What Have Ontologies Ever Done For Us – Potential Applications at a National Mapping Agency

John Goodwin

Research and Innovation, Orndnance Survey of Great Britain, Southampton, UK
john.goodwin@ordnancesurvey.co.uk

Abstract.

Current developments within the information technology industry are creating new opportunities for traditional information suppliers such as Ordnance Survey. In particular the ability to electronically trade information creates opportunities to increase the use of Ordnance Survey's topographic information in ways that can be go beyond the delivery of such data in the form of a map. The future development of a semantic web may create new business opportunities particularly for the mobile connected user who may access Ordnance Survey information through an application without ever realising it. More generally the ability to semantically translate between one information source and others may reduce both the time and cost of services better enabling joined-up government and industry. In response to these challenges Ordnance Survey has embarked upon research to investigate the development of a topographic ontology to underpin our data and to investigate how it may be used to support interoperation with other information sources and information based services. This paper describes some of the potential applications for ontologies at Ordnance Survey.

1 Introduction

Ordnance Survey is Britain's National Mapping Agency. It currently maintains a continuously updated database of the topography of Great Britain. The database includes around 440 million man-made and natural landscape features. These features include everything from forests, roads and rivers down to individual houses, garden plots, and even pillar boxes. In addition to this topographic mapping, entire new layers of information are progressively being added to the database, such as aerial photographic images which precisely match the mapping; data providing the addresses of all properties; and integrated transport information.

More and more companies and public services are using computer-based geographical information systems (GIS) and web-enabled services which allow the rapid integration and analysis of information from many sources, effectively bringing maps to life in an interactive way. Currently the costs of data conflation and adaptation activities are a major barrier to the adoption and efficient exploitation of complex datasets such as Ordnance Survey MasterMap™. These costs are currently being reduced through data structure standardisation, however understanding both the actual content of the data, and how it can be used with other data sources remains expensive. To reuse and share geo-data successfully, integration has to be realised not just on a syntactical level but also on a semantical level. When exchanging information with customers or partners we could assume that everyone is talking about the same thing given a particular term, word or symbol. However, it seems unlikely that true harmonisation will ever be practical between multiple data suppliers. An alternative approach is to provide data with meaning defined using an agreed structured language – this is what lead us to investigate an ontology solution. Sharing information in this way enables integration and reuse of knowledge and data across various applications.

We are investigating how to make Ordnance Survey data more interoperable by translating between different semantics. Other benefits of our current work include better understanding and modelling of our data, improvements in our core database models, the development of intelligent web services, and understanding how data can be translated for many different user tasks. In this paper we discuss at a high level these, and other potential applications of ontologies at the Ordnance Survey. We will not go into complex scenarios or implementation solutions.

2 Data Modelling and Data Consistency

Ordnance Survey's current internal data model describes what we called "real world types" – these are things you see in the world around you, for example banks, churches, playing fields, vegetation etc... These can be describe in terms of a form (the physical structure, e.g building, inland water etc.) and function (the intended use of a geographic feature, e.g. education service, pond etc.). The geometries of geographic featuers are represented by points, line or polygons.

We currently have a rulebase stating which combinations of geometry, form and function are valid. For example you would not want to find a geographic feature that had function education service and form inland water. There are additional rules topological rules that check (among other things):

- Which line features can bound which area features
- Which area features can be contained within which area featuers

There are currently around seventy thousand of these rules, and at the moment they are hard coded into a software application. As such there is no way to verify that the entire rule set is consistent. Thankfully the rules can easily be expressed in OWL using axioms of the form:

$$\begin{aligned} \text{RealWorldTerm1} \sqsubseteq\ & \exists \text{hasFunction}.(\text{Function1} \sqcup \text{Function2} \sqcup \text{Function3} \ldots) \\ & \sqcap \forall \text{hasFunction}.(\text{Function1} \sqcup \text{Function2} \sqcup \text{Function3}\ldots) \\ & \sqcap \exists \text{hasForm}.(\text{Form1} \sqcup \text{Form2} \sqcup \text{Form3}\ldots) \\ & \sqcap \forall \text{hasForm}.(\text{Form1} \sqcup \text{Form2} \sqcup \text{Form3}\ldots) \end{aligned}$$

$$F1^{1} \sqsubseteq \exists validWithin.\top \sqcap \forall validWithin.(F2 \sqcup F3 \sqcup F4...)$$

With our rules coded in an OWL ontology we can check that they are all consistent using standard DL reasoners.

We can then use this ontology to check that the objects in our database conform to our rules. In theory these rules could also assist surveyors and data collectors in the field. Some geographic features can be hard to unambiguously classify, but a formally encoded rule about classifcation could aid the sureyor's decision making process.

In theory this should also provide a cheaper solution than the hard coding of rules in a software application. As the data model changes or gets up dated it is far easier to modify the ontology and check the modifactions are consitent than it is to update software code and (if necessary) debug said software code.

We should also be able to use this method to check conformance of our products to specification.

3 Adaptable Data

If ontologies are to fullfill their full potentiential there must be some way of linking them with legacy databases. There are currently reasonably mature tools on the market that enable this.

Databases are typically designed and implemented by IT personel. The database then has to be queried against the physical schema. This can often cause much confusion. Ontologies enable business, domain and policy knowledge to be captured in a far more intuative and portable information model creating a view onto the data that can be easily tailored and adapted for different user needs.

To achieve this we envisage having a layer of (at least) three ontologies sitting on top of a database. The first ontology, the data ontology, essentially provides an interface between the database schema and the domain ontology. The domain ontology describes the geographic and topographic domain from the Ordnance Survey view of the world.

With these two layers in place we query our database through the concepts described in our ontology instead of the traditional symbol matching approach. This allows us to more easily specialise and generalise our queries. For example, we could ask our database for all the "Education Services" in Southampton and it could return all the schools, colleges and universities even though nothing is explicitly typed as "Education Service" in the database.

The third ontology layer, "the application ontology", describes a more specialised ontology based on either a particular application or a third party view of the world. Here new concepts can be constructed from or mapped to concepts in the domain ontology. One might imagine in a flood disaster management scenario defining new concepts such as "emergency accomodation" from the concepts "school" and "hospital". Instances in the database can then be dynamically reclassified based on the axioms in the application ontology.

We hope that this stacked ontology approach will help our data be more adaptable for different users needs.

---

1 F1, F2 etc. are either form or function classes

4 Interoperability

Our initial motivation for studying ontologies was to enable better data interoperability.

4.1 Interoperability at the Schema Level

As we stated above ontologies provide a more intuitive interface to a database schema. If two datasets have their schemas exposed through an ontology then this could greatly help the data integration process. Mappings can be created between two ontologies creating a centralised view on heterogeneous databases (see figure 1).

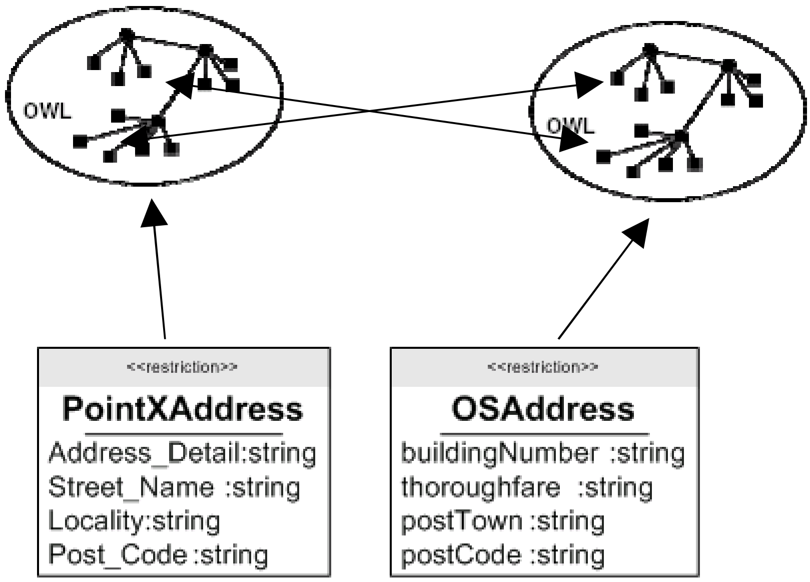

Fig. 1.

However, in reality it is likely that we will not be able to create simple mapping between schemas in this way. A good example of this would be address data. The definition and format of addresses is different in most organisations. However, having a formal definition of an address in an ontology may well help humans with their decision making processes in more manual integration efforts.

4.2 Classiciation allignment

Different organisations tend to have different views of the world. Often the meaning of a word can differ subtly or classification schemes can be at different levels of semantic grandularity. Figure 2 contrasts two fragments of classification schemes from the Ordnance Survey and Valuation Office.

Valuation Office	**Ordnance Survey**
School_and_Premises	EducationServices
LocalAuthoritySchool	School
IndependentSchool	PrimarySchool
	SecondarySchool
	PrivatePrimarySchool
	PrivateSecondarySchool

Fig. 2. Two taxonomic views of the world

Both classification schemes provide information on schools, but they are clearly structured differently. Here we can use a simple set of axioms (see figure 3) to provide semantic interoperability between the Ordnance Survey and Valuation Office data. However, this is a relatively simply example. Current research at Ordnance Survey is looking at how to best interoperate between two, potentially very large, ontologies where the mappings might not be quite as straightforward as those shown in figure 3.

We not propose that ontologies will solve all of our interoperability problems, but they are certainly an important part of the solution.

Valuation Office	**Ordnance Survey**
School_and_Premises	EducationServices
LocalAuthoritySchool	School
IndependentSchool	PrimarySchool
	SecondarySchool
	PrivatePrimarySchool
	PrivateSecondarySchool

Taxonomy Allignment

School $\equiv$ School_and_Premises
PrimarySchool $\sqsubseteq$ LocalAuthoritySchool
SecondarySchool $\sqsubseteq$ LocalAuthoritySchool
PrivatePrimarySchool $\sqsubseteq$ IndependentSchool
PrivateSecondarySchool $\sqsubseteq$ IndependentSchool

Fig. 3. The aligned taxonomies

5 Smart Queries

A more exciting application of ontologies would be inferring new information from combined datasets.

One might imagine a scenario where we have successfully used ontologies to integrate pollution data from the Environment Agency with topographic data from the

Ordnance Survey. We would now like to use this combined data to find topographic areas that are potentially at risk from pollution events. We could create an axiom of the type:

$$\text{RiskArea} \sqsubseteq \text{Area} \sqcap \exists\text{connectedTo}.(\text{Area} \sqcap \exists\text{contains}.\text{Pollutant})$$
$$\text{connectedTo}^{+} \sqsubseteq \text{connectedTo}$$

saying that all risk areas are those areas connected to areas that contain pollutants. “connectedTo” is a transitive property, and Pollutant is a concept defined in a seperate pollution ontololgy, which might contain axioms of the form:

$$\text{Organophosphate} \sqsubseteq \text{Pollutant}$$
$$\{\text{diazonin}\} \sqsubseteq \text{Pollutant}$$

One dataset will provide information about which areas contain which pollutants and another dataset will provide information about waterbodies and how they are topologically connected. If, say, “area1” could be inferred to be connected to “area2” and “area2” contains diazonin then we can infer that “area1” is of type RiskArea. This inference would be done at query time, unlocking “hidden” information in the data.

6 Ontologies and Spatial Databases

At Ordnance Survey we are interested in combining semantic technologies with spatial technologies to provide really power location based services (LBS).

Spatial databases allow one to use spatial functions and operators to answer queries like:

- “list all school zones crossed by this railway line”
- “find all pizza parlors within this area of interest”
- “return the 10 hotels which are closest to the airport, and the distance to each in miles”

So one can imagine that a simple example of combining spatial queries with semantics would be generalising the query “find all pizza parlors within this area of interest” to, say, “find all restaurants within this area of interest”.

A more interesting application, going back to the scenario of flood disaster management, trying to find all emergency accomodation in the event of a flood. One could define “potential emergency accomodation” as being all buildings that are within 10 miles of a hospital. The inference engine will figure out which objects in the database are buildings (as they may well be classified at a lower level as schools, churches etc.) and the spatial engine will then determine which of these are with 10 miles of a hospital. Objects in a database could then be dynamically reclassified as emergency accomodation according to these rules.

A skeptic might argue that finding all emergency accomodation might be done more easily using standard SQL with spatial operators. However, we expect it to be more useful to have the information about, for example, “emergency accomdation” captured in the portable and reusable form of an ontology.

Using currently technology is it not obvious[2] to see how such a spatial-semantic solution might be implemented and in [1] we will be looking at how it might be achieved.

Conclusions

In this paper we have discussed some of the potential uses of ontologies at Ordnance Survey, but there is still a lot of work that needs to be done before this can be done. It remains to be seen whether the current technology is mature enough to provide us with the solutions we require. When handling spatial data there is a clear need for OWL to understand spatial datatypes such as geometries and coordinates. Given that spatial data is notoriously hard to manage and that ontologies potentially enforce a highly normally database structure there is clearly much work that needs to be done to ensure efficient database implementations.

References

1. Goodwin, J., Lopez, X., Pollock, J.: Geographic System Grid Mediator: A Solution Design for Distributed and Heterogeneous Geospatial Data – work in progress

2 at least to the author