Tales of the River Bank,
First Thoughts in the Development of a Topographic Ontology

Glen Hart[1], Sarah Temple[1], Hayley Mizen[1,2]
glen.hart@ordnancesurvey.co.uk, hayley.mizen@students.plymouth.ac.uk,
sarah.temple@ordnancesurvey.co.uk
[1]Ordnance Survey of Great Britain,
Romsey Road, Southampton, Hampshire, England SO16 4GU
www.ordnancesurvey.co.uk
[2]University of Plymouth, Drake Circus, Plymouth, PL4 8AA
www.plymouth.ac.uk

ABSTRACT

This paper describes the initial thoughts and experiences of Ordnance Survey in the development of a topographical ontology. It describes the construction of a subset of a full topographical ontology, one restricted to inland water. The paper also describes our approach to the construction of such an ontology, one which attempts to maximise the ability to move shared concepts between ontologies. Lastly, it describes the issues raised when attempting to represent the ontology in DAML+OIL.

KEYWORDS: *Ontology, Data Interoperability, Semantic Interoperability, Topography, Inland water, DAML+OIL*

"I am not yet so lost in lexicography, as to forget that words are the daughters of the earth, and things are the sons of heaven"

Samuel Johnson

INTRODUCTION

The data that Ordnance Survey (Great Britain's national mapping agency) collects is being transformed from that suitable only to generate mapping, to discrete feature based information suitable for use in a wide range of digital applications. In parallel with the continual development and improvement of the information content that Ordnance Survey holds, there is a growing appreciation of the need to ensure this information is capable of being maximally interoperable with an end user's own information. This appreciation has naturally led us to investigate ontologies as a means to make our information more explicitly meaningful. Our end goal therefore is to ensure that Ordnance Survey's information can be understood and used by a third party system with minimal human intervention, and to establish a topographic ontology that can be used as part of a "semantic reference system" (Kuhn 2003). This paper outlines the initial phase of an exploratory investigation into the development of ontologies and research into how they may be used to enable information from heterogeneous sources to interoperate. Specifically, it describes our approach and initial experiences in constructing a subset of a full topographic ontology – one based largely around inland hydrology.

OVERVIEW

The research that we embarked upon is much wider than the work discussed here which concentrates on issues relating to the construction on an ontology. The overall scope of the work encompasses the understanding of not just how to construct ontologies but also how to represent instant data within databases and how to use multiple ontologies to perform cross domain tasks (Hart and Greenwood, 2003). The project structure is outlined in Figure 1.

Investigating the manner in which ontologies should be constructed and represented is being performed at OS with the production of a topographic ontology based around inland water. A scenario based around Water Quality Improvement that will use this ontology is to be developed at Oxford Brookes University, they will also be responsible for constructing a freshwater ecology ontology using the approach developed by the OS team. Ontology integrations and semantic translation will be investigated by University of Muenster using, and further developing, the task scenario developed at Oxford Brookes. Implementation of the instance databases formed around the task scenario will be performed by OS. Lastly, an independent ontology based around a scenario of freshwater leisure activities will be developed independently by the University of Brighton to provide a test ontology that may be used to explore how receptive the OS ontology is to adopting or exporting concepts from another ontology.

The core of the research centres around the development of ontologies and the approach we are adopting

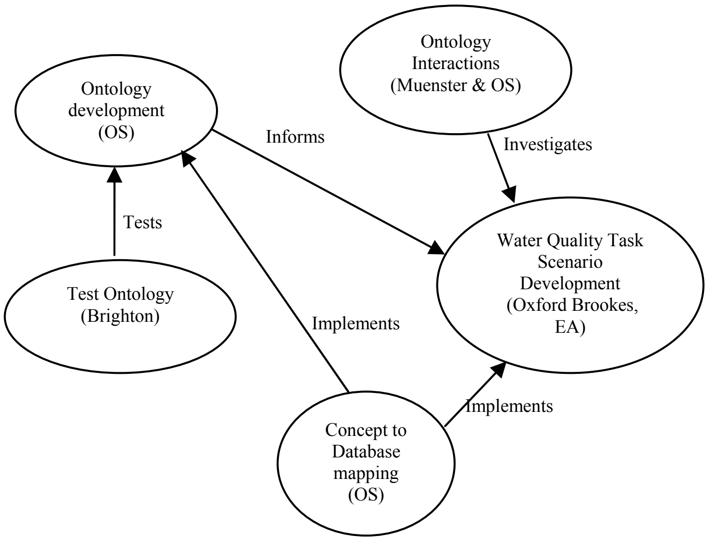

Figure 4 Research Activities

is to map concepts to instances based on the notion of disaggregated features comprising collections of components that have the potential to be distributed across databases whilst sharing common identity (Greenwood and Hart 2003).

The work described in this paper is focused upon the production of the inland water topographic ontology. Our initial aim was limited to constructing an ontology of inland water entities that are typically recorded by Ordnance Survey surveyors.

APPROACH

When constructing ontologies we are doing so as a research exercise: to investigate how they may be constructed and represented. Although the work is research it is also our intention to make the scenarios that the research is based around as realistic as possible. Hence the method adopted by the project as a whole has been to develop scenarios based around real world problems in real world situations, hence the choice of Water Quality. Within Ordnance Survey we therefore wanted our ontology to reflect the manner in which Ordnance Survey surveyors recognise and classify the world around them. However, we did so with modification as it became apparent during the development of the ontology that a small part of the classification scheme was not properly representative of reality. We also chose to initially develop a domain ontology as opposed to one that represented the data as held within the main topographic database. This choice was because the database representation lost a lot of the implicit information defined by the survey specification and was essentially a geometry specification. Thus although we will need to produce a data ontology to represent this model in order to support instance data it was felt that the initial work would be more fruitful if conducted at the domain level with its richer set of concepts.

Man and Machine

Ontologies serve two masters man and machine. For us the initial importance was on capturing an accurate description of the domain without too much concern for any limitations imposed by needing to use languages designed to be machine readable. Thus crudely the development of the ontology could be divided into two stages:

1. Development of idealistic[13] ontology;
2. Translation of the ideal ontology into a machine readable form.

In reality these two stages proceeded not sequentially but in parallel, with the second task starting shortly after the first.

To express the ontology in machine readable form we chose to use the ontology language DAML+OIL and the OilEd editing tool. This choice was made largely on the basis of the popularity that this language currently has and the tools available for it. There was also a pragmatic acceptance that any deployment of semantically enabled services in the near future were likely to be built using OWL the W3C ontology language which is based almost entirely on DAML+OIL[14].

DEVELOPING THE IDEALISTIC ONTOLOGY

In constructing our "ideal ontology" we set ourselves a number of constraints. These were all designed to maximise the ability to migrate concepts into other ontologies and to accept concepts from other ontologies.

Gruber neatly identifies many important characteristics that a good ontology should possess (Gruber, 1995). These may be summarised as:

- *Clarity*, the ability to effectively and accurately communicate meaning.
- *Coherence*, this stresses that concepts should be defined consistently and that the description (natural language form) should agree with the formal component.
- *Extendibility*, It should be extensible, enabling new terms to be easily added and concepts allowed to evolve.
- *Minimal encoding bias*, essentially meaning that ontologies should maintain a respectful separation

[13] Ideal in the sense that it is a true reflection of the way we understand the domain without making comprises forced by tool and language limitations.

[14] OWL itself was not chosen at the time as it had neither stabilised at that point nor did we know of any tool support. However, at the time of writing we are now considering the move to OWL.

from application implementations.
- *Minimal ontological commitment*, the ontology should only contain the minimum information necessary to support it's intended use. This strategy is designed to maximise the extensibility and reusability of the ontology.

These characteristics are useful guides and good measures of an ontology. An important aspect of what we wanted to achieve was extendibility, as well as the ability to provide definitions that could be easily absorbed by other ontologies. Gruber not only identifies extendibility, which supports our design criteria, but also minimal ontological commitment. We have tried to support these criteria by introducing two constraints on the way we build our ontologies:

- First, only introducing terms that are commonly used in the vernacular, by avoiding the introduction of new or invented terms to resolve particular issues of clarity the ontology is more likely to be interpretable by others. We used Ordnance Survey documentation as a beginning point for our ontology as it is Ordnance Survey's duty to observe and report, and we do so using the vernacular. It is not a specialist's view, but that of a generalist.
- Second, to limit the depth of hierarchies wherever possible. We are theorising that by using taxonomic relationships sparingly we will maximise the possibilities of being able to share concepts across ontologies. It also reflects our belief that the proper construction of ontologies will require the development of a new set of skills, certainly drawing on existing methodologies such as object-oriented design which is highly hierarchical, but also adopting new methods where appropriate.

Content Structure

Ontologies have been described as a "structured collection of defined objects" (Kuhn, 2001), although we would modify this to replace "objects" with "concepts". To fulfil this design purpose ontologies should therefore comprise at least the following for each concept: a natural language definition, attributes of the concepts, and all relationships between all concepts in the ontology (which will realise a semantic net). This last component of the ontology structure we believe should be the primary concern in all ontologies as this provides the links between concepts, as well as differentiating between them. Five main relationships were identified and then divided into sub-categories during the construction of the ontology:

- taxonomic,
- synonym,
- topological,
- mereological and
- affordance.

The process of generating sub-categories was by its nature evolutionary and led to periods where new sub-categories were added followed by periods of consolidation, rationalisation and reduction.

The relationships and their sub-categories as currently recognised within our ontology are summarised in Table 1 below.

Table 1 Relationships and Sub-categories

Relationship Type	**Sub-categories**	**Definition/Examples**
Taxonomic		Generalisation and specialisation relationship. E.g. A duck-pond is a special type of pond.
	Is a	One concept is a specialisation of another, for example: , duck-pond is a pond, Salt lagoon is a lagoon, a weir is a dam
	Is an instance of	A real world phenomenon is an instance of a concept. For example a The William Roy Building (OS HQ Building) is an instance (realisation) of the concept Building. Northern England is an instance of locality.

Relationship Type	**Sub-categories**	**Definition/Examples**
Synonym		The use of an alternative term for a concept
	Same as	Used for “pure” synonyms where terms are unconditionally interchangeable. For example river stretch and river reach.
	Same as and Local to	Used for synonyms where the use of one term is conditional on location For example a Loch is the same as lake and local to Scotland. Beck is the same as stream local to Northern England, sike is the same as stream and local to the Pennines; haughland is the same as meadow local to Scotland
	Same as in Language	Used for synonyms where the use of one term is conditional on language. For example an afon is the same as river in Welsh. It can be argued that many “local to” terms are in fact language based. The differentiator that we have applied is to determine whether a term originating from another tongue has been absorbed by the natural language used by the ontology or not. If it has then the condition is local to, if not it is language. So Loch is local to not language because Loch is now a well established term within English. E.g. ynys is the same as island in Welsh; nant is the same as stream in Welsh; llyn is the same as lake in Welsh
Topological		Spatial relationships
	Next to	Indicates one concept is always found next to another. Examples: Saltmarsh is next to estuary; foreshore is next to land; agricultural drain is next to agricultural land;
	Bounded by, Bounds	A specialisation of the relationship “next to”, which also implies confinement, and its inverse relationship. Examples: bank bounds water; ait bounded by river; silt lagoon bounded by artificial banks
	Contains, Contained in	Used when one concept is within the specified limits of another, and its inverse relationship. Examples: tarn is contained in upland, salt lagoon contains salt bed; oxbow lake is contained in floodplain; cataract is contained in river; sluice gate is contained in sluice
	Empties into, Feeds from	Examples: river mouth empties into sea area, river feeds from channel(watercourse), canal feeder empties into canal;
	Built across	Examples: bridge built across river; dam built across river
Mereological		Part / whole relationships such as a school comprises (has parts the) school buildings and grounds. Sub-categories taken from Tomai and Kavouras (2003).
	Part of	Examples: Riverbed part of river; channel(passage) part of bay; oxbow part of river
	Has part	Examples: River has part riverbank; canal(navigation) has part lock; cascade(ornamental) has part waterfall
Affordance	Affords May Afford	Affordances are deemed important to our model not just because they help to discriminate between concepts but also because the things that are afforded are often of interest to data collectors and ultimately end users. We differentiate between affordances that are necessary to define a concept and those that a concept may possess. This is seen as important due to imperfect knowledge we have of the world. Affordances in particular are often inferred rather than known

Relationship Type	Sub-categories	Definition/Examples
		as fact. Examples are: flood bank affords flood control, reservoir may afford power generation; storm drain affords drainage; canal(irrigation) affords irrigation; sluice affords flow control; inlet may afford shelter for ships.

Issues

Taxonomy: Taxonomies are inevitably hierarchical, and relationships are inherited down the hierarchy. Given our self-imposed constraint concerning the depth of hierarchies we therefore necessarily limited the use of this construct. It was used only where it was felt to be meaningful, e.g. a fishing pond "is a" pond. We did not use it to introduce high level classes such as Natural and Man-made, even though these exist within the current Ordnance Survey classification system, simply because they provide no useful function whilst introducing issues of categorisation: is a river bank that has been modified natural or man-made or both? Size modifiers were also introduced to this relationship where it was thought that the two concepts represented different sizes of the same concept, e.g. for a lake and a pond the only real differentiator is that typically lakes are larger than ponds

Synonym: There were found to be few true synonyms. The majority had some conditional qualifier. It can be argued that many "local to" terms are in fact language based. The differentiator that we have applied is to determine whether a term originating from another tongue has been absorbed by the natural language used in the ontology or not. If it has then the modifier is "local to" if not it is "language". So loch is "local to" not "language" because loch is now well established within English. By comparison afon, the Welsh for a river, has not been adopted within English.

Topology & Mereology: We found few direct issues with these relationships. Granularity was a problem with some topological and mereological relationships. Was a river bounded by the bank or the field? Was the bank part of the river or bounding it? However, these issues were normally resolved because the root cause was an inconsistent focus on granularity rather than any fundamental problem.

Affordance: We have attempted to differentiate between affordances that are necessary to define a concept and those that a concept may possess. Relationships such as affordances are deemed important to our model not just because they help to discriminate between concepts but also because the things that are afforded are often of interest to data collectors and ultimately end users. We find affordances, sometimes seductively easy to apply, only result in poorly defined affordances. For example we could say that rivers afford habitat, but what does this really mean? For an organisation like Ordnance Survey whose primary concern is topographic survey the term habitat has no use. Therefore in accordance with Gruber's criterion of minimal ontological commitment we should exclude it.

EXPRESSING THE IDEAL ONTOLOGY IN DAML+OIL

The "ideal ontology" was then expressed in DAML+OIL using OilEd, a freeware tool produced by Manchester University to aid the development of ontologies (van Harmelen, et al, 2001). As expected it was not possible to fully express the ontology as we would have liked; compromises had to be made. We developed mappings from the ideal relationships to DAML+OIL expressions as shown in Table 2 below.

Table 2 Using DAML+OIL to Express Relationships

Relationships	**DAML+OIL**
Taxonomic	
Is a	SubClassOf, e.g. <rdfs:subClassOf> <daml:Class rdf:about="file:/P:/Iridium/ont/field+hydro.daml#Ditch"/> </rdfs:subClassOf
Is an instance of	Type, e.g. <rdf:type> <daml:Class rdf:about="file:/P:/Iridium/ont/field+hydro.daml#Locality"/> </rdf:type>
Synonym	
Same as	SameClassAs, e.g. <daml:sameClassAs> <daml:Class rdf:about="file:/P:/Iridium/ont/field+hydro.daml#Reach"/> </daml:sameClassAs> Strictly this is incorrect. This relationship states that two different concepts share the same set of instances and is not the same as stating the concepts are the same concept. However, as DAML+OIL has no way of expressing a true synonym relationship this was used. OWL Full does differentiate having SameAs and EquivalentClass although OWL DL again does not support true synonyms.
Same as and Local to	SameClassAs and two restrictions containing the property "localTo", one which is "hasClass" and one which is "toClass" e.g. <daml:sameClassAs> <rdfs:Class> <daml:intersectionOf> <daml:List> <daml:first> <daml:Class rdf:about="file:/P:/Iridium/ont/field+hydro.daml#Confluence"/> </daml:first> <daml:rest> <daml:List> <daml:first> <daml:Restriction> <daml:onProperty rdf:resource="file:/P:/Iridium/ont/field+hydro.daml#language"/> <daml:toClass> <daml:Class rdf:about="file:/P:/Iridium/ont/field+hydro.daml#Welsh"/> </daml:toClass> </daml:Restriction> </daml:first> <daml:rest> <daml:List> <daml:first> <daml:Restriction> <daml:onProperty rdf:resource="file:/P:/Iridium/ont/field+hydro.daml#language"/> <daml:hasClass> <daml:Class rdf:about="file:/P:/Iridium/ont/field+hydro.daml#Welsh"/>

Relationships	DAML+OIL
	</daml:hasClass> </daml:Restriction> </daml:first> <daml:rest rdf:resource="http://www.daml.org/2001/03/daml+oil#nil"/> </daml:List> </daml:rest> </daml:List> </daml:rest> </daml:List> </daml:intersectionOf> </rdfs:Class> </daml:sameClassAs>
Same as in Language	SameClassAs and two restrictions containing the property "language", one which is HasClass and one which is "toClass"
Topological	
Next to	Used as the property of a restriction, which is a subclass of a class (usually hasClass), e.g. <rdfs:subClassOf> <daml:Restriction> <daml:onProperty rdf:resource="file:/P:/Iridium/ont/field+hydro.daml#nextTo"/> <daml:hasClass> <daml:Class rdf:about="file:/P:/Iridium/ont/field+hydro.daml#River"/> </daml:hasClass> </daml:Restriction> </rdfs:subClassOf>
Bounds, Bounded by	Used as the property of a restriction, which is a subclass of a class (usually hasClass), e.g. <rdfs:subClassOf> <daml:Restriction> <daml:onProperty rdf:resource="file:/P:/Iridium/ont/field+hydro.daml#boundedBy"/> <daml:hasClass> <daml:Class rdf:about="file:/P:/Iridium/ont/field+hydro.daml#ArtificialBanks "/> </daml:hasClass> </daml:Restriction> </rdfs:subClassOf>
Contained in, Contains	Used as the property of a restriction, which is a subclass of a class (usually hasClass).
Empties into, Feeds from	Used as the property of a restriction, which is a subclass of a class (usually hasClass).
Built across	Used as the property of a restriction, which is a subclass of a class (usually hasClass).
Mereological	
Part of	Used as the property of a restriction, which is a subclass of a class (usually hasClass).
Has part	Used as the property of a restriction, which is a subclass of a class (usually HasClass).
Affordance	
Affords	Used as the property of a restriction, which is a subclass of a class (usually hasClass).
May Afford	Uses "affords" as the property of a restriction, which is a subclass of a class where

<table>
<tr><th>Relationships</th><th>DAML+OIL</th></tr>
<tr><td></td><td>the cardinality is described as “min0”. This was as close a match to “may” as we could find using DAML+OIL. E.g.
<code><rdfs:subClassOf></code>
<code> <daml:Restriction daml:minCardinalityQ="0"></code>
<code> <daml:onProperty</code>
<code>rdf:resource="file:/P:/Iridium/ont/field+hydro.daml#contains"/></code>
<code> <daml:hasClassQ></code>
<code>   <daml:Class</code>
<code>rdf:about="file:/P:/Iridium/ont/field+hydro.daml#Stream"/></code>
<code> </daml:hasClassQ></code>
<code> </daml:Restriction></code>
<code> </rdfs:subClassOf</code></td></tr>
</table>

In addition to the issues indicated in Table 2 a number of other issues were also encountered of which the most important were:

- OilEd does not allow multiple definitions of a class name (Bechhofer el al, 2001), instead different concepts with the same name can only exist within different domains (or namespaces). In our experience this does not reflect reality; for example, simply within the domain of inland water the term channel is used four different ways, pool two, and so on. It is clear that context plays an important aspect in differentiating between like named concepts, but the use of namespaces is too blunt an instrument to provide the necessary level of discrimination.
- Concepts such as land and surface are potentially important, but we are unclear as to how we should define them and what relationships they have with other concepts. For example is there a single instance of land surface – the surface of the Earth, is it a single concept or is it a general concept with many instances?
- The manner in which geographic things have been classified is often not orderly and reflects the *ad hoc* nature of many of our common terms. Consider river and stream, generally we consider a river to be larger than a stream but it is clear that no precise definition exists and functionally there is little to differentiate them with geography. Languages such as DAML+OIL have no way to express fuzzy relationships such as rivers “are *typically* larger than” streams.
- In DAML+OIL we also encountered a problem with inherited relationships where the relationship is optional at a high level but not possible at the lower level. For example a possible affordance for a pond is fishing, it is mandatory for a fishing pond, but will not be possible for a heavily polluted pond. We have used to the “may afford” relationship, as shown in table 2, but inference engines for DAML+OIL cannot support this relationship.
- The use of DAML+OIL a description logic has been difficult simply because this means of reasoning is not familiar to us and has often conflicted with our common sense, especially when considering the use of terms such as “and” and “or” in logic and English. Often the use of “and” in an English statement requires the use of an “or” in logic – a problem that has caused us many problems. It is also clear that we are not alone in this, and that it seems description logics will be easier for machines to digest than people. This could be a serious inhibitor to their successful adoption. Extending description logics with rules bases could provide at least a partial solution. It is noted that some are already attempting to implement OWL in rule based systems (Voltz 2003), and such work may provide the required base for this extension.

CONCLUSION

There are clearly questions around whether DAML+OIL is sufficiently descriptive to express what we believe we need, though in part there may exist some solutions within DAML+OIL that our inexperience has as yet failed to reveal. It also seems apparent that the fuzzy world of geography presents some challenges that we are as yet unable to resolve, and therefore in the immediate future some compromises are required.

REFERENCES

Bechhofer, S. Horrocks, I. Goble, C. and Stevens, R. (2001) OilEd: a Reason-able Ontology Editor for the Semantic Web. Proceedings of KI2001, Joint German/Austrian conference on Artificial Intelligence, September 19-21, Vienna. Springer-Verlag LNAI Vol. 2174, pp 396--408..

Gibson, J. J. (1986). The Ecological Approach to Visual Perception. Hillsdale, NJ, Lawrence Erlbaum.

Greenwood, J. and Hart, G. (2003). Sharing Feature Based Geographic Information - A Data Model Perspective, GeoComputation 2003

Gruber, T. R. (1995). Towards Principles for the Design of Ontologies used for Knowledge

Sharing, International Journal of Human-Computer Studies, 43:907-928, 1995.

Hart, G and Greenwood (2003), J. A Component Based Approach to Geo-Ontologies and Geodata Modelling to Enable Data Sharing, Agile 2003

Horrocks, I, Voltz, R, Decker, S, Grosof, B, Motik, B. (2001) Rule Language, WonderWeb: Ontology Infrastructure for the Semantic Web.

Kuhn, W. (2001) Ontologies in Support of Activities in Geographic Space, University of Muenster, Germany

Kuhn, W. (2003) Semantic Reference Systems. International Journal of Geographical Information Science, 17(5): 405-409.

Tomai, E. & Kavouras, M. (2003) "Pivotal Issues in Designing Geographic Ontologies", COSIT-03 - Workshop on Fundamental Issues in Spatial and Geographic Ontologies", COSIT-03, Ittingen, Switzerland, 24-28

van Harmelen, F, Patel-Schneider, P. F., and Horrocks, I, (editors) (2001) Reference description of the DAML+OIL ontology markup language, www-rdf-logic

Voltz, R, Decker, S and Oberle, D. (2003) Bubo – Implementing OWL in rule-based systems