Exploring GeoMarkup on the Semantic Web

Femke Reitsma
Institute of Geography, University of Edinburgh
Edinburgh, UK
femke.reitsma@ed.ac.uk
Kaoru Hiramatsu
NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation
Kyoto, Japan
hiramatu@cslab.kecl.ntt.co.jp

SUMMARY
Marking up geographic data on the semantic web requires new tools that recognize the special nature of spatial data and the unique relations among geographic features. The potential of geographically referenced data is a dimension of the semantic web that has not yet been well developed. This paper presents initial research into the development of a GeoMarkup tool and associated ontology that enable users to express the spatial nature of their geographically referenced data. The tool generates RDF of the location of geographically referenced information and defines the spatial relations of topology and direction among selected geographic features. The advantages of semantic web technology for geographically referenced information over the current approach of gazetteers and web portals is discussed, and some of the opportunities of the Semantic Web for distributing and utilising geographic information is considered.

KEYWORDS: *semantic web, semantic geospatial web, ontologies*

INTRODUCTION
The World Wide Web is increasingly becoming the primary resource for knowledge acquisition and dissemination. It is a knowledge source that we view through our browsers as a hodge-podge of information that remains largely interpretable by humans alone. With the advent of the Semantic Web, at least a small part of this vast morass is becoming comprehensible to computers via standardized syntax and formalized semantics (Berners-Lee *et al.*, 2001). As more information is semantically marked up and advances in reasoning services for that information are developed, interoperability is enhanced and knowledge discovery becomes smarter and faster.

While the Semantic Web is still in its infancy, there are many potential advantages for adding a geographic component to the future Semantic Web. For example, I walk into the airport, intending to take a flight, and ask my location aware PDA (Personal Digital Assistant) where I should check in. The PDA determines that given my coordinates I am within the Dulles airport building. Given my flight details, which it downloaded from my travel agent, and its Dulles airport semantically marked up map, provided by a local spatially aware service, I need to travel East 50 meters to check in at the right desk. Such spatially aware services take advantage of a geographically referenced Semantic Web, or semantic geospatial web (Egenhofer, 2002).

The objective of this paper is to describe the advantages that ontologies have for storage and querying of geographic information as a new data schema for online gazetteers, and present the initial developments of a GeoMarkup tool developed for presenting that information on the Semantic Web. An ontology for geographic information and spatial relations and a geographic markup tool (GeoMarkup) has been developed to markup geographic information, expressing qualitative and quantitative aspect of geographic information.

GAZETTEERS AND ONTOLOGIES FOR GEOGRAPHIC INFORMATION

In general, the advantages that ontologies provide for geographic information processing include the enhancement of communication, systems engineering, and interoperability (Visser *et al.*, 2002). Their potential largely lies in the ability to reason about the spatial and spatio-temporal relationships. In GIScience, ontologies have been developed for geographic features (Manov *et al.*, 2003), and their uses include ontology-based discovery of geographic information services (Klein *et al.*, 2005), semantics enabled knowledge discovery for earth observation data (Durbha and King, 2005), integration of spatial information (Fonseca *et al.*, 2002), intelligent spatial search on the web (Fu *et al.*, 2005;Jones *et al.*, 2003), and the use of an ontology as the basis of agent communication in an Agent-based GIS (Nolan *et al.*, 2001). Yet these approaches do not allow for the semantics of the spatial data itself to be expressed (Tomai and Kavouras, 2003).

The Semantics of Online GeoInformation Extraction

Gazetteers are used as the basis for most approaches to geographic information retrieval, forming the index for geographic information, such as the World Wide Gazetteer or The Alexandria Digital Library (ADL) Project's Gazetteer service. Although portals utilizing gazetteers typically adhere to syntactic metadata standards, such as ISO 19115 Geographic Information Metadata standard, and syntactic spatial data standards, most notably GML 3.0 (Geographic Markup Language), there are no formal semantics for spatial data or the metadata describing that data.

The primary advantage of using ontologies for marking-up and presenting information on the web are its formal semantics, which support the use of semantic query languages, such as RQL (RDF Query Language), providing the means to access ontology descriptions with minimal knowledge of the schema(s) employed (Karvounarakis *et al.*, 2000). In developing a data format that is flexible enough to represent any form of domain knowledge, ontology based data models can potentially be used to query any form of information (Decker, 2002).

Flexible Accumulation

As the volume of information increases, we need to be able to easily add to this mass in a structured manner facilitating smarter searches across distributed sources of information. Asserting new classes and properties can easily incorporate additions to the ontology. This has implications for spatial database management, where only the ontology needs to be updated rather than the database structure.

Spatial Reasoning

Current approaches for retrieving geographically referenced information with gazetteers are not able to reason about spatial relationships or infer spatial relationships that are not explicitly defined within the database. Furthermore, to build expressive spatial queries about spatial relationships between geographic features, we need constructs describing mereological, topological, direction, and distance relationships (Stuckenschmidt *et al.*, 2001). For example, searching a spatially extended Semantic Web allows us to intelligently search for "hospitals in Washington D.C." where the "in" is interpreted as topological containment (Egenhofer, 2002).

Geographic Information Constraints

Marking up geographically referenced information with a standard ontology allows for checking the spatial consistency of expressions. This will be useful in constraint checking for data input for both the purely spatial aspects of the data and the evolution of represented objects (Borges *et al.*, 1999;Claramunt and Parent, 2003). Furthermore, it may solve problems in metadata description where redundancy, error, and missing information may arise without explicit knowledge of the user, leading to interpretation errors (Pouchard *et al.*, 2003).

GEOONTOLOGY

In order to markup web based information with geographic handles, two test ontologies were developed, an ontology of geographic features and an ontology of spatial relationships. The ontology of geographic features was developed to provide appropriate geographic references, and is focused on the data used to test the GeoMarkup tool. The relationship ontology was developed in order to express the topological, mereological, direction, and distance relationships among geographic features.

The ontologies are expressed in OWL (Web Ontology Language) with the namespaces described in Table 1 below. Some of the advantages of using OWL for expressing the ontologies is that it has an XML based syntax, formal semantics, associated reasoners, and a burgeoning development and user group, with new support tools being developed such as the Jena Java libraries for handling RDF[20].

Ontology Name	Namespace
geoFeatures	http://www.mindswap.org/2003/owl/geo/geoFeatures.owl
geoRelationships	http://www.mindswap.org/2003/owl/geo/geoRelationships.owl

Table 1: Ontologies and their Namespaces

Geographic Features

The feature classes in the ontology were partly derived from a dataset that included spatial information regarding the world's continents, countries, regional administrative units, and major cities. This data was converted to GML and then translated to OWL with an XSLT style sheet. Table 2 below compares the two data models used to express the same information. In OWL, the individual Tokelau is a type of "Country" that has an ObjectProperty "shape". The shape property maps the country domain to a spatial descriptor, MultiPolygon, which defines a polygon with multiple "xyCoordinates" datatype properties.

```
<featureMember>
 <cntry02>
  <_SHAPE_>
   <MultiPolygon srsName="">
    <polygonMember>
     <Polygon>
      <outerBoundaryIs>
       <LinearRing>
        <coordinates>
          -171.84805297851562…
          9.218889236450195
        </coordinates>
       </LinearRing>
      </outerBoundaryIs>
     </Polygon>
    </polygonMember>
   </MultiPolygon>
  </_SHAPE_>
  <CNTRY_NAME>Tokelau</CNTRY_NAME>
 </cntry02>
</featureMember>
```

```
<geo:Country rdf:ID="Tokelau">
 <rdfs:label>Tokelau</rdfs:label>
 <geo:shape>
  <geo:MultiPolygon rdf:nodeID="TokelauShape">
    <geo:xyCoordinates>
      -171.84805297851562…
      9.218889236450195
    </geo:xyCoordinates>
   </geo:MultiPolygon>
  </geo:shape>
 </geo:Country>
```

[20] http://jena.sourceforge.net/

Table 2: A comparison of the same geographic information in GML (on the left) and in OWL (on the right)

Spatial Relations

The spatial relationships incorporated into the ontology are topological, direction, distance, and mereological relationships, which are defined as classes (owl:Class). Their association with geographic features is via the property hasSpatialRelation (owl:ObjectProperty). Following the OpenGIS Simple Features Specification[21] of topological relations based on the Dimensionally Extended 9-Intersection model (DE-9IM) (Clemintini and Di Felice, 1995;Egenhofer and Herring, 1990), the ontology includes the following eight topological relations: equals, disjoint, intersects, touches, crosses, within, contains, and overlaps.

Within the ontology, the elements needed to establish a distance relation are specified, that is, a primary object, reference object, and a frame of reference (Hernández *et al.*, 1995). These distance relationships are further refined as spatial, temporal, and qualitative distance relations. The spatial and temporal relations can be expressed with the classes of distances, such as "WithinMinutesOf" or "WithinMetersOf". Qualitative relations can be expressed as close, medium, and far, and are restricted with owl:Restriction to incorporating a unit reference frame, such as kilometers or seconds. We also include a distance datatype property (owl:datatypeProperty) that allows for the specification of quantitative distances.

We have applied the 8-sector model to express the cardinal directions North, NorthEast East, SouthEast, South, SouthWest, West, and NorthWest, including their inverse (owl:inverseOf) and transitive nature (owl:TransitiveProperty) (Hong *et al.*, 1995). These directions have been defined as properties (owl:ObjectProperty) with two levels of spatial restriction. For example, we define A isNorthOf B to be true if the northern most point of A is further north than northern most point of B as the first level of direction relations, and its subproperty A is CompletelyNorthOf B to be true if the southern most point of A is north of the northern most point of B as the second level of spatial direction relations. Therefore, if A isCompletelyNorthOf B, the relation A isNorthOf B is also true. We have also included relations defined by Freksa's conceptual neighborhood (Freksa, 1992).

The mereological relations included in the ontology are used to define the part-whole inverse relation between things and the transitivity of these relations (Simons, 1987).

GEOMARKUP

A tool was developed that allows us to create well formed RDF for expressing topological and direction relations among spatial objects. This tool utilizes the ontologies discussed above and a number of open source APIs in order to markup spatial features. In particular, Jena[22], the JUMP Unified Mapping Platform (JUMP), and the Java Topology Suite (JTS)[23] were used in developing the tool, which is available online at: http://www.mindswap.org/2004/geo/geoStuff.shtml. The GeoMarkup tool has been developed as a plug-in of JUMP, which is an extendable lightweight GIS (Geographic Information System) for viewing, editing, analyzing, and processing spatial data, and is accessible via a menu from the JUMP Workbench. Thus far, only the topological and direction relations have been implemented with this tool. Development of distance relations depends on defining appropriate algorithms for varying coordinate systems, which have not yet been developed in the ontology.

[21] www.opengis.org
[22] http://jena.sourceforge.net/
[23] http://www.jump-project.org

GeoMarkup Implementation

The JUMP Workbench is presented in Figure 1 below with the sample output. The top left pane displays two geographic features that have been drawn with the editing tools (top right pane) and are selected for spatial markup. The central bottom pane displays two automatically generated URIs for the three selected features, the sets of spatial relations that may be displayed, and the output RDF of the topological, direction, and complete direction relations.

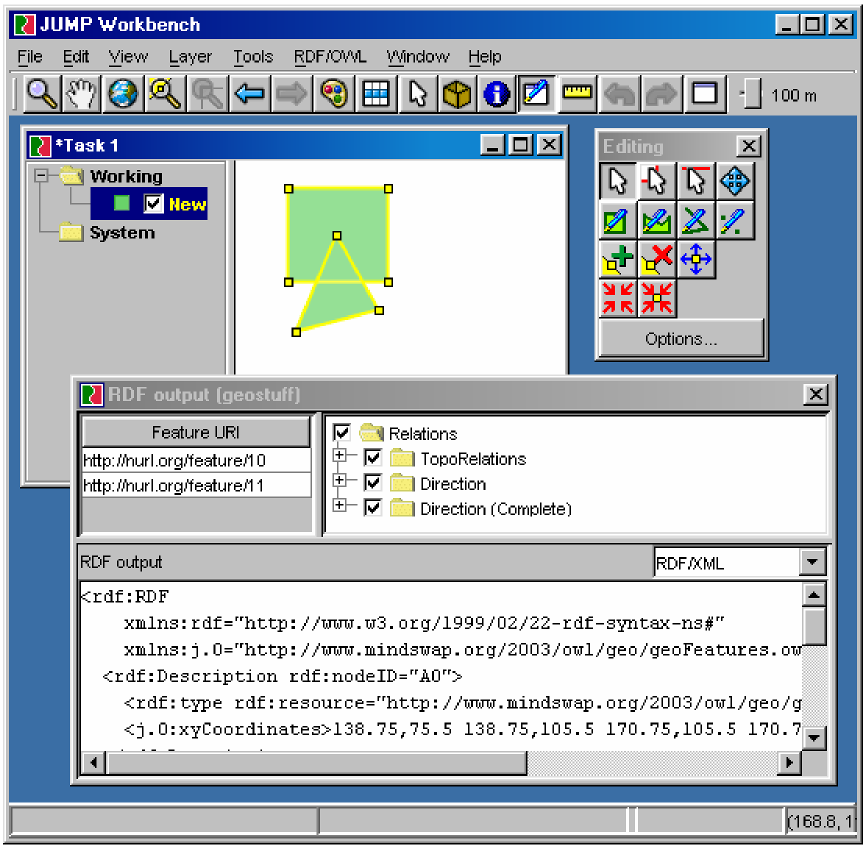

Figure 1: GeoMarkup Tool

The topological and direction relations applied with the GeoMarkup tool are those defined in the ontology, as described above. The direction relations include both the general direction relations (i.e. isNorthOf) and the complete direction relations (i.e. isCompletelyNorthOf). The tool includes added functionality that allows the URI of the object to be edited and automatically responds to edits of or additions to the feature shapes by updating the RDF output. Furthermore, the relations expressed may be limited to only those selected with the checkbox and can be output in different forms, such as in full RDF, abbreviated RDF, N-Triple, and N3, as defined by Jena.

GeoMarkup Application Example

A small dataset of downtown Washington D.C. was selected in order to provide an application example (Figure 2). These shapefiles include road features, prominent buildings, and a fictional target feature located on The Mall, which can be imagined to represent the location of a tourist.

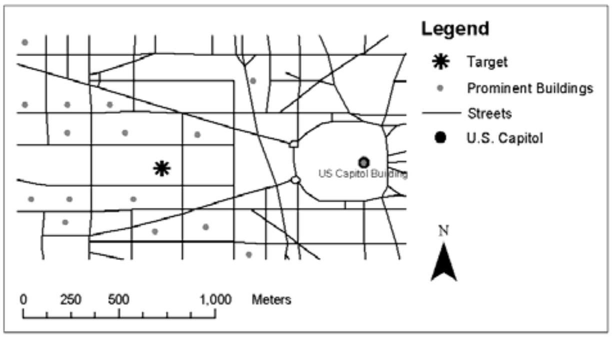

Figure 2: Example data set

After selecting the Target feature and the U.S. Capitol Building feature with the Editing tools within the JUMP Workbench, the topological and direction (basic and complete) relationships between these features can be expressed in RDF (Table 3). Each feature is defined as a resource of RDF:type point, with properties of xyCoordinates and the spatial relationships defining its relation to the other selected feature. Thus the Target point is described as being disjoint with the U.S. Capitol, and lying to West and South of the Capitol.


```
<rdf:RDF
   xmlns:j.0="http://www.mindswap.org/2003/owl/geo/geoRelations.owl#"
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:j.1="http://www.mindswap.org/2003/owl/geo/geoFeatures.owl#" >
 <rdf:Description rdf:nodeID="A0">
  <j.1:xyCoordinates>-77.0092,38.8896 </j.1:xyCoordinates>
  <rdf:type rdf:resource="http://www.mindswap.org/2003/owl/geo/geoFeatures.owl#Point"/>
 </rdf:Description>
 <rdf:Description rdf:nodeID="A1">
  <rdf:type rdf:resource="http://www.mindswap.org/2003/owl/geo/geoFeatures.owl#Point"/>
  <j.1:xyCoordinates>-77.01869267796447,38.889341320146244 </j.1:xyCoordinates>
 </rdf:Description>
 <rdf:Description rdf:about="http://nurl.org/feature/U.S.Capitol">
  <j.0:isSouthWestOf rdf:resource="http://nurl.org/feature/Target"/>
  <j.0:isSouthOf rdf:resource="http://nurl.org/feature/Target"/>
  <j.0:disjoint rdf:resource="http://nurl.org/feature/Target"/>
  <j.1:shape rdf:nodeID="A1"/>
  <j.0:isWestOf rdf:resource="http://nurl.org/feature/Target"/>
  <j.0:isCompletelySouthOf rdf:resource="http://nurl.org/feature/Target"/>
  <j.0:isCompletelyWestOf rdf:resource="http://nurl.org/feature/Target"/>
 </rdf:Description>
 <rdf:Description rdf:about="http://nurl.org/feature/Target">
  <j.0:isCompletelyNorthOf rdf:resource="http://nurl.org/feature/U.S.Capitol"/>
  <j.0:isNorthEastOf rdf:resource="http://nurl.org/feature/U.S.Capitol"/>
  <j.0:isCompletelyEastOf rdf:resource="http://nurl.org/feature/U.S.Capitol"/>
  <j.1:shape rdf:nodeID="A0"/>
  <j.0:isEastOf rdf:resource="http://nurl.org/feature/U.S.Capitol"/>
  <j.0:isNorthOf rdf:resource="http://nurl.org/feature/U.S.Capitol"/>
  <j.0:disjoint rdf:resource="http://nurl.org/feature/U.S.Capitol"/>
 </rdf:Description>
</rdf:RDF>
```


Table 3: RDF output of example scenario

Generating the semantics of spatial data allows us to publish this information on the web for its incorporation into other applications such as semantic web services. For example, if the Target was a tourist, submitting their semantic location to a service via some form of wireless device, would allow for automatic composition with tourist information services. The underlying formal semantics permit reasoning about those services and any information the user requires.

GeoMarkup Discussion

Thus far, the GeoMarkup tool allows us to generate well formed RDF for expressing the spatial location and spatial relations of geographic features. This provides advantages over non-semantic approaches such as gazetteers, as discussed above. However, in order to utilize this marked up information, geographic services for the Semantic Web need to be developed that take advantage of its semantic content.

In a sense, expressing the semantics of spatial data allows it to be treated like any other kind of information on the web. However this belies the special nature of spatial data, which is not accommodated for with current Semantic Web languages and their underlying logics. The types of reasoning available, beyond transitivity for part-whole relations, does not allow us to do the types of spatial reasoning that would be of value for spatial information, be that quantitative or qualitative reasoning. For example, incorporating topological relations such as the RCC calculus or the 9-Intersection model (Egenhofer and Herring, 1990;Randell *et al.*, 1992). Furthermore, there are several aspects of spatial relations that are difficult to express in OWL, such as a restriction that only one type of topological relation can hold between two features, that is, if feature A overlaps with feature B, no other topological relation can hold between those two features. Thus far, the authors are not aware of a single appropriate language that allows for the expression of, and reasoning with necessary spatial concepts.

CONCLUSION

This paper has described the need for a tool to markup geographically referenced information and developed a prototype of such a tool. Important aspects of geographically referenced information, such as their spatial description and the spatial relations among geographic features, can now be expressed with the ontology and topological and direction relations defined with the GeoMarkup tool. This enhances the spatial extension of the Semantic Web, allowing for the expression of spatial relations.

BIBLIOGRAPHY

Berners-Lee, T., J. Hendler, and O. Lassila, 2001. The Semantic Web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *The Scientific American*, 284**:** 34-43.

Borges, K. A. V., A. H. F. Laender, and C. A. Davis, 1999: Spatial Data Integrity Constraints in Object Oriented Geographic Data Modeling. *Proceedings of the Seventh ACM International Symposium on Advances in Geographic Information Systems, pp 1-6, ACM Press, New York.*

Claramunt, C. and C. Parent, 2003. Modelling Concepts for the Representation of Evolution Constraints. *Computers, Environment and Urban Systems*, 27**:** 225-241.

Clemintini, E. and P. Di Felice, 1995. A Comparison of Methods for Representing Topological Relationships. *Information Sciences*, 3**:** 149-178.

Decker, S., 2002: Semantic Web and Databases: Relationships and Some Open Problems. *Proceedings of the NSF-EU Workshop on Database and Information Systems Research for Semantic Web and Enterprises*, Amicalola Falls and State Park, Georgia, USA.

Durbha, S. S. and R. L. King, 2005. Semantics-Enabled Framework for Knowledge Discovery From Earth Observation Data Archives. *IEEE Transactions on Geoscience and Remote Sensing*, 43**:** 2563-2572.

Egenhofer, M., 2002: Toward the Semantic Geospatial Web. *ACM-GIS 2002*, McLean, VI.

Egenhofer, M. and J. Herring, 1990: A Mathematical Framework for the Definition of Topological Relationships. *Fourth International Symposium on Spatial Data Handling*, K. Brassel and H. Kishimoto (eds) Zurich, Switzerland, pp. 803-813.

Fonseca, F., M. Egenhofer, P. Agouris, et al., 2002. Using Ontologies for Integrated Geographic Information Systems. *Transactions in GIS*, 6**:** 231-257.

Freksa, C., 1992. Using Orientation Information for Qualitative Spatial Reasoning. *Theories and Methods of Spatio-Temporal Reasoning in Geographic Space*, A. Frank, I. Campari, and U. Formentini, Eds., Springer-Verlag.

Fu, G., C. B. Jones, and A. I. Abdelmoty, 2005: Building a Geographical Ontology for Intelligent Spatial Search on the Web. *Proceedings of IASTED International Conference on Databases and Applications (DBA2005), February 14-16, 2005*, Innsbruck, Austria, 167-172.

Hernández, D., E. Clemintini, and P. Di Felice, 1995: Qualitative Distances. *Proceedings of Conference on Spatial Information Theory (COSIT'95)*, Semmering, Austria, September 21-23, Lecture Notes in Computer Science 988, Springer.

Hong, J., M. Egenhofer, and A. Frank, 1995: On the Robustness of Qualitative Distance- and Direction-Reasoning. *In Proceedings of Auto-Carto 12, Charlotte, North Carolina, pp. 301-310*.

Jones, C. B., A. I. Abdelmoty, and G. Fu, 2003: Maintaining Ontologies for Geographical Information Retrieval on the Web. *Ontologies, Databases, and Applications of Semantics (ODBASE) 2003 International Conference, 3-7 November 2003. In Proceedings of On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE: Confederated International Conferences CoopIS, DOA, and ODBASE 2002.*, Catania, Sicily (Italy), Springer-Verlag, 934-951.

Karvounarakis, G., V. Christophides, D. Plexousakis, et al., 2000. Querying Community Web Portals, Technical report, Institute of Computer Science, FORTH, Heraklion, Greece. See http://www.ics.forth.gr/proj/isst/RDF/RQL/rql.pdf

Klein, E., M. Lutz, and W. Kuhn, 2005. Ontology-Based Discovery of Geographic Information Services - an Application in Disaster Management. *Computers, Environment and Urban Systems*, in print.

Manov, D., A. Kiryakov, B. Popov, et al., 2003: Experiments with Geographic Knowledge for Information Extraction. *Proceedings of Workshop on Analysis of Geographic References*, Edmonton, Canada.

Nolan, J. J., R. Simon, and A. K. Sood, 2001: Developing an Ontology and ACL in an Agent-based GIS. *Proceedings of the Workshop on Ontologies in Agent Systems, May 29, 2001, Montreal, QC, Canada.*

Pouchard, L., L. Cinquini, and G. Strand, 2003: The Earth System Grid Discovery and Semantic Web Technologies. *Proceedings of the Workshop of Semantic Web Technologies for Searching and Retrieving Scientific Data, The Second International Semantic Web Conference (ISWC-03), Sanibel Island, FL, October 20.*

Randell, D., Z. Cui, and A. G. Cohn, 1992: A Spatial Logic Based on Region and Connection. *Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning*, pp. 165-176.

Simons, P., 1987. *Parts: a study in ontology*. Clarendon Press, Oxford.

Stuckenschmidt, H., C. Schlieder, U. Visser, et al., 2001: Spatial Reasoning for Information Brokering. *Proceedings of the Florida Artificial Intelligence Research Society Conference (FLAIRS), Key West, FL*, pp. 568-573.

Tomai, E. and M. Kavouras, 2003: Pivotal Issues in Designing Geographic Ontologies. *Workshop on Fundamental Issues in Spatial and Geographic Ontologies*, COSIT'03, Ittingen, Switzerland, 24-28 September 2003.

Visser, U., H. Stuckenschmidt, G. Schuster, et al., 2002. Ontologies for Geographic Information Processing. *Computers & Geosciences*, 28**:** 103-117.