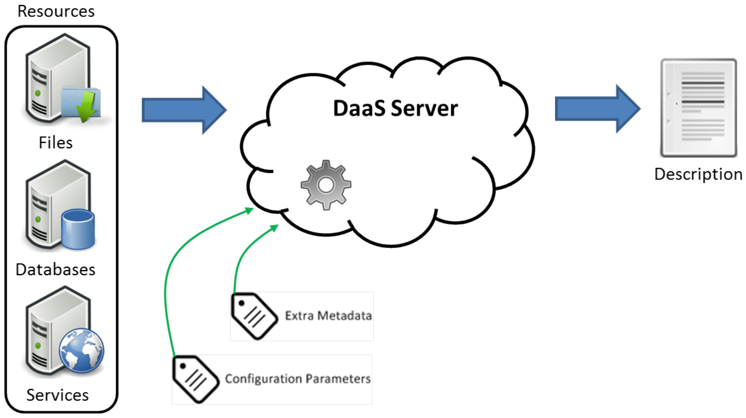
In general, two main issues occur when an exchange of resources has to be interoperable on the semantic level. On the one hand, knowledge representations appropriate to a specific purpose have to exist. On the other hand, such knowledge representations have to be attached either to the exchanged resource or to the participants of the communication process[2]. In this paper, we cover the second issue and highlight problems of this domain.

The next section gives some background information about

[1] A resource can be a physical as well as a virtual good.

[2] Directional communication processes require only the attachment of the vendor with semantic descriptions.

Figure 1: The general view on the DESCaaS paradigm (from [2]).



concepts, strategies, and technologies being vital regarding the approach. Section 3 introduces a concept to enable semantic annotations to any type of resource by providing automatic creation of metadata. The concept is a combination of the presented strategies and technologies developed and applied in the ENVISION project [12]. The presentation of the conceptual approach is followed by a concrete example demonstrating the intensive need of semantic descriptions for binary resources provided as simple downloads on Web-sites. Existing prototypes are described to prove the feasibility of the concept. Finally, we conclude our work and give some insights into future tasks.

2 Background

2.1 Description as a Service

The evolution of the Web raises various challenges which have to be met. The growing volume of resource cries for improvements and innovations in the data management field [3], for example. Within this research track a new paradigm called "DESCription as a Service" (DESCaaS) is proposed [2]. DESCaaS aims to extend the architecture of information systems by adding resource description functionality. The objective is to provide uniform descriptions of resources and their content to improve data accessibility, interoperability, and discovery. By offering descriptions, DESCaaS is covering the lack of mechanisms that allow users to publish, find, and access distributed resources efficiently.

We envision its core application in the area of distributed information systems and data infrastructures. The abstract and general definition of the DESCaaS paradigm extends the field of application to many use cases. Figure 1 presents a general view on the DESCaaS paradigm. A DESCaaS compliant service receives as a basic input the resource itself. Additional optional inputs like extra metadata or some configuration parameters, e.g. level of description or type of metadata extraction, can be considered by the user. The need of additional inputs and the way of providing them is always dependent on a specific implementation of the DESCaaS. The paradigm supposes a huge improvement to obtain resource descriptions by enabling the publication and re-use of metadata generation processes. A given resource is analysed and metadata is generated by the process implemented in the service to build the resource description according to the requested output format. A prototypical implementation is briefly explained in section 4.

2.2 Semantic Annotations

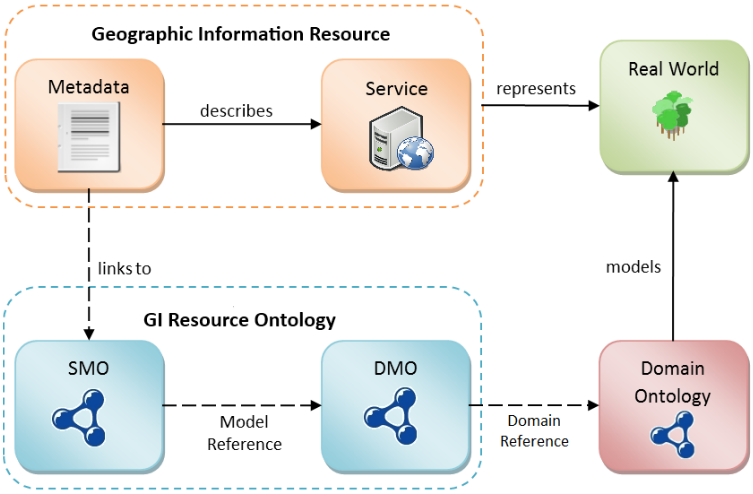
Semantic annotations establish links between Information Technology (IT) resources and explicit vocabularies explaining what the resource represents in the real world. Semantically annotated resources improve information retrieval, discovery, or the validation of requested resources to mention only some benefits [13]. Annotations can be applied versatile (on different levels) depending on the type of resource and its utilization. In [13], three valid annotation levels (service metadata, data model, data entities) are presented regarding the annotation of Web services specified by the Open Geospatial Consortium (OGC). Satisfying results are obtained only when performing the annotation on level two, i.e. the data model level. Going away from OGC Web services, this approach seems to be stable. Annotating directly resource entities (data entities in [13]) (level 3) comes with two disadvantages. On the one hand, it produces a mess of overhead which requires storage and reduces performance. On the other hand, some types of resources, for example, unstructured or binary ones are hard or even impossible to annotate. The annotation of a service holding resources (level 1) is efficient regarding performance. The increase of performance reduces the flexibility of the annotations which becomes an issue if the service provides resources of different types and content. As a consequence, annotating the descriptions of resources is the most appropriate solution in two aspects. (1) This strategy is independent of the type of the raw resource. (2) Metadata are mostly based on a principle that allows machines to process them automatically, i.e. they are semi-structured.

2.3 Service Models and Data Models

Next to the thematic aspect of resources, formal aspects are interesting during discovery as well to reach syntactic interoperability. Following the categorization of ontologies proposed by Guarino [7], application ontologies are an appropriate instrument to describe the model of resources. We refine and adapt the strat-

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

Figure 2: A general overview about the annotation strategy proposed in ENVISION [14].



egy of the ENVISION project by using non-shared application vocabularies working as a proxy between shared domain ontologies and the convenient resources. A Geographic Information (GI) resource represents a data resource of one particular type served by a Web service[3]. It is generated from the service metadata and the service itself offering access to the resource. Each type of resource is remodelled by a GI resource ontology (application ontology). This ontology consists of a service model ontology (SMO) and a data model ontology (DMO). The SMO describes common capabilities of the Web service, i.e. the different service operations and the corresponding inputs and outputs. Inputs and outputs are further specified by the DMO representing the structure of the resource, i.e. the schema. Figure 2 illustrates how Web services are annotated in the ENVISION project. The GI resource is linked to its ontological representation by adding a reference in the service metadata to the SMO. The link between a SMO and a DMO is called model reference. As described in [11], the semantic annotation is an extension which links elements of GI resource ontologies with appropriate concepts of domain ontologies (which are ideally grounded in top-level ontologies). This connection is called domain reference and is provided through rules that are part of the DMO.

3 Enable Semantic Annotations for Resources

A strategy to automatically obtain metadata describing different types of resources is satisfactory to several use cases. It improves the appropriate usage of resources as well as their discovery. However, using only non-semantic metadata as resource descriptions is limited. As argued in the introduction, conversations via machines evoke several semantic conflicts ending in misunderstandings. Non-semantic metadata are not sufficient to overcome these issues. This section presents how the introduced DESCaaS conception can be extended to enable semantic annotations for resources independent of their type.

The semantically enhanced DESCaaS (SemDESCaaS) borrows the core concept from the original DESCaaS. SemDESCaaS consists of a DESCaaS service and a Resource Model Translator (RMT). The former provides uniform descriptions of resources. The RMT is an Java application programming interface (API) and extends the Service Model Translator (SMT) [14] which is further described in Section 4. It translates resource descriptions provided by the DESCaaS into a GI resource ontology and a Web Service Description Language (WSDL) [5] document[4]. The automatic creation of WSDL descriptions allows the integration of simple files[5] into Web service compositions based on a workflow language like the Business Process Execution Language (BPEL) [1]. Next to WSDL files, the GI resource ontology is the required document to enhance the resource with a semantic description. As afore-mentioned, a service model and a data model ontology are integrated in such a GI resource ontology. Both represent all necessary parts of the service capabilities and the resource. The SMO includes functional aspects explaining how to access and, if needed, execute the resource. It has references to the DMO which contains a description of the served resource including the inner structure as well as quality information. The GI resource ontology, respectively the DMO, is linked to domain ontologies via logical rules. Such a domain reference further specifies the semantics of the entire resource. It is important to mention that we propose a conceptual approach of how

[3] We follow the definition of the W3C available at http://www.w3.org/TR/ws-gloss/ .

[4] WSDL is an international accepted standard of the World Wide Web Consortium (W3C) to describe Web services.

[5] Here, we refer to resources which a not provided by Web service, i.e. resources just downloadable via Web sites.

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

Figure 3: The general view on the SemDESCaaS paradigm.

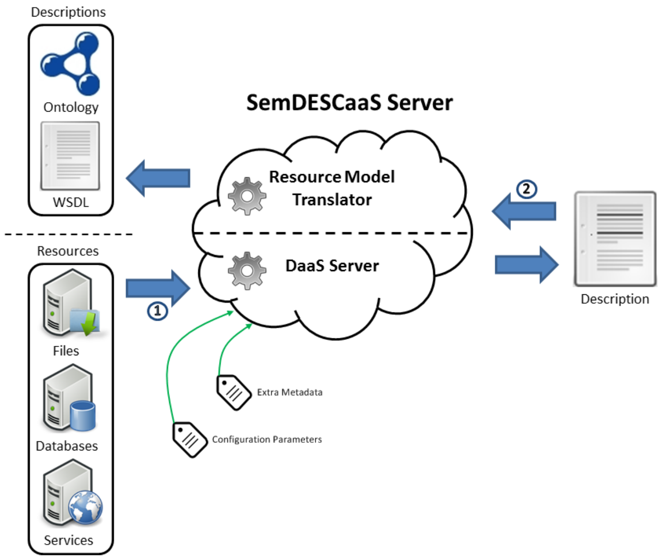


Figure 4: Example URLs to access the various resource descriptions.

```
(1) http://sem-descaas.com/execute/description?url=<url of the resource>
(2) http://sem-descaas.com/execute/giro?url=<url of the resource>
(3) http://sem-descaas.com/execute/wsdl?url=<url of the resource>
```

to semantically annotate the resources. The implementation does not automatically inject references into the GI resource ontologies pointing to domain ontologies. We expect that domain ontologies, which are supposed to capture knowledge specific to a certain domain, already exist. In section 5, we introduce opportunities and current issues of the creation and injection of domain references.

Implementations of the SemDESCaaS concept are supposed to be Web services to offer a flexible and open architecture which is independent from platforms and programming languages. Corresponding to [4], the type of Web service can be REST-compliant[6] or arbitrary. Figure 3 illustrates the general idea of the semantically enhanced DESCaaS concept which generates the three different types of metadata documents required for meaningful semantic annotations. The documents are created in two stages.

A SemDESCaaS service gets as input a resource (see step 1 in Figure 3). Within the capabilities of a specific SemDESCaaS implementation the input type is further specified. It can be an URL pointing to the resource which is locally or remotely accessible, obtained from a database, or a response of a further service. It is also allowed to send the resource directly to the service, either as stream or as file. A SemDESCaaS service offers different levels of description through additional configuration parameters. In addition, users can include their own metadata to obtain custom descriptions if it is specified in the service capabilities. The output, generated by a DESCaaS service, is a description of the resource and of common capabilities of its source. Our concept leaves open the format to remain flexible and to avoid restrictions to specific types of resources. It always depends on the use case which formats have to be provided by the service, e.g. raw metadata, XML, or specific standards. A generated description is annotated with references linking to the other two metadata documents and can be retrieved by using the first URL in Figure 4. The reference locations are represented by the URLs (2) and (3)[7]. Each of the three URLs follows a pattern by using the original request parameter and the description type. In Figure 4, (2) represents the location of a GI resource ontology. The WSDL description of the service can be requested using the URL (3).

Each of both files is generated in a second step if it is needed. The request of such a metadata document triggers automatically the RMT which translates the resource description offered by the DESCaaS service into the desired document (WSDL or GI resource ontology). If a resource description does not exist yet, it will be created "on the fly" and cross-linked to the other metadata documents. Cross-linking has several advantages. On the one hand, it allows to reason on the semantic description in BPEL compositions since the WSDL files are linked to the GI resource ontology. On the other hand, the original resource description (generated by the DESCaaS service) includes the location of the original resource which enables backtracking. Backtracking is required to provide users with the original resource when they use semantic discovery to find convenient resources.

[6]Web services compliant to the representational state transfer style of software architecture. Further information can be found at: `http://en.wikipedia.org/wiki/Representational_state_transfer`

[7]These URLs are just examples which are not supposed to work.

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

4 Applying the SemDESCaaS to a Pilot Case

The ENVISION project aims to provide a Web-based, environmental decision support platform assisting non ICT-skilled users to create and execute environmental models by integrating semantic technologies. Three demonstrators are set up to approve the suitability of the platform. One of the scenarios predicts the drift of an oil slick spreading out after an oil spill and the effect on cod populations. Predicting oil drifts requires weather and current forecasts. In the project, such information is provided by NetCDF files downloadable via a Web site. NetCDF is a binary machine-independent format for representing time series with three or four dimensions. The discovery of convenient forecasts is an issue domain experts have to deal with. This is mostly tedious and time-consuming, especially, if metadata is insufficient or missing. A further issue covers the integration of such resources into a workflow. They lack of syntactic and semantic interoperability due to missing well-defined interfaces and semantic descriptions.

Within the project, they use BPEL for the specification of workflow compositions and their execution. Such workflows expect chains of W3C compliant Web services to ensure syntactic interoperability. Semantic interoperability is reached through the enhancement of Web services with semantic descriptions encoded in the Resource Description Framework (RDF) and based on the Web Service Modelling Ontology (WSMO) [6]. The link between domain specific vocabularies and the resources (Web services consistent with the OGC and W3C specifications) is established by a GI resource ontology acting as a proxy. In order to meet the current trend in building RDF descriptions of every kind of Web resource, the GI resource ontologies are encoded in RDF as well. The SMO is based on the Procedure-Oriented Service Model (POSM)[8] providing a simple vocabulary for creating RDF-based Web service descriptions. The DMO is also encoded in RDF and linked to the according messages (input/output) specified in the POSM. The semantic annotation is, finally, a relation between elements of the DMO and concepts in a domain ontology specified by logical rules.

The dependence on external forecasts provided as NetCDF files on a Web site results in a big effort for the oil spill modellers. They have to find convenient forecasts (NetCDF files) and migrate them into a W3C compliant service every time, they want to execute the oil spill model. Generating automatically the required Web service descriptions (WSDL file, SMO, DMO) reduces the effort twofold. (1) The resources are prepared to get semantically annotated with meaningful descriptions, and (2) to be used directly within workflows.

First prototypes are implemented supporting domain experts in their modelling tasks. A DESCaaS service, implemented as an OGC Web Processing Service[9] (WPS), generates automatically NcML[10] descriptions for NetCDF files. Indexed NcML files allow users to discover appropriate weather and current forecasts via a catalogue. In addition, they are able to easily analyse the content and the characteristics of the forecasts through the human-readable format. A translation to the various service models is currently achieved by the Service Model Translator (SMT). The SMT is a Java API [11] generating WSDL descriptions and GI resource ontologies for several OGC Web services. The resources (original and generated) are automatically interlinked to provide backtracking. The aforementioned RMT will be based on the SMT and provides the translations for any type of resource instead of OGC services only.

[8]The vocabulary is available at: http://www.wsmo.org/ns/posm/0.1/

[9]The implementation is available for use at: http://giv-wfs.uni-muenster.de/52n-wps-webapp/. The process is called org.n52.wps.server.algorithm.envision.NetCDF2NcMLAlgorithm.

[10]NcML is a standard XML dialect to encode present information about NetCDF files. The specification is available at: http://www.unidata.ucar.edu/software/netcdf/ncml/

5 Conclusion and Future Work

In this paper, we presented an approach to enhance the DESCaaS concept with functionalities enabling semantic annotations for any type of resource. We drawed the substantial need for semantically described resources in the growing Web by raising current interoperability and discovery issues. The concept is supposed to be implemented as a Web service that takes into account the DESCaaS for uniform resource descriptions and the RMT, which generates a WSDL document and a GI resource ontology for each resource.

We introduced existing solutions used in the ENVISION project that can be reused to implement a SemDESCaaS service. We selected exemplarily the oil spill scenario and used our concept to semantically annotate four dimensional NetCDF files. The use case has shown that the approach of creating interlinked WSDL files and ontological descriptions for any type of resource is substantially required. WSDL is an accepted standard and facilitates the integration into commonly used Web service compositions specified, for example, in BPEL. GI resource ontologies provide information about the resource structure itself and establish the link to concepts in a domain ontology. SemDESCaaS is a basis to provide semantic descriptions for all types of resources (including unstructured or binary resources). When metadata descriptions and semantic annotations are in place, they facilitate indexing and cataloguing resources more accurately. This in turn improves the recall, relevance, and accuracy of search results and, finally, the discovery of resources.

The next step is to put together all the described working pieces and implement a SemDESCaaS prototype. We expect that the prototype will play a major role in the on-going work of the ENVISION project. This will allow us to assess the benefits of the proposed solution. However, our approach is still limited since the semantic annotation of resources has to be done by expert users which is tedious and time-consuming. Adapting SemDESCaaS to the Semantic Annotations Proxy (SAPR) will reduce the effort enormously [11]. In the future it should be the objective to fulfill the desire of automatically annotating resources. One example is proposed in [16] to automate the semantic annotation of RESTful services using vocabularies like DBpedia or GeoNames.

Acknowledgment

This work has been partly funded by the Universitat Jaume I predoctoral fellowship (PREDOC/2008/06), by the Fundació Caixa

[11]The API is available at: http://kenai.com/projects/envision/pages/SpecServiceSmt.

Castelló-Bancaixa mobility grant (E-2011-12), and partly by the European research project ENVISION (FP7 249170).

References

[1] T. Andrews, F. Curbera, H. Dholakia, Y. Goland, J. Klein, F. Leymann, K. Liu, D. Roller, D. Smith, S. Thatte, and et al. Business Process Execution Language for Web Services Version 1.1. *Standards proposal by BEA Systems International Business Machines Corporation and Microsoft Corporation*, 2003.

[2] A. Beltran and H. Michels. Description as a Service: Improving data discovery. Currently, the paper is internally reviewed. A rejected version can be found at: `http://www.geoinfo.uji.es/files/arturo/daas_paper.pdf`, 2011.

[3] T. Berners-Lee. Linked data-the story so far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22, 2009.

[4] D. Booth, H. Haas, F. McCabe, E. Newcomer, M. Champion, C. Ferris, and D. Orchard. Web Services Architecture. Technical report, W3C, 2004. `http://www.w3.org/TR/ws-arch/`.

[5] E. Christensen, F. Curbera, G. Meredith, and S. Weerawarana. Web Services Description Language (WSDL) 1.1. Technical report, W3C, 2001. `http://www.w3.org/TR/wsdl`.

[6] J. De Bruijn, C. Bussler, J. Domingue, D. Fensel, M. Hepp, U. Keller, M. Kifer, B. König-Ries, J. Kopecky, R. Lara, et al. Web service modeling ontology (WSMO). Technical report, World Wide Web Consortium (W3C), June 2005. `http://www.wsmo.org/wsml/wsml-syntax`.

[7] N. Guarino. Formal ontology in information systems. In *Formal Ontology in Information Systems: Proceedings of the 1st International Conference, Trento, Italy*, pages 3–15. IOS Press, 1998.

[8] K. Janowicz. Semantic interoperability. `http://musil.uni-muenster.de/images/musil/sage_janowicz_semantic_interoperability_draft.pdf`, 2010.

[9] M. Kavouras and M. Kokla. *Theories of geographic concepts: ontological approaches to semantic integration*. CRC Press, 2007.

[10] W. Kuhn. Semantic interoperability. `http://musil.uni-muenster.de/images/musil/kuhn_portrait.pdf`, 2010.

[11] Maué, H. Michels, and R. Roth. Injecting semantic annotations into (geospatial) web service descriptions. *Semantic Web Journal (SWJ)*, 2011.

[12] P. Maué and D. Roman. The envision environmental portal and services infrastructure. In *Proceedings of International Symposium on Environmental Software Systems (ISESS)*, Brno, Czech Republic, June 2011 27-29.

[13] P. Maué, S. Schade, and P. Duchesne. Semantic Annotations in OGC Standards. Technical report, Open Geospatial Consortium (OGC), 2008. `http://portal.opengeospatial.org/files/?artifact_id=34916`.

[14] Mladenić, D. and Moraru, A. and Škrjanc, M. and Rihtar, M. and Grčar, M. and Roth, M. and Maué, P. and Michels, H. and Llaves, A. ENVISION Deliverable 4.2 - Model annotation components and guidelines. Technical report, Institute for Geoinformatics (ifgi), Westfälische Wilhelms-Universität, 2011. `http://www.envision-project.eu/wp-content/uploads/2011/03/D4.2-1.0.pdf`.

[15] C. K. Ogden and I. A. Richards. *The Meaning of Meaning*. Harcourt Brace Jovanovich, 1946.

[16] V. Saquicela, L. M. Vilches-Blazquez, and O. Corcho. Lightweight semantic annotation of geospatial REST ful services. In *Proceedings of the 8th extended semantic web conference on The semanic web: research and applications - Volume Part II*, ESWC11, pages 330–344. Springer-Verlag, 2011.

[17] S. Scheider. *Grounding Geographic Information in Perceptual Operations*. PhD thesis, Institute for Geoinformatics, August 2011.

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke