OWL-based formalisation of geographic databases specifications

Nathalie Abadie
Institut Géographique National,
Laboratoire COGIT
73 Avenue de Paris
94160 Saint-Mandé, France
+33 1 43 98 80 03
nathalie-f.abadie@ign.fr

Ammar Mechouche
Institut Géographique National,
Laboratoire COGIT
73 Avenue de Paris
94160 Saint-Mandé, France
+33 1 43 98 80 00 + 71 25
ammar.mechouche@ign.fr

Sébastien Mustière
Institut Géographique National,
Laboratoire COGIT
73 Avenue de Paris
94160 Saint-Mandé, France
+33 1 43 98 81 49
sebastien.mustiere@ign.fr

ABSTRACT

The ability to share and combine geographic data from different information sources in a consistent way is a key issue for enabling successful implementation of Spatial Data Infrastructures (SDIs). This can only be done through a deep understanding of databases structure and content. In this poster, we propose to do that through the elicitation and formalisation of geographic database specifications, relying on OWL ontologies, as recommended in the semantic Web community. We thus propose a general ontology for eliciting key concepts manipulated by data specifications, and rules to build local ontologies representing knowledge contained in specific data specifications.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Databases Applications – *Spatial databases and GIS.*

General Terms

Management, Standardization.

Keywords

Geographic Database Specification, Ontologies, Geo-data Semantics, OWL

1. WHY FORMALISING SPECIFICATIONS?

In the last decades, the increase of geographic data acquisition campaigns has resulted in a huge amount of diverse, heterogeneous and distributed geographic data sources. However, even if these data represent the same geographic real world, there is a great heterogeneity between them. Consequently, the ability to share and combine geographic data from different sources in a consistent way is a key issue for enabling their efficient usability.

Previous geo-data integration efforts mainly focused on syntactic heterogeneities through the development of standards. Semantic interoperability, which addresses more complex problems, is still investigated. Actually, recent works mainly focused either on geo-data discovery and retrieval or on transformation of geo-data schema. In the former case, most of the proposed approaches [1][2][3] use a global domain ontology to specify the precise meaning of geo-data, either by renaming feature classes with ontology labels, or thanks to semantic annotations. They rather aim at helping a user in retrieving geo-data that represent a specific geographic concept, such as '*buildings'*, even if feature class names of available datasets are totally different. In the latter cases, recent approaches [4][5][6] provide geo-databases experts with a graphical interface to help them in manually describing their schemas and specifying mappings between source and target schemas.

However, each geo-data producer has its own rules for data capture, and its own point of view about the geographic real world [7]. As an example, if a feature class is named 'Building', it may actually designate only permanent buildings, or include precarious buildings, such as cabins, or huts. Besides, a geographic database is produced at a specific scale of analysis and geographic features are then captured in the database consistently with this specific level of detail. For example, only buildings of area greater than 50 m^2 may be captured. Furthermore, the geometric representation of a given geographic feature may vary: a building may be represented by a polygon representing its perimeter or by a point captured at its centre.

All these selection and representation criteria are stored in specific textual documents, used as guideline for data capture, namely the database specifications. They are a very rich source of knowledge about geo-data semantics and their use in a schema matching process could help in identifying and solving complex heterogeneities. Let us consider two different databases covering the same geographical space. The first one has a feature class named 'Building' which represents only "buildings of area greater than 20 m^2", while the second one has a feature class named 'Built-up area' which represents "buildings of area greater than 50 m^2". Comparing these feature classes' specifications enables to find the following mapping rule: 'Building' instances of area greater than 50 m^2 represent the same real world buildings as 'Built-up area' instances. Providing a schema matching application with formal specifications would therefore enable to automatically find such complex mapping rules between heterogeneous geo-databases.

2. THE SPECIFICATIONS ONTOLOGY

Several formal models for geographic database specifications have already been proposed [8][9]. As formalisation of data specifications in SDIs is a kind of elicitation of data semantics in a Web environment, we propose to rely on semantic Web standards to do so: our approach is based on ontologies developed with the Ontology Web Language (OWL 2 [10]).

A first step to formalise specifications is to define unambiguously key concepts commonly used in geo-database specifications. In other words, we define a domain ontology, named “Specifications Ontology” (SO, see Figure 1). This ontology SO only contains concepts specific to geographic data specifications. It relies in turn on more general ontologies, for example for defining basic geometric types [11]. For example, this domain ontology SO formalises the concepts of data source and centreline, which are commonly used in many data specifications.

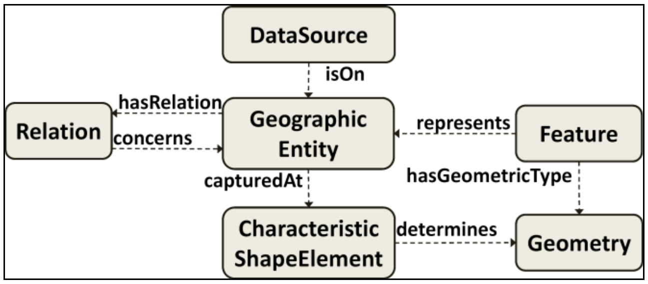

Figure 1. The Specifications Ontology's main classes

3. HOW TO FORMALISE A GEO-DATABASE SPECIFICATION?

Besides we propose to formalise each database specification by means of an application ontology, named “local specification ontology” (LSO). This ontology imports SO and extends it to describe real world geographic concepts and database classes.

In order to clearly separate database concepts from real world concepts encountered in specifications, two main classes from SO are used: *GeographicEntity* and *Feature*. On the one hand, *GeographicEntity*’s subclasses are concepts imported from a domain ontology of topographic concepts. They may be created from the specifications text thanks to natural language processing tools [12]. On the other hand, *Feature*’s subclasses are concepts directly derived from the corresponding database schema. They contains information such as selection criteria used to populate the database, such as the fact that “only habitation buildings of area greater than 20 m^2 are represented in the feature class ‘Building’ of the database”. We thus require feature classes such as ‘Building’ to be modelled as “classes” in the OWL language, and selection constraints to be modelled as “axioms” including rules that restrict the possible interpretations for the defined term, those axioms being defined by means of concepts and relations defined in SO and LSO. Considering the example above, the feature class ‘Building’ of this geo-database will be defined in LSO as follows:

```
Class: lso:db_Building
        EquivalentTo: so:represents some
        (lso:Habitation and so:area some
        double[>20.0])
```

4. CONCLUSION

In this poster we proposed an OWL 2 based model for geographic database specification formalisation, which aims at eliciting geographic databases semantics by describing the link between data and what they represent. Key concepts used in data specifications are specified in a specifications domain ontology (SO), whereas knowledge contained in one given database specification is described in a specification application ontology (LSO) which uses SO’s concepts. A tool enabling automatic comparison of formal specifications is being implemented. It aims at providing expressive schemas mappings between geographic heterogeneous databases, for schema translation or schema integration purposes.

5. ACKNOWLEDGMENTS

This work is partly funded by the French Research Agency through the GeOnto project ANR-O7-MDCO-005 on “creation, alignment, comparison and use of geographic ontologies” (http://geonto.lri.fr/).

6. REFERENCES

[1] Paul, M., and Ghosh S.K. 2006. An Approach for Service Oriented Discovery and Retrieval of Spatial Data, In *Proceedings of International Workshop on Service Oriented Software Engineering*, Shangay, China, 84-94.

[2] Nambiar, U., Ludscher Ludäscher, B., Lin K., and Baru, C. 2006. The GEON portal: accelerating knowledge discovery in the geosciences, Procedings In *Proceedings of the 8th ACM International Workshop on Web Information and Data Management*, Aarlington, Virginia, USA, 83-90, (2006).

[3] Klien E.M. 2008.*Semantic Annotation of Geographic Information*. Phd thesis, Institute for Geoinformatics, University of Muenster. Muenster, Germany.

[4] Balley, S. 2007. *Aide à la restructuration de données géographiques sur le Web - Vers la diffusion à la carte d'information géographique*. PhD Thesis, University of Marne-La-Vallée, France.

[5] Schade, S. 2009. Translation of Geospatial Data, Challenges, Solution and Vision, In *Proceedings of the 12th International Conference on Geographic Information Science (AGILE'09), Pre-Conference Workshop “Challenges in Spatial Data Harmonisation”*, Hannover, Germany.

[6] Reitz, T., de Vries, M., and Fitzner, D. 2009. *A5.2-D3[3.3]Conceptual Schema Specification and Mapping*, HUMBOLDT Technical Report.

[7] Fonseca, F., Clodoveu, D., and Camara, G. 2003. Bridging Ontologies and Conceptual Schemas in Geographic Information Integration. In GeoInformatica, 7(4), 355-378.

[8] Gesbert N. 2005. *Etude de la formalisation des spécifications de bases de données géographiques en vue de leur intégration.* Phd thesis, University of Marne-la-Vallée.

[9] Christensen, J. V. 2006. Formalizing Specifications for Geographic Information. In *Proceedings of the 9th AGILE Conference on Geographic Information Science. College of Geoinformatics, University of West Hungary*, 186-194.

[10] W3C. 2009. *OWL 2 Web Ontology Language, Primer*, W3C Recommendation 27 October 2009. http://www.w3.org/TR/owl2-primer/

[11] Lieberman J., Singh R., Goad C. 2007. *W3C geospatial ontologies*, W3C incubator group report, 23 october 2007.

[12] Kamel M., Aussenac-Gilles N. 2009. Ontology Learning by Analysing XML Document Structure and Content, In Proceedings of Knowledge Engineering and Ontology Development, Madère, Portugal.