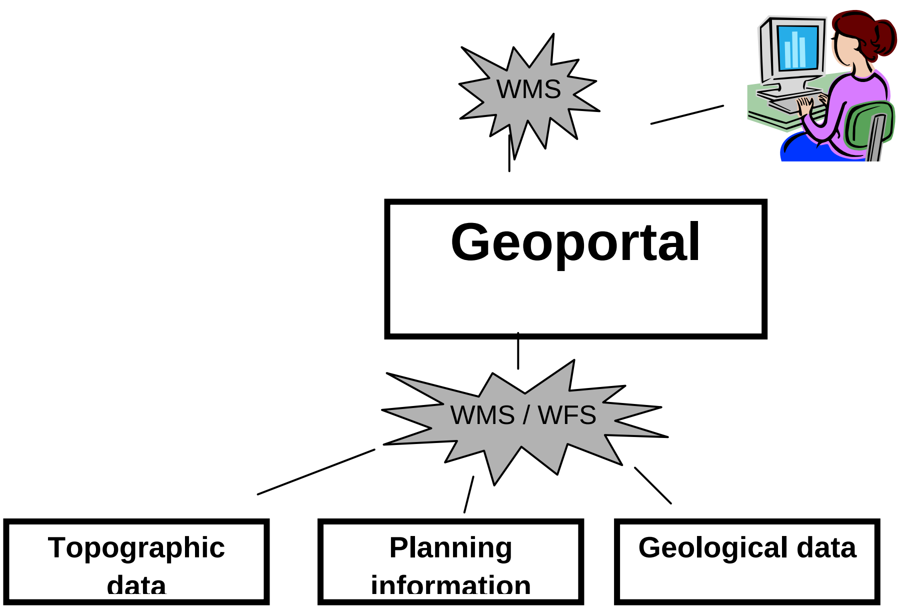
Cartographic and semantic aspects on web services

Discussion group at Dagstuhl – Generalisation of Spatial Information

Lars Harrie, Lund University, Sweden
Heiner Stuckesnchmidt, University of Mannheim, Germany

1 Background

Several countries are currently working on setting up geoportals as part of their national spatial data infrastructure (SDI) (and this is also a requirement of the Inspire initiative). A key ability of these geoportals is that the user should be able to view (and download) data from several sources from one access point. This will certainly make the access to geospatial data easier. However, there are also cartographic and semantic challenges that have to be solved. In this discussion group we discussed some topics concerning both download services and view services (as in the figure below) and some possible solutions.



Dagstuhl Seminar Proceedings 09161
Generalization of spatial information
http://drops.dagstuhl.de/opus/volltexte/2009/2134

2 Download services

Problems can arise when the user uses different download services to combine data from different, possibly heterogeneous sources and wants to combine them into a single view in the geo-portal. In principle, the problems that can occur in this situation are the typical problems of data integration that can be found in any domain, i.e. inconsistency, redundancy and differences in granularity and conceptualization. In the context of geo-information, these general problems manifest themselves as follows:

1. *Inconsistency in data + redundancy in data:* In some cases several organizations have same object types (e.g road data is stored both by NMA and road administrations). Then inconsistency / redundancy of data could be a problem due to e.g. different update cycles.

2. *Different Levels of Detail:* Organisations are using different level of details in their data, which will cause problems when the user is merging data from different sources.

3. *Different classifications:* Organisations are using different classification schemas which of course are problematic when merging data.

3 View services

View services have the same problem as download services, but come with some additional problems. In particular, in the case of download services, the way, information is represented in the portal was defined by the user accessing the different sources who could ensure a uniform representation. In the case of view services, not only the data, but also the way it is represented is defined by the local sources leading to the following additional problems:

4. *Symbology:* The base data (e.g. topographic data) is often rendered to optimize its own visualization. This is problematic when other data (additional data) is put on top of the base data. E.g. there could not be any suitable color left. Another problem is that different servers might use the same symbols for different object types.

Possible solutions: The user must be given the possibility to change the symbology in the view services e.g. with SLD (styled layer descriptor). Another possibility is that the user can choose among a set of symologies (stored e.g. on the server or geoportal level).

5. *Overlapping problem*: Data can overlap or be too congested for cartographic visualization.

4 Potential Role of Semantic Technologies

It has been widely acknowledged, that semantic technologies can play a major role in overcoming data integration problems in different domain. Research in the database as well as the semantic web area have developed technologies for describing the intended meaning of data in different sources and using these definitions for defining semantic relations between different sources that can be used to integrate information in a meaningful way. Thus a natural question is whether semantic web technologies can also help to overcome the problems we identified above. In the following, we discuss the different problems and identify the potential contribution of semantic technologies to solving the problem. As we will see, some of these problems can naturally be addressed using semantic web technologies while others elude a solution involving semantic technologies.

1. *Inconsistency and redundancy in data:* Checking consistency of definitions is a basic functionality of semantic web technologies and can be implemented using the web ontology language OWL [Horrocks et al 2003]. This, however, requires the data from the

different sources to be described as instances of a common ontology. This ontology has to specify explicit consistency constraints for the data and is limited to certain types of inconsistency. In particular, semantic technologies can only be used to identify conceptual inconsistency such as legal combinations of types the same object can have or legal types of objects it can be in a certain relation with. Other types of inconsistency resulting, for example from outdated data cannot always be found as semantic technologies cannot check data against the real state of the world, but only against data from another source. Checking redundancy is not directly supported in OWL.

2. *Different Levels of Detail:* Levels of details are a problem that is very characteristic for geo-data. While the granularity of data is also an issue in other domains, the issue of granularity has some very specific properties in the domain of geo-services. Here the issue of granularity is a geometric rather than a semantic problem. If the data from two sources do not have the same level of detail, the more detailed map has to be abstracted. This can be done using existing abstraction methods that are mostly geometric by nature. It is not entirely how semantic technologies can help in this case.

3. *Different classifications:* Using semantic technologies to integrate heterogeneous classifications used by different data sources is one of the more promising applications of semantic technologies in this context. In fact the use of semantic technologies for integrating heterogeneous object catalogues has already been described in the literature [Stuckenschmidt and van Harmelen 2004] and there is a rich literature on matching ontologies [Euzenat and Shvaiko 2007] that can also be applied to object catalogues. A problem not adequately addressed by current semantic web technologies is vagueness of concepts in the geographic domain (what is a mountain as opposed to a

hill?). Recent work combining OWL and fuzzy reasoning [Straccia 2005] addresses this problem to some extend but so far, this extension is not an official language and there are no experiences with using Fuzzy OWL for modeling geographic concepts.

4. *Symbology:* the problem of in compatible symbology that arises from the use of different view services cannot be resolved using semantic technologies; however, it could be possible to use semantic technologies for detecting problems with symbology that might not be spotted by the user at first sight. In particular, it is possible to build a semantic model of types of objects shown in the geo-portal. The description of such concepts could contain a relation that links the type with the symbol used to represent objects belonging to that type. In order to make sure that no symbol is used for different concepts, this relation could be specified as being one-to-one. When a new data source is included in the portal, it needs be represented as an instance of that model. If a symbol is now used for different object types or the same object type is represented by different symbols, this causes and inconsistency in the semantic model that can be reported to the user.

5. *Overlapping problem*: This problem again is a typical example of a problem with no obvious use of semantic technologies. It might be possible to also build a semantic model of the data that constrains the configuration of polygons on the screen in such a way that potential problems result in inconsistencies in the semantic model. Such a model, however, will be mainly concerned with spatial constraints that cannot easily be encoded in semantic web languages. This makes it unlikely that semantic technologies are a good choice for solving this particular problem.

5 Conclusions

The integration of multiple spatial data sources into geo-portals comes with a number of potential problems related to mismatches in the data to be combined in the portal. We distinguished between problems related to download services and additional problems arising from the use of view services where conflicts can also include the representation of spatial objects. As briefly explained, some of these problems could be addressed using semantic technologies. In particular, semantic technologies can be used to detect inconsistencies across data sources provided that the data from different sources is described using a common semantic model. Further, semantic technologies provide support for the integration of heterogeneous classifications of objects which is a fundamental requirement for having a meaningful integration of different datasets. Beyond these possible applications there are also a number of problems like different levels of detail and overlapping of polygons that should not be addressed using semantic technologies but should rather be addressed using computational geometry and related methods.

6 References

Ian Horrocks, Peter F. Patel-Schneider, Frank van Harmelen: From SHIQ and RDF to OWL: the making of a Web Ontology Language. J. Web Sem. 1(1): 7-26 (2003)

H. Stuckenschmidt and F. van Harmelen. Information Sharing on the Semantic Web. Advanced Information Processing, Springer Verlag, Berlin, Heidelberg, 2004

Jérôme Euzenat, Pavel Shvaiko. Ontology Matching. Springer-Verlag, Berlin Heidelberg (DE), 2007.

Umberto Straccia: Towards a Fuzzy Description Logic for the Semantic Web (Preliminary Report). ESWC 2005:167-181