(3) OIL has a ***layered architecture***, avoiding the temptation to throw everything into the core language, mixing up features that cannot be reasoned over with those that can be. Thus the limits are clear and explicit. For example, OIL-Core includes all the mainstream primitives and is entirely computationally viable; OIL-Standard includes instances so they can be included but reasoning over them is not supported computationally. OIL does have limitations: it currently does not include default reasoning, arbitrary axioms, or second order expressivity. The idea is that these are included as layers, carefully preserving the bounds of what can be reasoned with but without excluding features that are desirable.

OIL is lightweight enough to represent simple taxonomies or frame-like ontologies such as GO without obstruction. It is also expressive enough to present logic-based models such as the TAMBIS Ontology, and all points in between. Some parts of the ontology can be simple, others complex; moreover, the language offers the ontologist an evolutionary development path whereby they can progressively introduce more expressive constructs as required. Although OIL is a powerful language it isn't necessary to use all the expressive power.

OIL has been developed as an international collaboration as part of an effort to realize the W3C vision of a semantic web through the explicit representation of their meaning [3]. The issues to be addressed by the biology E-Science community are a version of issues to be addressed by the whole web community—specifically, how to move the web from one where information is machine readable by humans to one where information is machine processable by intelligent services. Current OIL activity is focused on:

- ***Example applications:*** Manchester has already produced a version of the Gene Ontology in OIL and has developed the second generation TAMBIS ontology in OIL.
- ***Tools*****:** The FaCT reasoner has an implementation efficient enough to make reasoning over OIL empirically tractable. Other tools under development include OIL editors (OILEd at Manchester, OntoEdit at Karlesruhe and an extension of Protégé 2000 at Stanford), and the adaptation of ontology integration tools, such as Chimaera at Stanford, as well as parsers between OIL and RDF/XML etc;
- ***Project proposals and consortia*****:** An EU Network of Excellence and an EU IST FET-O proposal have been submitted, with over 20 commercial partners as supporters, including major pharmaceutical companies and biotech web portal information providers.
- ***WWW metadata language standardisation:*** The OIL effort is a major contributor to the DARPA Agent Markup Language initiative (DAML) in their attempt to define an ontology language, DAML-ONT. In addition, the W3C have been developing an extension to RDF to include logical inferencing (RDF-Logic); OIL's RDFS mapping gives RDF a semantics that it is otherwise absent. At the time of writing, the three communities—OIL, W3C and DAML—have come together in a joint DAML Language Committee to produce a universal, scalable yet technically sound language for describing knowledge on the web.

At the workshop I will motivate for the necessity of ontologies to interoperability between bioinformatic resources. I will present the OIL language, what it can and cannot do giving examples from the GO and TAMBIS OIL representations, and report on OIL's current status and future plans. For more discussion and motivation, and for papers that give informal and formal descriptions and a grammar specification, see http://www.ontoknowledge.org/oil/.

References

[1] Ashburner et al *Gene Ontology: Tool for the Unification of Biology* Nature Genetics 25: 25-29 (2000)
[2] Baker PG, Goble CA et al *An ontology for bioinformatics applications* Bioinformatics 15(6) 510-520 (1999)
[3] Berners-Lee T, *Weaving the Web*, Orion Business (1999)
[4] Calvanese D, *Reasoning in expressive description logics* in Handbook of Automated Reasoning. (1999).
[5] Fensel D, Horrocks I et al *OIL in a Nutshell* Proc EKAW-2000, LNAI http://www.ontoknowledge.org/oil/
[6] Horrocks I and Sattler U *A description logic with transitive and inverse roles and role hierarchies* Journal of Logic and Computation 9(3):385-410, (1999).
[7] Karp PD *An Ontology for Biological Function Based on Molecular* Bioinformatics 16(3):269-85. (2000)
[8] Karp PD et al *XOL: An XOL-based ontology exchange language* http://www.ai.sri.com/~pkarp/xol/
[9] Stevens R, Goble CA et al *Ontology-based Knowledge Representation for Bioinformatics* Briefings in Bioinformatics *to appear*
[10] Baker PG et al *TAMBIS: Transparent Access to Multiple Bioinformatics Information Sources, An Overview* Proc ISMB 98 (1998)
[11] Goble CA *Supporting Web based Biology with Ontologies* in Proc ITAB-2000 (2000)

OIL: A Language for Representing and Sharing Ontologies

Carole Goble[(1)], Ian Horrocks[(1)] and the OIL Consortium[(2)]

(1) The Department of Computer Science,
The University of Manchester,
Oxford Road,
Manchester, M13 9PL, UK
<carole/horrocks>@cs.man.ac.uk
http://img.cs.man.ac.uk/

(2) The University of Manchester, Vrije Universiteit Amsterdam Nederland, AIdministrator Nederland, Institute AIFB, University of Karlsruhe, Stanford University, Bell Labs Research, Massachusetts Institute of Technology and others at
http://www.ontoknowledge.org/oil/misc.shtml

There is a temptation to think that exchanging information depends on sharing a common syntax. However, information exchange needs more than syntax—even when two applications use XML as their interchange format how can we be sure that they use the same vocabulary and that the words in the vocabulary mean the same thing? *Terminologies* provide shared and common vocabularies of a domain, so search engines, agents, curators, authors and users can communicate. However, using a common vocabulary is no good unless everyone means the same thing. *Ontologies*, "a specification of a conceptualization of a domain", provide a consistent, shared and common understanding of a domain that can be communicated across people and applications (e.g. agents). Consequently, ontologies play a major role in supporting information exchange and discovery in E-Commerce and E-Science [11]. Conceptual database schemata can be viewed as a form of ontology.

Ontologies in bioinformatics are used as repositories of potentially reusable biological knowledge, as a common framework for multi-database queries [10] or as controlled vocabularies for genome annotation [1]. The success of many biological databases depends upon their fidelity to, and the clear communication of, their ontologies to prevent errors of data entry and interpretation. E-commerce uses ontologies to describe service provision and contractual information.

Ontologies take a variety of forms, from classification schemes such as the Gene Ontology [1], to frame-based knowledge models such as EcoCyc [7] and logic-based models such as the TAMBIS Ontology [2]. Despite all these forms they all include a vocabulary of terms and some specification of the meaning of the terms.

A prerequisite for a widespread use of ontologies is a joint standard for their description and exchange. The ontology needs to be specified in some language that can comfortably cater for the variety of forms. The Ontology Inference Layer (OIL) is a new language proposed as a knowledge representation language for the web and web-based applications [5].

(1) OIL brings together ***the best of frames and description logics*** bridging the two communities.

- It provides the *modelling primitives* commonly used in frame-based ontologies, based on but considerably extending XOL [8] which in turn was based OKBC-Lite;
- It has simple, clean and *well-defined semantics* based on those of Description Logics (DLs) [4];

OIL has the "intuition" and familiarity of frame systems but also a clear definition of what the language really means and the ability to compose classes and slots to form new expressions. For example, we are able to define sufficiency conditions for concepts as well as necessary conditions. Moreover, we are able to use the *automated reasoning support* to automatically compute subsumption (isa) relations between concepts, or to check the consistency and coherency of the classification and its concepts. This turns out to be really useful when collaboratively building substantial ontologies or when ontologies are reused or merged [9]. This reasoning is provided in a computationally efficient manner by state of the art DL classifiers such as FaCT [6].

(2) OIL has a machine-readable ***syntactic encoding in the languages of the web*** including the syntactic ones (XML DTD, and an XML Schema definition) and the semantic metadata ones (an RDFSchema definition). In fact an ontology in OIL can be used by an agent that is not OIL aware but is RDFS aware. Thus OIL can be thought of as an extension of RDFS.