Requirements for Geospatial Ontology Engineering

Eva Klien and Florian Probst
Institute for Geoinformatics
Münster, Germany
{klien, f.probst}@uni-muenster.de

SUMMARY

Ontologies have been acknowledged to be the core methodology for capturing and sharing semantics of geospatial information (GI). Ontologies, specifically domain-specific ontologies, are at the heart of most semantic approaches to interoperability. In this paper we want to make a strong case for the importance of domain ontologies in the context of geospatial web service environments. We present an ontology application example and derive from this a requirement specification for geospatial ontologies and the ontology architecture they are embedded in. We claim that the lack of a supportive environment for ontology engineering and maintaining decelerates the efficient use of ontologies in the GI community. Taking into account the requirements we identify a research action line which will help to establish such an environment.

KEYWORDS: *Geospatial ontologies, geospatial web service, semantic matchmaking*

INTRODUCTION

Geospatial information is the key to effective planning and decision-making in a variety of application domains. It also plays an important role as integrative factor across applications. Ontologies have been acknowledged to be the core methodology for capturing and sharing semantics of geospatial information. Ontologies, specifically domain-specific ontologies, are at the heart of most semantic approaches to interoperability. Domain ontologies help to manage semantics of terms used in application schemas and they may enable semantic matchmaking. This is crucial for realising semantic interoperability between different information communities (IC). For the description of geospatial web services we need geospatial domain ontologies as common ground to which members of different communities can commit.

In this paper we want to make a strong case for the importance of domain ontologies in the context of geospatial web service environments. Work on geospatial ontologies is conducted in several research groups and projects[1]. The workshop on Geo-Ontologies 2002 organized by Ordnance Survey (Harding, 2002), the workshop on Action-Oriented Approaches in Geographic Information Science in Maine (ACTOR, 2002), and the workshop on Fundamental Issues in Spatial and Geographic Ontologies held at the COSIT 2003 (COSIT, 2003) showed the variety of approaches and perspectives on ontologies for geographic information in the GI science community. Current research is focused on modelling geospatial ontologies and adequate representation of space and time (Arpinar et al., 2004; Frank, 2003; Grenon & Smith, 2004; Tomai & Kavouras, 2004), theories of vagueness, uncertainty and granularity (Bennett & Cristani, 2004), ontologies for discovery and retrieval of GI (Hiramatsu & Reitsma, 2004; Klien, Lutz, Einspanier, & Hübner, 2004; Lemmens & Vries, 2004; Lutz & Klien, 2005), ontologies for mediation and transformation (Bowers & Ludäscher, 2004; Fonseca, Egenhofer, Agouris, & Camara, 2002), and ontology grounding (Kuhn, 2003, 2005).

---

[1] MUSIL (http://musil.uni-muenster.de/), OntoGeo (http://ontogeo.ntua.gr/), OntoSpace (http://134.102.58.154/), SEEK (http://seek.ecoinformatics.org), SWEET (http://sweet.jpl.nasa.gov/index.html)

The remainder of the paper is structured as follows: we first introduce a GI web service discovery example to illustrate how ontologies are used for inferring the compatibility of offers and requests. Based on this reasoning task, we then specify the requirements a geospatial ontology should meet. Finally we formulate an agenda listing problems and research questions which have to be tackled in order to fulfil the specified requirements.

ONTOLOGY APPLICATION EXAMPLE

We apply ontologies in order to realise ontology-based discovery in geospatial web service environments. The matchmaking, which underlies the ontology-based discovery, is a reasoning process with the goal of deciding, which of the available information sources match the request. Reasoning is the fundamental procedure enabling matchmaking (Sycara, Klusch, Widoff, & Lu, 1999). The main task of the matchmaking process is to resolve semantic heterogeneities between the request and the offer (Klien et al., 2004). This reasoning perspective emphasizes the need for approaches that go beyond the mere construction of ontologies and involve their use in discovery, evaluating, and combining geospatial information (Kuhn, 2005). Semantic matchmaking mechanisms will (a) lead to enhanced usability of heterogeneous and distributed GI sources and (b) facilitate the task of automatic service composition.

In order to illustrate the matchmaking process which underlies the ontology-based discovery we introduce an example in Figure 1. The domain ontology contains the basic terms of a certain domain (in our case hydrology). It is assumed that all actors within a domain share a common understanding of the concepts provided on the domain level (Wache et al., 2001). These concepts are combined and extended in the application ontologies in order to describe the information sources. In our example the information source is a Web Feature Service (WFS) that provides features representing water level measurements. The user in our example searches for water level measurements. He formulates his request for *a water level at time x at control point y* on basis of the concepts of a geospatial ontology. Note, that the application ontologies describing the available information sources are also described using the concepts provided in the same geospatial domain ontology. As a consequence, the user's query becomes machine-comparable to all application ontology concepts in this catalogue. By subsumption reasoning, a terminological reasoner can automatically infer if application concepts are equivalent or sub-concepts to the query concept. As shown in (Klien et al., 2004) the integration of the matchmaking capability into Spatial Data Infrastructures overcomes some of the semantic heterogeneity problems in service discovery and thus leads to increased recall and precision.

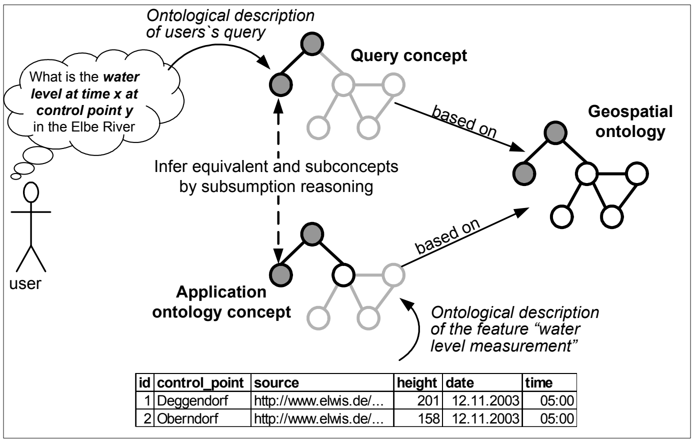

Figure 1*: Example to illustrate semantic matchmaking for ontology-based service discovery.*

REQUIREMENT SPECIFICATION FOR GEOSPATIAL ONTOLOGIES

What needs to be semantically defined in order to support the matchmaking approach introduces in our example? The decisions on what and how things are represented in an ontology are design decisions (Gruber, 1993). In the following we identify the core requirements geospatial ontologies should meet in order to be employed successfully.

Separation of Real World Phenomena and Data Representation

According to the OGC Reference Model, a geographic feature is the starting point for modelling geographic information. They define a feature as an abstraction of a real world phenomenon and a geographic feature as a feature associated with a location relative to the Earth (OGC, 2003).

Analogous, we model *conceptualisations* of real world phenomena that can be located relative to the earth in geospatial ontologies. We use the term *geospatial concept* to refer to these conceptualizations. It is important to note, that data representation features (like point, line, and polygon) that are needed to abstract the real world phenomena, are not part of geospatial ontologies since they deal with the implementation structure of data and not with the semantics of a term referring to a real world phenomenon (Figure 2).

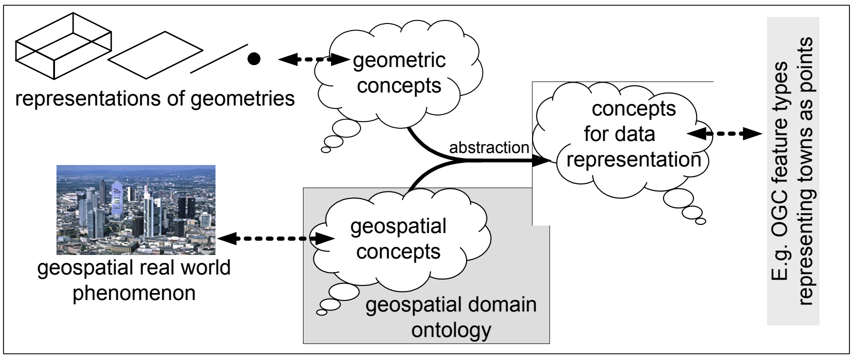

***Figure 2**: The distinction between three types of concepts leads to geospatial domain ontologies which are not biased by implementation needs.*

For example, a town is often represented as a point feature in geospatial applications. But in the first place, the "real world" town has no ontological relation to the representational structure of a point. The domain of geospatial concepts should thus be strictly separated from the domain of data representations. If towns are *modelled* in an application by representing them as points, then this relation between town and its geometrical representation will be part of the application ontology. This view is also reflected in Figure 3, where the domain concepts and representation concepts are distinguished by their colourings. The requirement of keeping geospatial ontologies independent from the implementation view is also a strong argument for introducing a layered ontology architecture as shown in Figure 4**.**

Geospatial Sub-Domains

In the definition above, a distinction is made between concepts for real world geospatial phenomena and concepts for representing them. Defining the scope of the latter ontologies is relatively simple as they are based on existing models for implementing geographic information, e.g. the specifications of the Open Geospatial Consortium (Lemmens & Vries, 2004; Probst et al., 2004).

In contrast, defining the extent of a geospatial ontology is much more difficult since ontologies on the domain level claim to comprise the basic concepts of a common conceptualisation. Great care must be taken to define the concepts and relations on an appropriate level of expressiveness. The terms have to be general enough to allow the annotation of all information sources, but specific enough to make meaningful definitions possible (Schuster & Stuckenschmidt, 2001). In consequence, geospatial ontologies require to be defined within a certain context and for a well-known user community, i.e. we have to come up with adequate and manageable subsets of the geospatial domain.

Moreover, to serve as source for building application ontologies, the domain ontology needs to meet the requirement of high stability. This is, the ontologies should reach after an iterative development phase a status comparable to a standard. Frequent changes in the domain ontologies would discourage service providers to reference their application ontologies on them.

Internal Ontology Structure

The structure of efficiently applicable geospatial ontologies has to meet the requirements of the semantic matchmaking approach in the example. Taxonomic reasoning is useful but not sufficient. Equally, or more important are non-taxonomic relationships, e.g. that a *quantity has a unit of measure*. Consequently, we need ontologies that describe not only simple taxonomic relationships but provide suitable axioms to express other relationships between concepts and to constrain their intended interpretation (Guarino, 1998).

Non-taxonomic relationships play a central part in ontology engineering and should be used wherever possible for defining concepts (Hart, Temple, & Mizen, 2004; Lutz & Klien, 2005; Tomai & Kavouras, 2004). This strategy leads to domain ontologies, which contain not only taxonomic but also non-taxonomic relationships. Figure 3 depicts extracts from domain ontologies for Measurements and Hydrology. In this ontology, taxonomic as well as non-taxonomic relations are defined. Thus, a concept does not have to be given a fixed position in a static hierarchy. Rather, its position in the hierarchy can be dynamically inferred based on existing concept and role definitions using subsumption reasoning. This is fundamental for enabling the ontology-based search for unknown information sources. Some guidelines for the formalisation of domain ontologies are proposed in (Lutz & Klien, 2005).

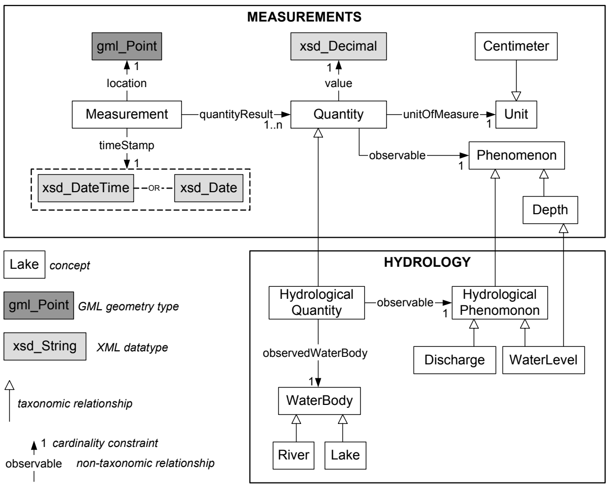

Figure 3: *Schematic representation of extracts from the domain ontologies of Hydrology and Measurement* (Lutz & Klien, 2005).

Representation Language

The selection of the ontology representation language should be based on the inference mechanisms needed by the application that uses the ontology. For achieving semantic interoperability in web service environments, crucial requirements regarding the representation language are the availability of a reasoning engine, the ability to scale up with the requirements of web applications, and the expressiveness to meet the ontology engineering criteria.

Currently, only with Description Logic (DL)-based languages, the inference engine (reasoner) can infer concept taxonomies at run time (Gómez-Pérez, Férnandez-López, & Corcho, 2003), which is needed for semantic matchmaking. These requirements are partly met by the Web Ontology Language (OWL) in its DL or Lite version (W3C, 2004b). OWL-S is a semantic markup language for Web Services and defines an OWL ontology of services (W3C, 2004a). It enables users and software agents to automate the process of discovering, invoking, composing, and monitoring Web resources that offer particular services and have particular properties. More comprehensive service modelling efforts like the Web Service Modelling Language (de Bruijn, 2005) are under way.

Usability

Users on the application level are usually not involved in the development process of ontologies on the domain level. They have to face the task of exploring and understanding the ontology in order to be able to commit to it or not. The application of ontologies will only become widely accepted if methods and tools are provided that support creation and usage of ontologies. Such support is given by possibilities to visualize, browse and query the internal structure of ontologies and by support for implementing a multi-level ontology structure.

Whether these possibilities do exist depends on the representation language. Currently, the open source tool Protégé (http://protege.stanford.edu/) has a rapidly growing user community. It offers a number of functionalities like visual ontology navigation, consistency checking and importing/ exporting different representation languages.

Knowledge Sources

In order to achieve the highest acceptance possible in a user community, it is crucial to base the ontology development on agreed upon knowledge sources.

Standards are sources for the concepts used to describe the models of representing and implementing geospatial information. This has been shown for ISO/OGC standards in (Lemmens & Vries, 2004; Probst et al., 2004).

Geospatial ontologies should be build on agreed upon terminologies and domain expert knowledge whenever possible. Most natural sciences have well-defined terminologies (e.g. geology, hydrology, and meteorology) but not in formalised and machine-interpretable formats. That means, the expenses for creating geospatial ontologies are quite high as they have to be build from scratch. This is surely one of the reasons why only a few geospatial domain ontologies exist in this area.

Ontology Architecture

The approach of a three-layered ontology architecture (Figure 4) provides a solid foundation for ontology engineering which single or two-layered architectures are lacking.

The heart of this architecture is the geospatial ontology on the domain level. As indicated before, it is crucial to develop ontologies on the domain level with the right granularity and with a high level of stability. The domain level contains basic terms of a domain which are combined and extended in the application ontologies in order to describe more complex semantics. With respect to our hydrology example, the concepts *water level*, *water body* and *discharge* are formalized on this level. It is stated that every water body has a water level and a discharge and that these qualities can be observed and measured. This general description provides an entry point for the semantic search. How measuring and representing is done for a specific water level measurement service is then formalized on application level.

Once the domain level is settled, application ontologies can be added or removed without the need of modifications on domain level which makes the application level highly flexible. Their commitment to the same domain level makes the application ontologies comparable. Also, ad-hoc concepts (like query concepts) that are build on basis of the domain ontologies become comparable to all application concepts. The task of constructing an application ontology lies in the responsibility of the provider of the information source whereas the construction of (geospatial) domain ontologies is a joint effort of domain experts. In our hydrology example, the fact that the *water level* is measured in centimetre and not in feet is stated on application level. Additionally all other peculiarities of this specific water level measurement service are describe on application level. This could include legal information of how to use the provided information and data representation issues. The concepts used to describe non geospatial aspects are taken from other types of domain ontologies such as measurement ontologies or data representation ontologies (Figure 4).

The third layer of the architecture represents the ontological backbone. So far, we have talked little about the philosophical aspects of geospatial concepts. The introduction of an upper level (e.g. DOLCE (Gangemi, Guarino, Masolo, Oltramari, & Schneider, 2002)) could help in achieving not only logical consistency (which is provided by the reasoner) but also ontological consistency. Putting the domain ontologies on the foundation of upper level ontologies could enhance the quality of the domain and application ontologies. This can be shown with the concept *measurement (*Figure 3). On domain level, there are various possibilities for the semantics of *measurement*. A domain ontology can subsume *measurement* under the concept *magnitude* und this in turn under the DOLCE upper-level concept *quality,* indicating that the measurement has an attributive character to the thing being measured. Or it could also be modelled as the process which has as result a quantity. In this case, it would be subsumed under the DOLCE upper-level concept *process*. Both conceptualizations might account for the semantics of *measurement*. However, before a service provider is using the concept *measurement* in an application ontology, he would have to make sure which conceptualization is behind it. Otherwise the problem of implicit semantics would just have been shifted from application level to domain level with only partly solving it. This is currently done in the approach of shared vocabularies.

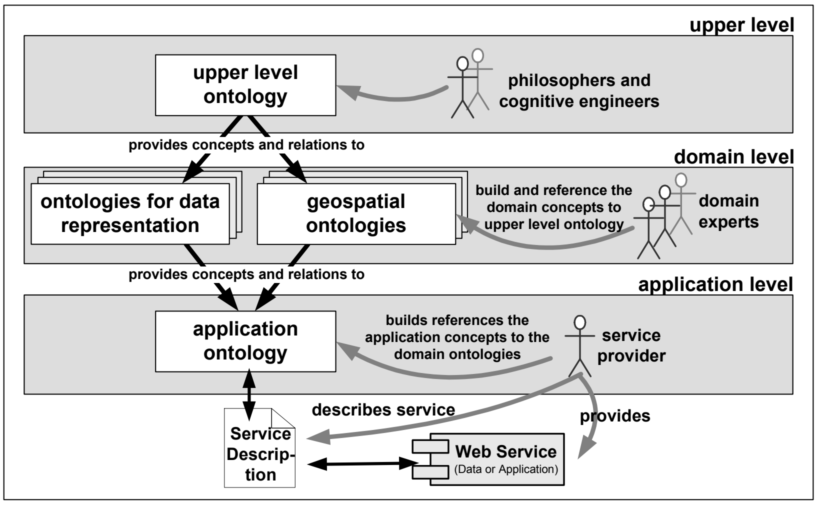

Figure 4*: The three-layered ontology architecture.*

DISCUSSION AND AGENDA

It has been shown that the application of ontology-based matchmaking technology may enhance GI service discovery with respect to precision and recall (Bernstein & Klein, 2002; Klien et al., 2004). Also, research is done on how to apply similar techniques for GI retrieval (Lutz & Klien, 2005) and transformation (Bowers & Ludäscher, 2004).

In 2001 (Gómez Pérez, 2001) stated that the number of ontologies developed is not large and their practical use in final and real applications is small. This is still true at least for the practical use of applications in the geospatial domain. We believe that part of the problem lies in the lack of a supporting environment for ontology engineering and maintaining. Taking into account the requirement specification for geospatial ontologies given above we propose to concentrate research on the following points.

Regarding the content of the domain ontologies, the separation of concepts for data presentation from geospatial concepts is a crucial requirement for consistent, implementation-independent ontologies. Therefore we need to:

⇒ Investigate to which extent data representation and real world geospatial concepts can be separated and maintained in different domain ontologies (e.g. a point is not a geospatial entity).

Regarding the internal structure of the domain ontologies, we need to:

⇒ Investigate domain ontologies centred on non-taxonomic relations and their potential in contrast to the wide spread concept-centred ontologies. Which requirements regarding the expressivity of the representation language need to be considered and do they collide with the requirements regarding reasoning and inference?

Regarding the challenge of revealing the ontological structure of the domain ontology concepts in a philosophical sense, alignment of the domain ontologies with the upper level ontology is needed. This involves:

⇒ Collaboration between the engineers of formal ontology (referring to a philosophical research field in the sense defined in (Guarino, 1998)) and engineers of geospatial domain ontologies.
⇒ Investigations on the type of upper-level ontology. Which existing upper-level ontology is most suitable for providing ontological consistency to geospatial domain concepts? This investigation is part of the SeReS (Semantic Reference Systems) project[2]. In this project the potential of theories of cognitive semantics to serve as upper-level is examined.

Currently the domain ontology engineers can be considered to be the only ones who really can commit to their ontologies. Users have a hard time in understanding the formal statements and fully grasp their meaning. Research on methods and tools is needed in order to:
⇒ Support application ontology engineering, e.g. by automating the process of creating application ontologies.
⇒ Support query formulation in an intuitive way, e.g. by hiding the logical statements from the user.
⇒ Provide means for visualising, browsing and exploring domain ontologies in an intuitive way.
⇒ Provide adapted evaluation methods as proposed by (Gómez Pérez, 2001) to support the user in deciding on the quality of available domain ontologies as well as evaluating the quality of their own application ontologies.

The approach of using ontologies for matchmaking during discovery, retrieval and evaluation of GI is essential for achieving semantic interoperability in web service infrastructures. The goal is to make this approach more widely accepted in the GI community. Specifying in more detail a supportive ontology environment which accounts for changing semantics by providing flexibility and which at the same time serves as semantic reference system by providing stability for the semantic annotations of geospatial applications would certainly support this goal.

ACKNOWLEDGEMENTS
Many thanks to our colleagues from MUSIL (http://musil.uni-muenster.de) for valuable input to this work. The work presented in this paper has been supported by the German Federal Ministry for Education and Research as part of the GEOTECHNOLOGIEN program (grant number 03F0369A) and by the German Research Foundation (DFG) through the Semantic Reference Systems Project (grant KU 1368/4-1).

BIBLIOGRAPHY
ACTOR. (2002). Workshop on Action-Oriented Approaches in Geographic Information Science, from http://www.spatial.maine.edu/~actor2002/motivation.html
Arpinar, B., Sheth, A., Ramakrishnan, C., Usery, L., Azami, M., & Kwan, M.-P. (2004). Geospatial Ontology Development and Semantic Analytics. In J. P. Wilson & A. S. Fotheringham (Eds.), Handbook of Geographic Information Science: Blackwell Publishing.
Bennett, B., & Cristani, M. (Eds.). (2004). Spatial Cognition and Computation: special issue on spatial vagueness, uncertainty and granularity. Spatial Cognition and Computation, Springer.
Bernstein, A., & Klein, M. (2002, June 9-12, 2002). Towards High-Precision Service Retrieval. Paper presented at the The Semantic Web - First International Semantic Web Conference (ISWC 2002), Sardinia, Italy.
Bowers, S., & Ludäscher, B. (2004, March 25-26, 2004). An Ontology-Driven Framework for Data Transformation in Scientific Workflows. Paper presented at the International Workshop on Data Integration in the Life Sciences (DILS'04), Leipzig, Germany.
COSIT. (2003). Workshop on Fundamental Issues in Spatial and Geographic Ontologies, from http://www.comp.leeds.ac.uk/brandon/cosit03ontology/

[2] http://seres.uni-muenster.de/

de Bruijn, J. (Ed.). (2005). The Web Service Modeling Language WSML. http://www.wsmo.org/TR/d16/d16.1/v0.2/
Fonseca, F. T., Egenhofer, M. J., Agouris, P., & Camara, G. (2002). Using Ontologies for Integrated Geographic Information Systems. Transactions in GIS*, 6*(3).
Frank, A. (2003). Ontology for spatio-temporal Databases. In M. Koubarakis et al. (Eds.), Spatiotemporal Databases: The Chorochronos Approach (Vol. 2520, pp. 9-77). Berlin: Springer.
Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., & Schneider, L. (2002). Sweetening Ontologies with DOLCE. Paper presented at the International Conference on Knowledge Engineering and Knowledge Management. AAAI, Madrid, Spain.
Gómez Pérez, A. (2001). Evaluation of ontologies. International Journal of Intelligent Systems, *16*(3), 391-409.
Gómez-Pérez, A., Férnandez-López, M., & Corcho, O. (2003). Ontological Engineering.
Grenon, P., & Smith, B. (2004). SNAP and SPAN: Towards Dynamic Spatial Ontology. Spatial Cognition and Computation, *4*(1), 69-104.
Gruber, T. (1993). A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition*, 5*(2), 199-220.
Guarino, N. (1998). Formal Ontology and Information Systems. In N. Guarino (Ed.), Formal Ontology in Information Systems (pp. 3-15). Trento, Italy: IOS Press.
Harding, J. (Ed.). (2002). Geo-ontology Concepts and Issues. Ikley, UK: Ordance Survey.
Hart, G., Temple, S., & Mizen, H. (2004). Tales of the Riverbank, First Thoughts in the Development of a Topographic Ontology. Paper presented at the 7th Conference on Geographic Information Science (AGILE 2004), Heraklion, Greece.
Hiramatsu, K., & Reitsma, F. (2004). GeoReferencing the Semantic Web: ontology based markup of geographically referenced information*.* Paper presented at the Joint EuroSDR/ EuroGeographics workshop on Ontologies and Schema Translation Services, Paris, France.
Klien, E., Lutz, M., Einspanier, U., & Hübner, S. (2004). An Architecture for Ontology-Based Discovery and Retrieval of Geographic Information*.* Paper presented at the 7th Conference on Geographic Information Science (AGILE 2004), Heraklion, Greece.
Kuhn, W. (2003). Semantic Reference Systems. International Journal of Geographical Information Science, *17*(5), 405-409.
Kuhn, W. (2005). Geospatial Semantics: Why, of What, and How? Journal of Data Semantics, accepted for publication.
Lemmens, R., & Vries, M. (2004). Semantic Description of Location Based Services using an Extensible Location Ontology. Paper presented at the Muenster GI-Days, Muenster, Germany.
Lutz, M., & Klien, E. (2005). Ontology-Based Retrieval of Geographic Information, *under review*.
OGC. (2003). Reference Model (No. OpenGIS Discussion Paper OGC 03-025): OpenGIS Consortium.
Probst, F., Gibotti, F., Pazos, A., Esbri, M. A., Benigno, M., Gutiérrez, M., et al. (2004). Connecting ISO and OGC Models to the Semantic Web. Paper presented at the Third International Conference on Geographic Information Science, Adelphi, MD, USA.
Schuster, G., & Stuckenschmidt, H. (2001). Building shared ontologies for terminology integration*.* Paper presented at the KI-01 Workshop on Ontologies, Vienna, Austria.
Sycara, K., Klusch, M., Widoff, S., & Lu, J. (1999). Dynamic Service Matchmaking Among Agents in Open Information Environments. SIGMOD Record*, 28*(1), 47-53.
Tomai, E., & Kavouras, M. (2004). From "Onto-GeoNoesis" to "Onto-Genesis": The Design of Geographic Ontologies. Geoinformatica*, 8*(3), 285-301.
W3C. (2004a). OWL-S: Semantic Markup for Web Services (W3C Member Submission 22 November 2004).
W3C. (2004b). W3C: OWL Web Ontology Language Overview (W3C Recommendation 10 February 2004).
Wache, H., Vögele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., et al. (2001). Ontology-Based Integration of Information — A Survey of Existing Approaches*.* Paper presented at the IJCAI-01 Workshop: Ontologies and Information Sharing, Seattle, WA.