# Using Structured English to Author a Topographic Hydrology Ontology.

Glen Hart, Catherine Dolbear, Katalin Kovacs, John Goodwin, Sheng Zhou

Ordnance Survey of Great Britain, Romsey Road, Maybush, Southampton SO31 9AJ England
{Glen.Hart, Catherine.Dolbear, Katalin.Kovacs, Sheng,Zhou, John.Goodwin}@ordnancesurvey.co.uk


## 1 Introduction

Ordnance Survey, Great Britain's national mapping agency, is currently in the process of building a topographic ontology to express the content of its topographic database with the aim of enabling the semi-automation of data integration, product repurposing and quality control. We are devising a methodology that enables domain experts working with ontology engineers to construct ontologies that have both a conceptual, human readable form, and a computation form that is interpretable by machines (Mizen, et al, 2006). A key part of the methodology is that it enables the first stages of ontology authoring to be conducted using a restricted form of English which enables the domain expert to be in charge of the authoring process and other domain experts to understand and verify the content. The computational aspect, expressed using OWL (W3C, 2004), is treated as an assembler code version of the conceptual ontology.

This paper introduces Rabbit, the constrained form of English that we are developing and gives examples of its use in constructing a hydrology ontology and how it maps to OWL.

## 2 Related research

Ever since OWL was conceived there have been concerns that its form makes it inaccessible to those without a good understanding of logics (Horridge, 2006). It is difficult for domain experts to use OWL to author or validate ontologies. This in turn creates a serious impediment to the adoption of OWL and semantic web. There have been a number of attempts to resolve this issue through the creation of grammars for OWL that attempt to make it more understandable. Such grammars include the Manchester Syntax (Horridge, 2006) which attempts to replace the abstract symbology of description logic. Whilst this is significantly more readable than the pure mathematical representation, the average domain expert will still struggle to understand what it means.

Other approaches are to use constrained forms of English, examples being ACE (Fuchs et al, 2005) and Processable English (PENG) (Schwitter et al, 2002) both of which provide grammars based on constrained English to represent First Order Logic (FOL) and both have now Description Logic (DL) subsets (Kaljurand et al, 2006) and (Schwitter et al, 2006). These do provide significantly more readable representations. For example in PENG DL simple statements such as "France is a country." can be made in. As these grammars are representations of OWL they also allow more complex statements to be made. These more complex forms also begin to sound a bit unnatural but are still readable: “If X has Y as a topping then X has Y as an ingredient and X is a pizza and Y is a pizza topping and Y is a topping of X.”

Lastly, it is worth mentioning that all these approaches are limited to representing only what can be stated in DL (or FOL). They will also reflect any optimisations or modelling “tricks” that an ontology engineer might apply and which whilst necessary for either efficient reasoning or accurate modelling will obscure the ontology from a domain expert’s point of view.

# 3 Rabbit – motivation and design principles

Our research has been focused on developing a language that overcomes some of the limitations described above. We have named this language Rabbit, after Rabbit in Winnie the Pooh, who was really cleverer than Owl. We have involved domain experts from the outset in the core language design decisions. It is a language in its own right, rather than merely a syntactic veneer on OWL. This means that it contains constructs such as “typically” (meaning an unquantified but significant majority but not all) that are needed by the domain expert, even though they cannot be expressed in OWL. These constructs are included to ensure that domain knowledge is not lost, even if it cannot be fully exploited.

The fundamental principles underlying the design of Rabbit are:

- Maximal Expressibility;
- Be based on a formal grammar;
- To hide ontology engineering optimisations and modelling tricks;
- To be independent of any specific domain.

## 3.1 Principle 1: Expression

Rabbit focuses on keeping the sentences natural sounding and where possible short: for example using “A River flows into a Sea” rather than the OWL-like “All Rivers flow into some Seas”. To make it easier for the domain expert to produce, several defaults are assumed for example “or” is assumed to be an exclusive or, and all concepts are assumed to be disjoint, unless specified otherwise. As domain experts tend to think in terms of the closed world assumption, these defaults are more natural for them. The open world assumptions of OWL have to be explicitly closed down when such Rabbit sentences are converted to OWL. This leads to relatively simple expressions in Rabbit having complex expression in OWL. For example the Rabbit statement:

A river flows into a sea or a lake or a river or a reservoir.

would result in the following OWL:

River -> flowsInto some (Sea or Lake or River or Reservoir) and not (flowsInto some Sea and flowsInto some Lake) and not (flowsInto some Lake and flowsInto some River) and not (flowsInto some River and flowsInto some Reservoir) and not (flowsInto some Reservoir and flowsInto some Sea) and not (flowsInto some Sea and flowsInto some River) and not (flowsInto some Lake and flowsInto some Reservoir).

Use of some constructs such as “typically” can be represented in Rabbit but cannot be represented in OWL. They are included in Rabbit to ensure that the domain can be captured as accurately as possible and to ensure that a better quantification of information loss can be made.

## 3.2 Principle 2: Grammar

Concepts in Rabbit may comprise one or more words such as “River” or “River Stretch”. Homonyms are differentiated with the addition of a disambiguation term following the concept name in brackets: Pool (River). This is converted to OWL as Pool_River and an rdf:label of “Pool”.

Rabbit has a few predefined relationships such as “Is a kind of” to introduce super and sub-class relationship. Mostly authors will define their own relationships such as “flows into”. Relationships may be modified using phrases such as “only”, “typically”, “at least” and “does not”. For example, “a braided river stretch flows in at least 2 channels”.

## 3.3 Principle 3: Interaction between domain expert & knowledge engineer

An important aspect worthy of a little more discussion is the interaction between the domain expert, the knowledge engineer and the translation process in “compiling” Rabbit into OWL. Translating Rabbit to OWL will have three significant effects on the modelling process. First, it is likely to expose flaws in modelling which will need correction in Rabbit. Second, it gives the potential to include optimisations in the OWL version (for reasoner efficiency) which need not be reflected in the Rabbit representation lest they obscure the clarity of the ontology. Third, the knowledge engineer may wish to use “modelling tricks” (again to help the OWL form) which need not necessarily be explicitly represented in Rabbit. An example would be the representation of transitive properties. These are interesting, because if we took “has part” as an example, we could say that “a car has a part engine” and that “an engine has a part piston” and so on. Normally if we asked what were the parts of a car, then we would expect a list of the major components. However, if “has part” is defined in OWL as transitive, then a reasoner would return all parts. This is probably not what most people would expect. One solution (W3C, 2005l) is to define the “hasPart" property as transitive and then to define a subproperty “hasDirectPart” which is not transitive. This enables a choice to be made at query time as to whether all parts or just “direct parts” are selected. Rabbit has adopted this convention in that all transitive relationships in Rabbit translate to OWL by defining a property and subproperty pair.

## 3.4 Principle 4: Domain independence

Rabbit has been developed over the course of authoring an ontology in the domain of hydrology and has been further tested with the buildings and administrative geography domains. We believe that it now contains most of the necessary constructs to author ontologies in any domain, including expressions beyond OWL or OWL 1.1.

# 4 Conclusions and future work

Currently, we are finalising the grammar and translation and are also working closely with the University of Leeds on the development of a tool to assist a domain expert to input an ontology in Rabbit, following our ontology authoring method. An important aspect of this is the implementation of a tool subset that supports the ontology author in the construction of Rabbit sentences and their automatic conversion to OWL. In the near future we will be conducting human subject testing of the grammar to ensure that the statement constructs are interpreted with a significant degree of accuracy. Here we do not expect to be able to design a completely unambiguous grammar; whenever using a natural language-like construct, one has to accept some misinterpretation. However, we do expect such work to show significant improvement in interpretability by domain experts over existing alternatives and to enable us to identify problem areas for resolution.

In conclusion we have introduced the work we are performing to develop a language to enable domain experts to better author and interpret ontologies. We see the advantages of Rabbit over other methods and DL syntaxes as being:

- Greater clarity of expression due to the involvement of the domain expert both in the development of Rabbit, its authoring and use.
- The ability to hide certain modelling complexities and optimisations.
- The encouragement, through its grammar, to produce short sentences that are more easily interpretable.

We would also stress that another important aspect of Rabbit is that authoring should be supported by a tool as an integrated part of any ontology authoring methodology.

# 5 References


W3C, 2004, OWL Web Ontology Language Guide, W3C Recommendation, http://www.w3.org/TR/owl-guide/

MIZEN, M., HART, G and DOLBEAR, C, 2006, A Two-Faced Approach to Developing a Topographic Ontology, - P227-231 Proceedings of the GIS Research UK 14th Annual Conference ISBN 0853582262

HORRIDGE, M., DRUMMOND, N, GOODWIN, J, RECTOR, A, STEVENS, R, WANG, H, 2006, The Manchester OWL Syntax, OWL-ED Workshop

FUCHS, N.E., HÖFLER, S., KALJURAND, K., RINALDI, F. and SCHNEIDER, G., 2005, Attempto Controlled English: A Knowledge Representation Language Readable by Humans and Machines. In Norbert Eisinger and Jan Ma'luszy´nski, editors, Reasoning Web, First International Summer School2005, Msida, Malta, Revised Lectures, number 3564 in Lecture Notes in Computer Science. Springer

SCHWITTER. R, 2002, English as a formal Specification Language. Proceedings of the Thirteenth International Workshop on Database and Expert Systems Application (DEXA 2002), pp. 228-232

KALJURAND K. and FUCHS. N.E. 2006, Mapping Attempto Controlled English to OWL DL. In 3rd European Semantic Web Conference. Demo and Poster Session, Budva, Montenegro

SCHWITTER R., TILBROOK, 2006, M. Let's Talk in Description Logic via Controlled Natural Language, in: Proceedings of the Third International

Workshop on Logic and Engineering of Natural Language Semantics (LENLS2006) in Conjunction with the 20th Annual Conference of the Japanese Society for Artificial Intelligence, Tokyo, Japan, June 5-6, pp. 193-207,)
W3C, 2005, Simple part-whole relations in OWL Ontologies, W3C Editor's Draft 11, http://www.w3.org/2001/sw/BestPractices/OEP/SimplePartWhole/simple-part-whole-relations-v1.5.html

**Biography**

Glen Hart is the Principal Research Scientist working within Ordnance Survey's Research Labs. His research interests include spatial data modelling and integration geosemantics and vernacular geography.

Cathy Dolbear is a senior research scientist within Ordnance Survey's Research Labs with a particular interest in the application of the semantic web to GI Science. She is currently working on methods to enable semantic queries to be executed over spatial databases.