EXPERIMENT:

Interoperability of Protégé 2.0 beta and OilEd 3.5 in the Domain Knowledge of Osteoporosis

Franz Calvo, MD fcalvo@u.washington.edu and John H. Gennari, PhD gennari@u.washington.edu
Department of Medical Education and Biomedical Informatics, School of Medicine. University of Washington, Seattle, WA, USA

I. Introduction

Because the idea of building a single, overall ontology for the entire Semantic Web seems impossible, we believe that integration of the various standards and ontology building tools is an important goal. However, the lack of interoperability between the different knowledge engineering tools currently available constitutes one of the bottlenecks of the Semantic Web [1]. Yet shared ontologies, ontology extension, and most ontology tools exhibit a certain degree of interoperability. In this experiment, we evaluate the capabilities of two ontology tools—Protégé and OilEd—to successfully import an ontology originally developed using the other tool. In Figure 1, we show the tools we have evaluated, indicating relationships among the tools. The arrows show the sorts of output formats (languages) that each tool can produce.

OilEd (version 3.5) http://oiled.man.ac.uk is the *de facto* standard environment for the language which grew out of the combination of DAML and OIL and has been variously known as DAML+OIL and OWL. The Web Ontology Language (OWL) has recently been advanced to a W3C Candidate Recommendation status. Details are available under the Semantic Web activity of W3C at http://www.w3.org/2001/sw . OWL is based on description logics but has many of the syntactic and other features of Frame languages. As DAML+OIL, the native format for OilEd ontologies, is not readable by Protégé 2.0 beta, ontologies created with OilEd should be exported in RDFS format to be readable by Protégé. OilEd can export in OWL format but is unable to import ontologies in this format, so OWL will not be evaluated.

Protégé 2.0 beta http://protege.stanford.edu is an extensible ontology editor and a knowledge base editor. Protégé uses the Open Knowledge-Base Connectivity protocol (OKBC) model as the basis for its own knowledge model. OKBC is a common query and construction interface for frame-based systems. As an effort to be compatible with other ontology tools, Protégé can export its ontologies in RDFS format. The current version provides beta level support for editing Semantic Web ontologies in OWL. The PAL constraints and Queries Tabs, a plug-in to represent axioms, is not compiled for Protégé 2.0 yet.

DAML+OIL plug-in for Protégé (alpha version) http://www.ai.sri.com/daml/DAML+OIL-plugin. is developed at SRI. The plug-in generates ontologies in two formats simultaneously, PPRJ and DAML, which are readable by Protégé and OilEd respectively. The OWL format is not supported.

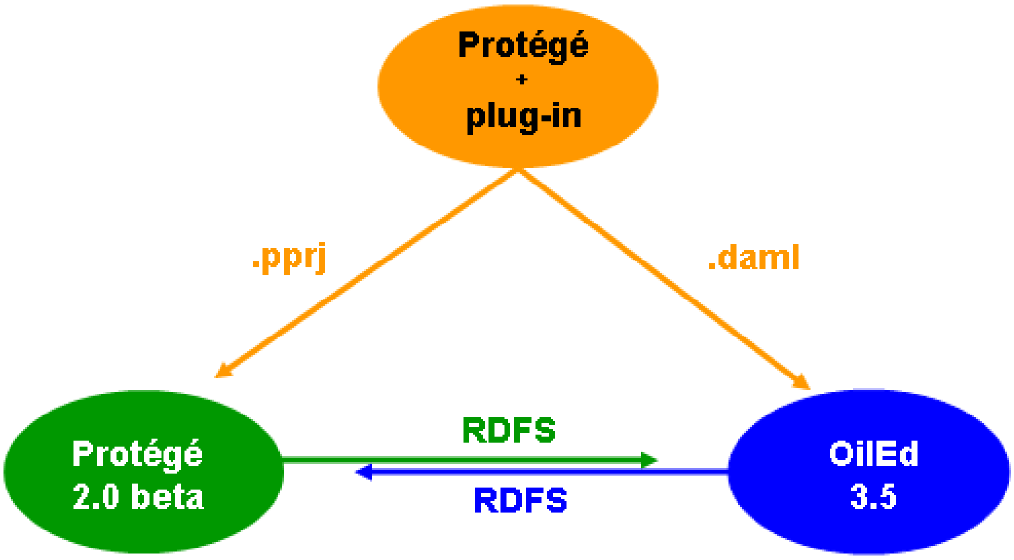

Figure 1. Relevant file formats for the two ontology tools being evaluated in this experiment.

II. Building the model

In order to test the interoperability of OilEd and Protégé, we have developed an ontology in the domain knowledge of osteoporosis, a common medical disorder. Our high level ontology has been modeled after the NLM's Unified Medical Language System (UMLS) Semantic Network http://www.nlm.nih.gov/research/umls , a freely available knowledge source which has been subject of numerous publications.

Our ontology contains over 200 concepts representing clinically-relevant aspects of osteoporosis, such as physical signs, symptoms, diagnostic tests and management options. Salient characteristics of knowledge to be represented in a biomedical ontology include:

- Preferred name. Several biomedical concepts are referenced by more than one name, and one of them is usually preferred over the others. For example "Postmenopausal osteoporosis" is also known as "Type I osteoporosis", but the former is the preferred one.
- Synonymy. There are biomedical concepts which have up to six synonymous (e.g. "Disease of hematopoietic system" has as synonymous "Blood dyscrasia", "Hematologic disease", "Disorder of hematopoietic system", "Hematopoietic disease", "Blood disorder", and "Hematopoietic disorder ").
- Disjoint concepts. Examples of mutually exclusive but not exhaustive concepts include "Medical device" and "Clinical drug", both subclasses of "Manufactured object".

- Partition. Examples of mutually exclusive and exhaustive concepts include "Organic chemical", "Inorganic chemical", and "Element, ion, or isotope", all of them subclasses of "Chemical viewed structurally".
- Defined and primitive classes. We have a defined class, when we are able to assign *sufficient* as well as *necessary* conditions for the class (e.g. "metabolic disease" and "disease of bone" for the defined class "metabolic bone disease"). In the case of most biomedical concepts, we can only assign them some *necessary* conditions. These classes are so-called primitive classes. (e.g. the class "metabolic bone disease" and some *necessary* but not *sufficient* properties build up the primitive class "osteoporosis").
- Multiple inheritance (polyhierarchy). Most biomedical concepts have more than one parent class.
- Abstract concepts. Some concepts, such as "Element, Ion, or Isotope" are used only for classification purposes. These abstract concepts can have subclasses, but not instances.
- Inverse relations. In some cases it is useful to represent relations that have inverse meanings because both are useful. (e.g. "causes" and "has_etiology"). The ontology should be able to automatically assign values to the other relation when one of them is used.
- Relation hierarchies. The UMLS Semantic Network associates all its 54 relations in a hierarchy. For example, the relation "spatially_related_to" and "temporally_related_to" are both subclasses of the relation "associated with".

In Table 1, we show how each of these three tools represents each of these ontological characteristics.

	Protégé 2.0 beta	Protégé + plug-in	OilEd 3.5
Preferred name	Represented as a metaclass	Represented as the class name	Represented as the class name
Synonymy	Represented as a meta-class with multiple cardinality	Not satisfactory because custom-built metaclasses are not allowed.	Not satisfactory because custom-built metaclasses are not allowed.
Disjoint concepts	Not possible	Possible, as a "LogicalDefinition".	Possible, as axioms
Partition	Not possible	Not possible	Initially possible but a bug converts a partition into disjunctions when the ontology is saved to disk (Figure 6)
Defined and primitive classes	Not possible	Possible	Possible (Figure 5)
Polyhierarchy	Possible	Initially possible but one of the parent classes disappears when imported by OilEd (Figure 3)	Possible
Abstract concepts	Possible	Not possible	Not implemented; metaclasses are not supported
Inverse relations	Implemented in the tool but not useful in this ontology because slots are used in override mode.	Implemented in the tool but not useful in this ontology because slots are used in override mode.	Yes (Figure 4).
Relation hierarchies	Yes, hierarchies are graphically displayed	Yes, hierarchies are graphically displayed	Yes, but the hierarchy is not graphically displayed.

Table 1. Comparison of the ontology-building capabilities of Protégé, Protégé+plug-in, and OilEd.

Ontologies generated using the alpha DAML+OIL plug-in for Protégé can not represent properly multiple inheritance (see Figure 3). The ability to represent polyhierarchies is crucial for a biomedical ontology, so we choose not to further test this plug-in tool in our experiment.

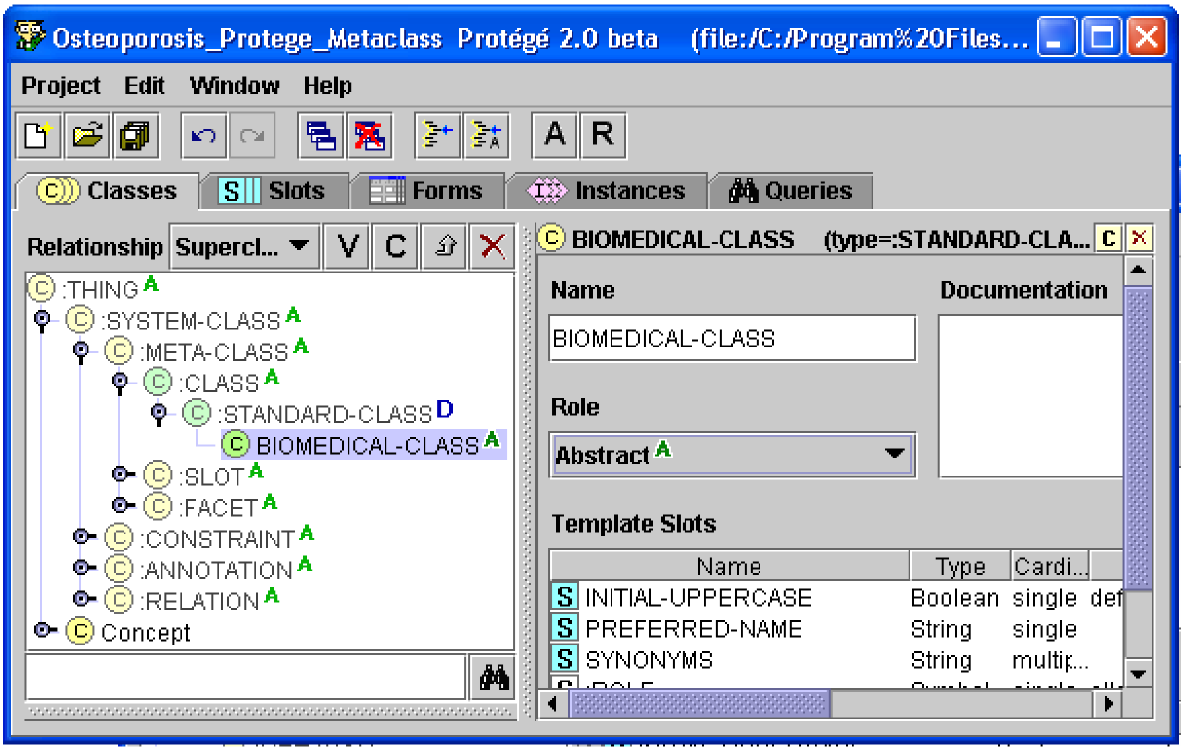

Figure 2. All concepts in our Protégé ontology have been modeled as subclasses of the *Biomedical-class* metaclass.

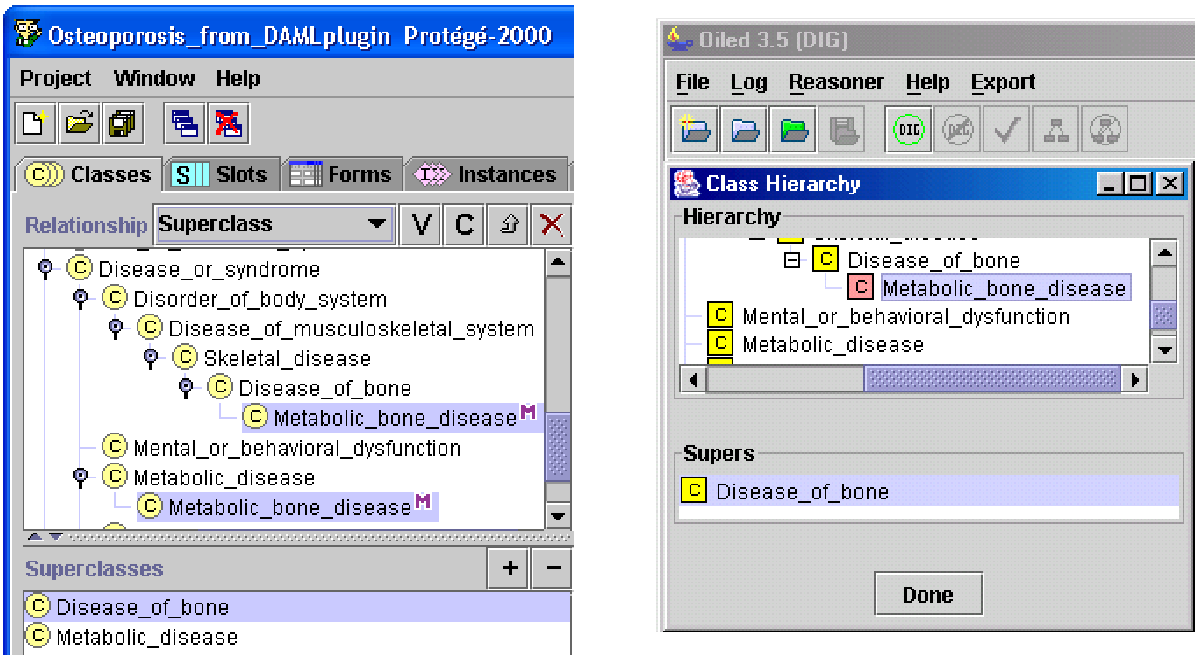

Figure 3. Polyhierarchies modeled with DAML+OIL plug-in for Protégé (left pane) disappear when imported by OilEd (right pane).

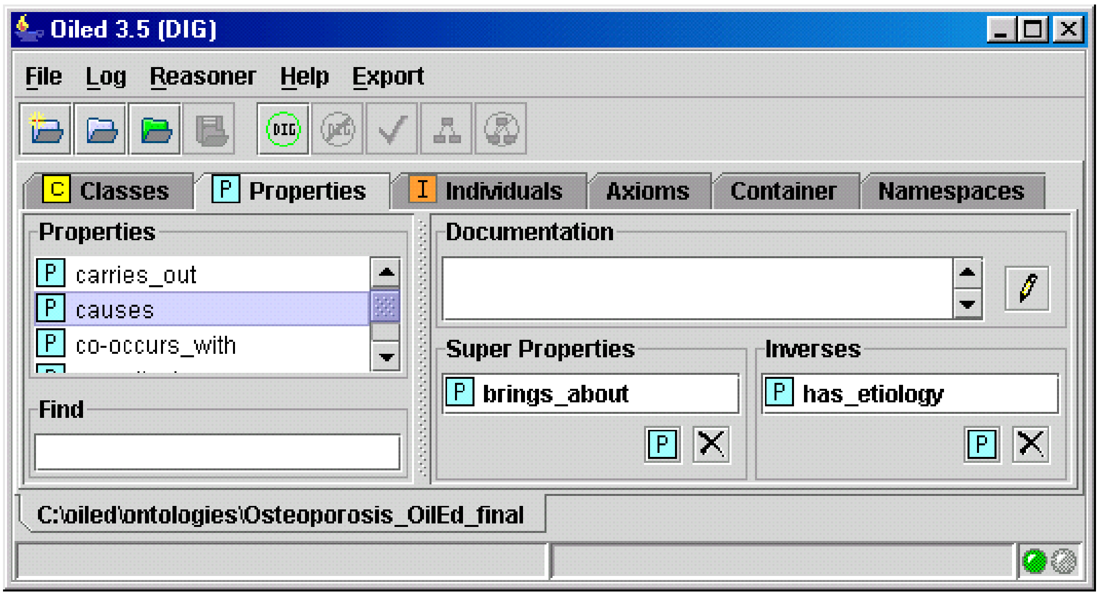

Figure 4. OilEd easily represents inverse properties (called slots in Protégé).

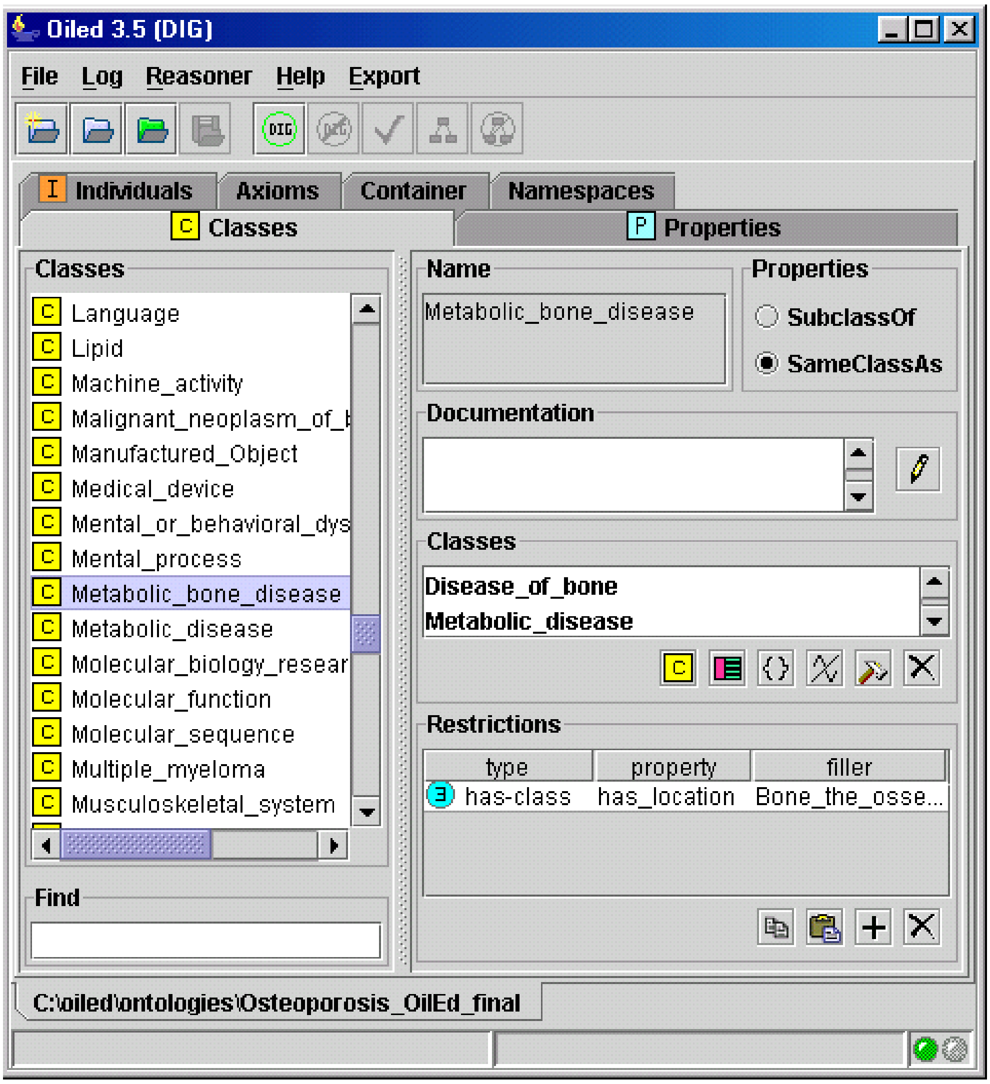

Figure 5. Example of a defined class (Metabolic_bone_disease) in OilEd.

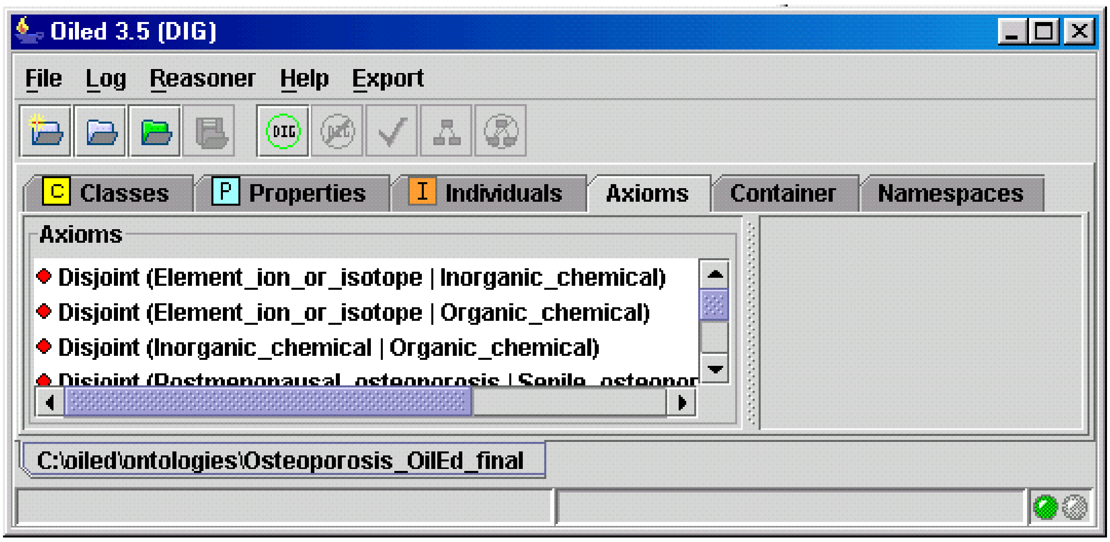

Figure 6. Representation of disjoint classes in OilEd. A bug prevents the definitive representation of partitions.

The ontologies created with Protégé, Protégé with the DAML+OIL plug-in, and OilEd are all available from: http://www.galenonet.com/Osteoporosis_Interoperability_experiment_Oct16.zip

III. From Protégé to OilEd

The ontology was exported from Protégé 2.0 beta in RDFS format. When opened with OilEd 3.5, all properties (slots in Protégé) were present but none of the classes. In additions, properties had lost its hierarchy. As classes were not present, we did not perform any further interoperability tests.

IV. From OilEd to Protégé and back

When the ontology—created with OilEd—is later imported by Protégé, all the class names are displayed with a prefix and we could not find a way to get rid of them (Figure 1.). However, the main limitation we found in this step of the interoperability evaluation of the tools is the disappearance of the restrictions modeled with OilEd. For example, Protégé's Template Slots window does not contain any representation of OilEd's Restriction "Metabolic_bone_disease" "has-class" "has_location" "Bone" (compare Figure 5 and Figure 7). Table 2 summarizes these changes.

We then saved to disk the imported ontology (originally created with OilEd), using the RDFS format. When this ontology (saved by Protégé, originally created with OilEd) was opened by OilEd, the classes pane was empty. The situation was similar to the one described above in section III. OilEd is not capable of successfully importing classes from ontologies saved with Protégé.

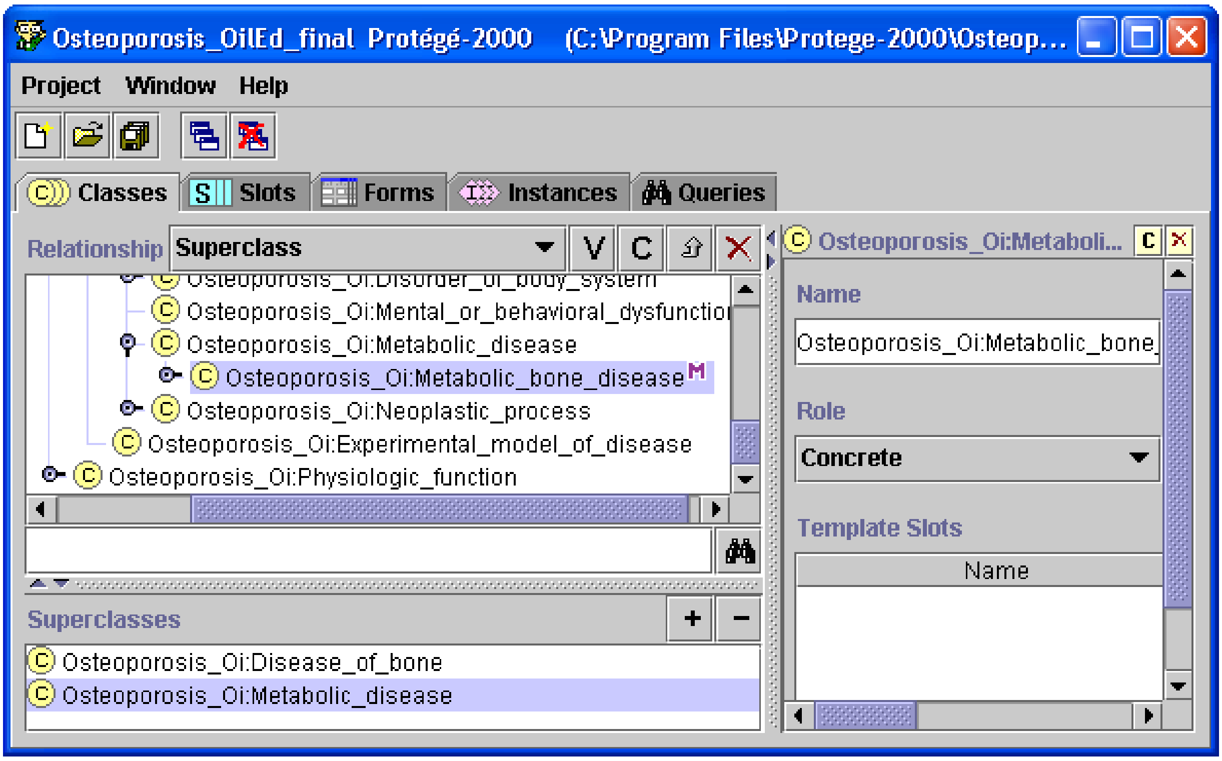

Figure 7. The ontology created with OilEd loses its restrictions (Template slots) when imported by Protégé.

	Ontology as imported by Protégé
Disjoint concepts	Not applicable
Defined and primitive classes	Not applicable
Polyhierarchy	Yes
Inverse relations	Disappear
Relation hierarchies	Conserved

Table 2. Changes found when the RDFS ontology is imported by Protégé.

V. Discussion

We designed an experiment to specifically evaluate the interoperability of Protégé 2.0 beta and OilEd 3.5, two promising tools to create ontologies. Our results demonstrate that interoperability is not possible between these tools, by way of the RDFS format.

The work here does not investigate the *causes* of these interoperation problems. In some cases, the problems we report may simply be due to immature tool development. The semantic web languages in particular are quite new, and it may take some time before robust and well-tested tools are available for these languages. However, in other cases, interoperation problems may be more fundamental, indicating a gap or discrepancy in the underlying knowledge models. For example, the inability of Protégé to understand and use disjoint concepts and defined concepts (see Table 2) may fall into this category.

Each one of the two ontology engineering tools analyzed in this experiment offer special capabilities to represent biomedical knowledge that the other tool cannot offer. Protégé's advantage over OilEd include the representation of preferred names, synonymous, and abstract concepts. On the other way, OilEd uniquely allows the representation of disjoint concepts, defined and primitive classes, and to more easily represent inverse relations.

Interoperability of ontology engineering tools is highly desired, in order to integrate the different knowledge representations developed by different groups and organizations. However, knowledge representation is an area so complex that, in general, the different tools available lack interoperability. The ongoing CO-ODE project, which will merge the best of Protégé and OilEd, promises to enable interoperability between knowledge representation tools by using OWL.

Acknowledgement

To Richard Phillips, MD, for his useful comments and review of the manuscript.

This work was supported by NIH grant D43 TW001286-04S1 Fogarty International Training in Medical Informatics.

Bibliography

Fensel D. Ontologies: Their Glory and the New Bottlenecks They Create. Workshop Semantic Web Action Line 2001. http://www.ontoweb.org/workshop/amsterdamdec8/ShortPresentation/swt_dieter.pdf

Noy N. F., Fergerson R. W., Musen M. A. The knowledge model of Protege-2000: Combining interoperability and flexibility. 2nd International Conference on Knowledge Engineering and Knowledge Management (EKAW'2000), Juan-les-Pins, France. 2000. http://protege.stanford.edu/publications/Knowledge_Model/protege-knowledge-model.html

Rector A. and colleagues. OilEd Normalised Ontology Tutorial – Biomedical version (for OilEd version 3.4). October 2002.
http://www.cs.man.ac.uk/mig/ontology-tutorial/oiled-biomedical-ontology-tutorial.zip

Rector A. CO-ODE/HyOntUse. Sixth International Protégé Workshop. July 2003. Manchester, England. http://protege.stanford.edu/workshop_vi/Alan_Rector_Co-Ode-Southampton-at-Manchester-2003-07-061.pdf