From Idea toward Ontology

Hardy Pundt

University of Applied Studies and Research Harz
Department for Automation and Computer Science
GIS-Laboratory
Wernigerode, Germany
Email: hpundt@hs-harz.de

Abstract. The paper describes an exemplary ontological engineering process based on the conceptualization of a spatial domain. A pragmatic view of ontologies is briefly introduced, whereas the focus is on the stepwise engineering process. An example is chosen that results from experiences gained during a lecture, including exercises, on semantics of geoinformation. The aim is to describe the main steps during the engineering process. The diffusion of methods that support ontological engineering is important due to the fact that ontologies will be used increasingly to support the access to spatial data sources and information sharing in future, especially within the framework of the development of spatial data infrastructures.

Keywords Ontologies, conceptual modeling, ontological engineering, geoinformation

1 INTRODUCTION

Interoperability between platforms and systems has made significant progress during the last decades, but semantic non-interoperability still hinders problem-free information sharing. Geospatial information is characterized through semantic heterogeneity and techniques are required that support users to identify "relevant" datasets in an increasing jungle of data sources, spread over the Internet worldwide. Ontologies are seen as an approach that can help to support the access to and sharing of information. Therefore it is necessary to formalize ontologies in such a way that they are readable and interpretable for computers. Considering this background, the specifications and standards of the Open Geospatial Consortium (OGC) and the World Wide Web Consortium (W3C) become an important basis to realize both, technical and semantic interoperability. Significant progress has been made concerning several technical issues, and standards such as XML (including derivates such as RDF, GML, and others), as well as OWL are a basis for semantic interoperability.

2 ONTOLOGIES

Meanwhile, ontologies have gained broad interest in Geoinformatics in the discussion about semantic interoperability. Following (Guarino and Giaretta, 1995), the convention in which the capital letter "O" is used to distinguish the "Ontology" in philosophy from others, is widely accepted. "Ontology" in a philosophical view aims at answering questions such as "what is existence", "what properties can explain the existence", "how these properties explain the differences of existence", and likewise. Therefore, the ontologies focus' is not the "existence" in general, but smaller and concrete things such as a specific (spatial) domain, an enterprise, a data source, or the solving of a concrete spatial problem. An ontology is defined as an explicit (and often less ambiguous) description of concepts and relations among them within a specific domain. Such ontologies exist as many as possible domains exist (Mizoguchi and Ikeda 1996). The well known definition of Gruber (1993) describes such "ontologies" adequately: An ontology is "an explicit specification of a conceptualisation". A formal ontology, then, uses concrete classes, properties, relationships, and axioms to describe a domain. It is based on the conceptualisation of the domain and its specification using adequate languages. It is aimed at the implementation of the domain's description in a computer-readable manner.

Formal ontologies can be used to describe specific domains comprehensively. Using languages that are readable by computers, such as the XML-based RDF and OWL, those information communities that collect data can provide formal ontologies and make them accessible via the Internet. In such a way, formal ontologies support the identification of *relevant* data sources, because they are not searching for data on an syntactical basis exclusively. Formal ontologies are thought to include information about the meaning of the terms that are used to describe a domain. Therefore, an ontology provides a semantically enriched domain model – which is a more comprehensive description as a "usual" data model such as an ER-Model or similar methods to conceptualize domains.

From a user's point of view, *relevant* means that a data source contains information that is usable for interested people outside a data provider's information community. Users outside a data provider's community search for information within a specific context. Only such information that fits this context is relevant. Therefore, users must consider the data provider's context to assess the usability of data. If data sources are described via formal ontologies, the processes of identification, access and information sharing can be supported significantly which should help to overcome the problems that are due to semantic variety, as it is typical for Geoinformation (Kuhn 1995, Mark et al. 1999, Harvey et al. 1999, Pundt and Bishr 2002, Giger and Najar 2003). That is the reason why formal ontologies are helpful "means" to achieve semantic interoperability. Klien et al. (2004) summarize this when citing Wache (2001): "One possible approach to overcome the problem of semantic heterogeneity is the explication of knowledge by means of ontologies, which can be used for the identification and association of semantically corresponding concepts". Heterogeneity is due to the specific world views of different geospatial information communities. (Fielding et al. 2004) see this situation as follows: "As ontologies and terminologies expand and as the drive to integrate them increases, it is natural that semantic consistency will become increasingly difficult to maintain. The root of this difficulty is typically the ambiguities and inconsistencies that result from the lack of a standard unified framework for understanding those basic relations that structure our reality."

Taking into account these considerations, it becomes clearer that software application ontologies have the potential to become the keystone in information management techniques. It is expected that these ontologies will support the sort of reasoning power required to navigate large and complex terminologies correctly and efficiently (Fielding et al. 2004). With formal ontologies, being identified as vehicles that could support effectively data identification-, access-, and sharing processes, we get closer to the goal of providing only those data to users that are relevant within a specific context (Pundt 2007). The roles of ontologies (as summarized, for example, in Mizoguchi 1996) are shown in figure 1.

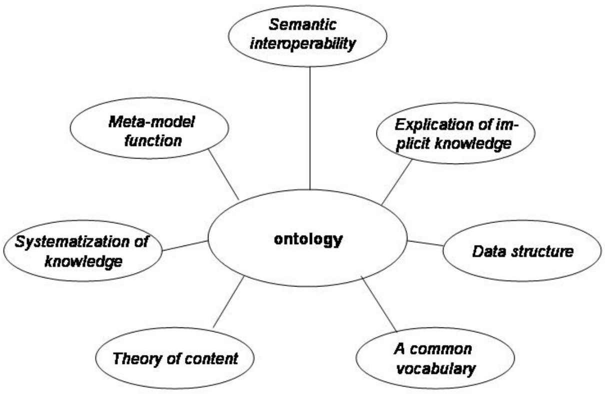

Figure 1: Roles of ontologies.

3 CONTEXT-DEPENDENCY

Ontologies have a large potential to support information search (Fensel 2001). A significant goal to enable the identification of the "right" data is to include information about the information community specific context into data descriptions (Hart et al. 2004). Context-sensitivity of spatial information is the reason for the different interpretations of (same) "things", such as spatial objects. Humans can interpret terms (used to describe "things", or "domains") taking into account experience as well as various knowledge sources they have in mind or can look for (such as metadata). Terms, used to describe "things", have an implicit meaning. Computerized tools haven't such human capabilities. Digital data have no explicit meaning which would be necessary to support readability for computers (Fonseca et al. 2002). Ontologies serve the goal of explication of implicit knowledge.

Metadata cannot replace ontologies. They represent a standardized way of describing data themselves (the well-known "data about data"-paradigm). The goal of metadata is different from the goal of ontologies because metadata are not aimed at the explication of implicit knowledge. Furthermore, metadata are often aimed at being interpreted through human users; in an Internet-connected world mechanisms are required that enable automatic data set identification and –interpretation effectively.

Taking these arguments into account, the goal must be to enable computers to *understand* the information that is provided by different geospatial information communities. If a search machine, for instance, looks for data within a specific context, the access to various ontologies enables the search machine to "understand" the meaning of data provided in the different data sources and therefore evaluate, whether data are usable, or not. The result is a set of data sources provided to the user that includes only "relevant" data, in the best case, but excludes such data sources that were assessed as irrelevant (within the specific context). Metadata can support such a goal but aren't enough to achieve it ultimately. Ontologies follow further goals; the important one is, as mentioned before, the explication of knowledge (Redbrake and Raubal 2004). They facilitate knowledge sharing and reuse by providing machine-processable semantics of information sources that can be communicated between different agents, software and humans (Fensel 2004). In any case, knowledge must be formalized in such a way that the domain and context of the user is made explicit, as well as the domain and context of the data providers. This requires different ontologies, in other terms: all "participants" must provide *their* ontologies. Meanwhile, we face several technologies to formalize ontologies, editors as well as XML-grounded languages that make metadata and semantics machine-processable.

4 A STEPWISE APPROACH

4.1 Introduction

Based on an example from a student's course, ontological engineering is presented as a stepwise process including various methods and techniques (as well as tools). Without claiming completeness, the road map underlying this stepwise approach was defined as follows:

- Definition of the domain, limitation on essential concepts; this can be supported through the formulation of context-specific questions
- Cognitive mapping of the domain; definition of essential concepts, and properties
- Hierarchical organization of the concepts; definition of relationships between them
- Possibly consideration of rules and axioms existing within the domain and affecting the relationships or other aspects
- Transformation of the concepts, properties, relationships and possibly rules and axioms into an ontology (e. g. using an ontology editor)
- Representing the ontology (e. g. using semantic web technologies, such as RDF and OWL)

It is not intended to describe the full engineering process in the following sections. Only some of the steps mentioned before are described, thus representing strongly required steps in the ontological engineering process.

4.2 Context-specific questions

The initiation of the ontological engineering process is a brainstorming in which people are involved that deal with a "common" problem that has to be solved. Collecting ideas, discussing and refining those leads to more concrete thoughts about the domain in which a specific group is interested in. A helpful means to support this process is the formulation of context-specific and competency questions. These are seen as being relevant within the domain and are aimed at giving a clearer idea of the problem itself. According to Noy and McGuiness (2001) the following general questions form the framework:

- What is the domain that the ontology will cover?
- For what we are going to use the ontology?
- For what types of questions the information in the ontology should provide answers (competency questions)?

The competency questions are context dependent, so they vary from domain to domain. The most important goal of the competency questions is to determine the concepts that are essential within the domain. A concept is an idea that characterizes a set, or category, of objects (Sloman et al. 1998), they provide the meaning to terms used in an information system (Probst 2006).

The domain is described via the concepts which are relevant within the context of the problem. The concepts have names, and the names are labelled using a specific language. The language must be understood by all those involved in the engineering process. This requires often an intermediate step: the agreement on a common language, defined via a vocabulary.

4.3 Vocabulary

The terms used to describe the domain must be defined in a vocabulary. Such a vocabulary is the result of discussion of all those that are involved in the ontology engineering process. The vocabulary represents a "common language", a consensus within an information community or between different partners that want to share information. The idea of "common languages" has been presented by Guarino and Giaretta (1998). It is a basis for the realization of semantic translators, and similar mechanisms that support information sharing. The model of Guarino has been modified in figure 2. It shows the intersection of three different information community-specific languages. The intersection, however, represents the space of "common language" that can be used to share information.

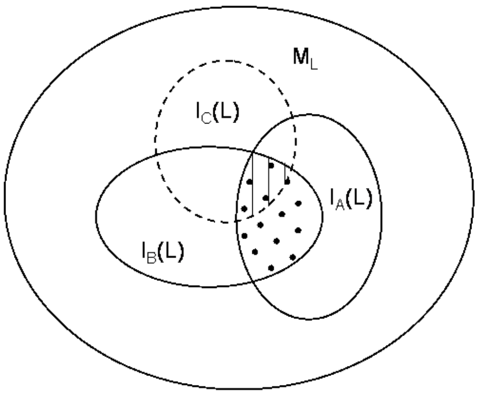

Figure 2: Three geospatial information communities I_A, I_B, and I_C each one using a specific "language", (L). $I_A(L)$ and $I_B(L)$ (as well as $I_C(L)$) are part sets of M(L). A proper communication – information sharing instead of "pure data exchange" - between such communities will be enabled, if the development of a "common ontology" is possible. Such a common ontology is based on the overlapping area between $I_A(L)$ and $I_B(L)$ (dotted area). If several geospatial information communities identify such overlapping areas, then an additional $I_C(L)$ (possibly more) can be integrated in the development of such a common ontology (area with dots and lines) (Guarino 1998; modified).

4.4 Concept definition

Cognitive mapping can be used during the initial phase of the engineering process. A cognitive map shows in a somehow structured view the relevant concepts, and possibly properties and relationships. Cognitive maps can be designed in different ways – what they have in common is that they support the structuring of the concepts that are used to describe the domain and which underlie specific definitions determined in the vocabulary. A hierarchical organization is usual. Figure 3 shows exemplary the result of a cognitive mapping exercise within the framework of a lecture during the winter semester 2006/07. The task was to model a domain of a specific spatial entity, in this case the "universities' environment", which was defined (by the students) as the area surrounding the building of the university (500m buffer), including the building and its human and physical interior. Different groups were built before. They discussed the problem, the objects to be included and developed a common vocabulary within the groups.

Even the groups had the same tasks, very different results were produced and shown in a variety of models. Different interpretations (and uses) of terms occured. This way, the experience of the heterogeneity of world views (of different "information communities", represented by the groups) became transparent for everyone. Doing it this way, the OGCs' information communities' model and the semantic heterogeneity problem was introduced through the experience of a stepwise ontological engineering process.

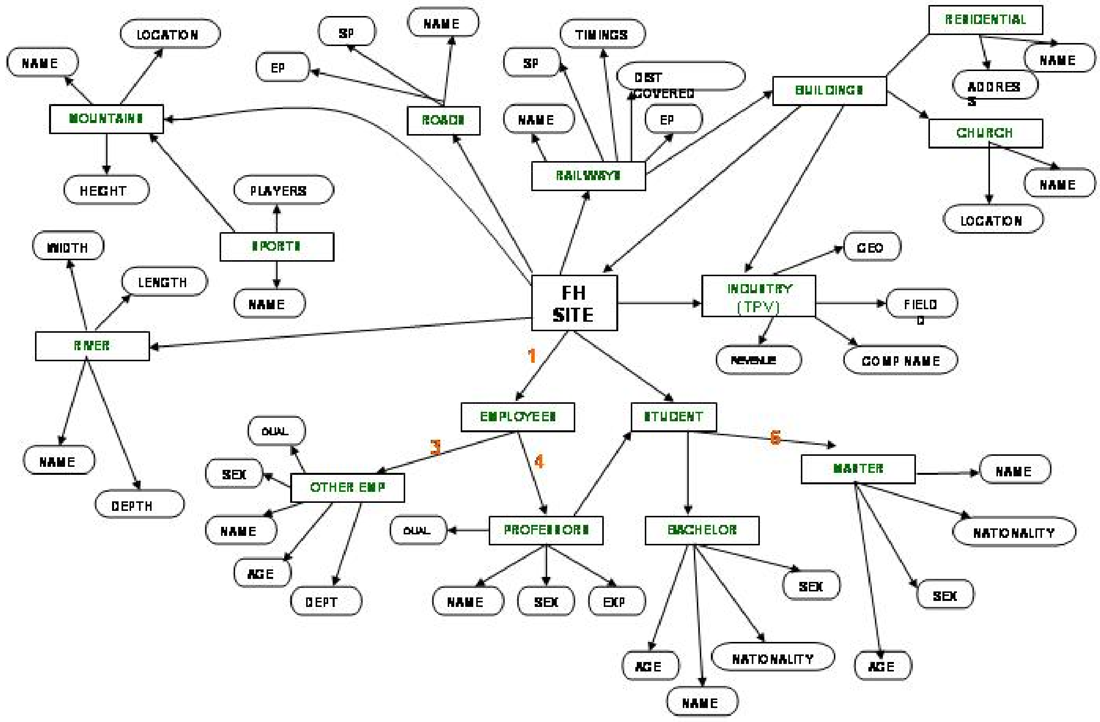

Figure 3: A draft of a cognitive map of "the universities' environment", representing a specific view of a students' group; the numbers represent the meaning of relationships which were defined clearly, as well as the definitions of terms used to describe the concepts and properties (the vocabulary is not included here).

4.5 Hierarchical organization

The cognitive map presented before gives an idea of the hierarchical organization of concepts. Hierarchical organization is typical for many domains. Classes and sub-classes can represent such a hierarchical organisation adequately. The following figure shows just one branch of the ontology resulting from the domain representation shown in figure 3. At this stage of the engineering process, the domain model must be accepted by all those involved in the engineering process. There must be a commitment about the terms that describe the concepts, their properties and relationships, and other aspects being identified as significant for the domain model.

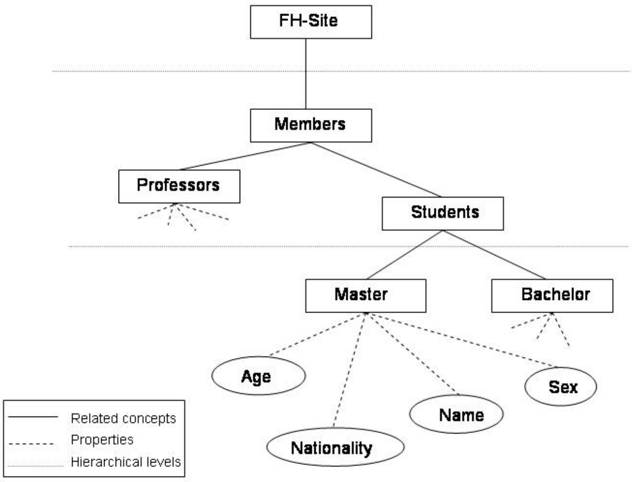

Figure 4: An exemplary hierarchical order of concepts (after modifications).

4.6 Relationships

The concepts are related, thus representing dependencies between them. The relations have a specific meaning that has to be defined exactly. Referring to the cognitive map, the concept "FH Site", for instance, shows various relationships that were defined in a vocabulary and added in the map via a legend, giving each relationship a concrete meaning. Examples are the "`is_nearby`"-relationship to the concept "River" or the "`consists_of`" relationship that exists between "FH-Site" and the "Members".

4.7 Rules and axioms

Axioms can be part of an ontology, thus representing sentences in first order logic that are assumed to be true without proof. Axioms can describe classes as part of an ontology that can't be represented using attributes (or slots), and values. Axioms are used to enable the representation of specific conditions existing in a domain and can enhance the expressiveness of a domain ontology. The ontology editor Protégé includes functionalities that enable the user to define not only classes, attributes, slots, values, and relationships, but also to express axiomatic relations. The Web Ontology Language OWL includes elements that enable engineers to describe axioms, as well as rules, as parts of ontologies (W3C 2007a). This makes ontologies a stronger "means" to describe domains, than other paradigms. Through axioms, specific conditions can be determined, through rules, the derivation of conclusions based on specific arguments is possible.

4.8 RDF, OWL

The Resource Description Framework (RDF) is a foundation for processing metadata. It is based on XML and provides interoperability between applications that exchange machine-understandable information on the

Web. A goal of RDF is to provide facilities to enable automated processing of Web resources. RDF has been enhanced by RDF Schema which defines elements that may be used to describe classes, properties and other resources. Both, RDF and RDF Schema are meanwhile well known standards that are used in various information communities to describe metadata for WWW-based data sources. Within the framework of the lecture and the exercises mentioned before, RDF was used to prototypically describe parts of the ontology. The RDF files were parsed which resulted in RDF-triples, as well as graph-based representations, which was a helpful exercise for the students to get an impression of such XML-based languages.

The Web Ontology Language (OWL) is aimed at describing domains based on the idea of formal ontologies. OWL has been developed based on the experiences of RDF and other languages. They gave insight into the limitations of RDF (and others) in describing ontologies. OWL was a result of the DAML+OIL initiative, and is the current description language for ontologies proposed by the W3C (W3C 2007b).

4.9 Protégé

To draft an ontology, resulting from the steps described before, the ontology editor Protégé was used. Protégé has got great attention during the last years (Noy and McGuiness, 2001). The tool is based on the definition of classes, subclasses, attributes (slots), relationships and further aspects in a hierarchical way. It supports RDF and OWL. Figure 5 shows exemplary the class hierarchy based on the draft in figure 2.

***Figure 5*:** A class structure in Protégé.

5. CONCLUSIONS

Spatial information is always embedded in a specific context. Such a context is specific to a geospatial information community that models reality according to its specific needs and thinking. Spatial information that is provided by such communities can only be used by others if the meaning of concepts is made public. This requires the explication of knowledge.

Ontologies that include a vocabulary of used terms, context information, and the hierarchical organization of concepts including their relationships represent an explication of the knowledge that describes a domain.

If GI-sources are described explicitly via formal ontologies, these can support the automatic identification of relevant information among different distributed data sources. Based on ontologies search machines (as parts of geoportals, for instance) can evaluate, whether data are usable within a specific context, or not. Protégé includes the “elements” necessary to develop such comprehensive data models; RDF and OWL are important languages to implement computer-readable ontologies.

From a practical point of view the methods and tools to design domain ontologies are applicable and there are various proposals how to carry out the ontological engineering process. Some of the important steps were presented in this paper. From the experiences in the course mentioned before it can be underlined that people that are familiar with basic programming paradigms as well as Internet technologies get relatively fast access to the idea of ontologies, their conceptualisation, and implementation. As grounding, the sensitisation of involved people for the problems of semantic heterogeneity, especially in the spatial domains, is indispensable.

The ontology engineering process is a time consuming one. It is encouraging that the idea of the next WWW generation (the “semantic web”), and the research on semantic interoperability in the geospatial domain show parallels. The semantic web should allow “a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing” (W3C 2007b). These activities should lead to synergies that can be fruitful for both, the spatial and the non-spatial communities in future. This will help to pave some more meters on the road toward the “geospatial semantic web”; however, up to now, formal ontologies describing geospatial domains still rarely exist.
.

Thanks to
the students from India, Italy, France, and Austria that participated in a course on “Semantics” at the University of Applied Sciences, Technikum Kärnten, Villach (Austria), in February 2007. Dealing with complex problems, they delivered several ideas to approach a pathway from idea toward ontology.

References

Fensel D. 2004. Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce. 2nd edition. Berlin, New York: Springer.

Fielding, J.M., Simon, J., Ceusters, W., Smith, B. 2004. Ontological Theory for Ontological Engineering: Biomedical Systems Information Integration Proceedings of the Ninth International Conference on the Principles of Knowledge Representation and Reasoning (KR2004), Whistler, BC.

Fonseca F., Egenhofer M., Agouris, P., Camara, G. 2002. Using Ontologies for Integrated Geographic Information Systems. Transactions in GIS 6 (3): 231-257.

Guarino, N., Giaretta, P. 1995. Ontologies and knowledge bases towards a terminological clarification, Proceedings of KB & KS'95, pp. 25-32.

Hart G., Temple S., Mizen H. 2004. Tales of the River Bank, First Thoughts in the Development of a Topographic Ontology. In Toppen, F., Prastacos, P. (eds) Proceedings of the 7th AGILE Conference on Geographic Information Science. Heraklion, Crete: University Press: 169 – 178.

Harvey, F., Kuhn, W., Pundt, H., Bishr, Y., Riedemann, C. 1999. Semantic interoperability: A central issue for sharing geographic information. In: The Annals of Regional Science, Volume 33, Number 2, Springer-Verlag, Berlin, Heidelberg, pp 213 - 232.

Klien, E., Einspanier, U., Lutz, M, Hübner, S. 2004, An Architecture for Ontology-Based Discovery and Retrieval of Geographic Information. In: Toppen, F., Prastacos, P. (eds.): Proceedings of the 7th AGILE Conference on Geographic Information Science, Crete University Press, pp 179 – 188.

Kuhn, W. 1995. Semantics of Geographic Information: GeoInfo 7. Technical University of Vienna, Austria, Department of Geoinformation, 108 p.

Mark, D.M., Smith, B., Tversky, B., 1999. Ontology and Geographic Objects: An Empirical Study of Cognitive Categorization. Freksa, Chr., Mark, D.M., (eds.), Spatial Information Theory. Lecture Notes in Computer Science 1661, Springer, Berlin, New York, pp. 283 – 298.

Mizoguchi, R., Ikeda, M. 1996. Towards Ontology Engineering. Technical Report AI-TR-96-1, I.S.I.R., Osaka University, Japan.

Mizoguchi, R. 1998. Tutorial on ontological engineering. The Institute of Scientific and Industrial Research, Osaka University. Technical Report.

Noy, N., McGuiness, D. 2001. Ontology Development 101: A Guide to Creating Your First Ontology.

Probst, F. (2006). Ontological Analysis of Observations and Measurements. In: Raubal, M., Miller, H.J., Frank, A.U., Goodchild, M.F. eds., 2006. Geographic Information Science. Springer Heidelberg, New York, pp. 304 – 320.

Pundt, H., Bishr, Y. 2002. Domain Ontologies for Data Sharing – An Example from Environmental Monitoring Using Field GIS. Computer & Geosciences, Vol. 28. No 1, pp. 95 – 102.

Pundt, H. 2007. The semantic mismatch as limiting factor for the use of geoinformation in disaster management. Accepted for publication, Taylor & Francis Publishers.

Redbrake D., Raubal M. 2004. Ontology-Driven Wrappers for Navigation Services. In: Toppen F., Prastacos P. (eds), Proceedings of the 7th AGILE Conference on Geographic Information Science. Heraklion, Crete: University Press: 195 – 205.

Sloman, S.A., Love, B.C. 2004. Feature centrality and conceptual coherence. Cognitive Science 22 (cited based on Probst 2006, see above).

W3C 2007a. Web Ontology Language. http://www.w3.org/2004/OWL/ (accessed 21.03.2007)

W3C 2007b. Technology and Society Domain. Semantic Web Activity. http://www.w3.org/2001/sw/ (accessed 28.02.2007)