

Metarel: an Ontology to support the inferencing of Semantic Web relations within Biomedical Ontologies

Ward Blondé^{1,2}, Erick Antezana¹, Bernard De Baets², Vladimir Mironov³, Martin Kuiper³

¹Department of Plant Systems Biology, VIB, Ghent, Belgium; ²Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Belgium; ³Department of Biology, Norwegian University of Science and Technology, Trondheim, Norway

Abstract

While OWL, the Web Ontology Language, is often regarded as the preferred language for Knowledge Representation in the world of the Semantic Web, the potential of direct representation in RDF, the Resource Description Framework, is underestimated. Here we show how ontologies adequately represented in RDF could be semantically enriched with SPARUL. To deal with the semantics of relations we created Metarel, a meta-ontology for relations. The utility of the approach is demonstrated by an application on Gene Ontology Annotation (GOA) RDF graphs in the RDF Knowledge Base BioGateway. We show that Metarel can facilitate inferencing in BioGateway, which allows for queries that are otherwise not possible. Metarel is available on <http://www.metarel.org>.

Introduction

Ontologies have become one of the cornerstones of Knowledge Management (KM) in the Life Sciences.¹ They are increasingly used for annotating and integrating biomedical data, including genomic data, patient data, disease data, molecular data and more. For ontologies to fulfill their intended role, it is mandatory that both the ontologies and the data are modeled with the use of technologies that enable efficient integration and querying. In addition, these technologies should allow inferencing of new knowledge, one of the great promises of Knowledge Representation (KR).

The Semantic Web provides such technologies, the most important ones being the Resource Description Framework RDF and the Web Ontology Language OWL.^{2,3} The life sciences, in particular the domain of systems biology, are expected to be among the early adopters of these technologies.⁴ While a number of successful applications of the Semantic Web technologies in the life sciences have been reported (GenoQuery, LinkHub, Thea-online, BioDash), the field is still in its infancy and a number of technical hurdles need to be overcome.⁵⁻⁹ For example, while

OWL allows semantically rich knowledge representation, querying large knowledge bases represented in OWL poses a tremendous computational tractability challenge.¹⁰ Here we explore how a highly optimized RDF implementation can be used to alleviate some of the hurdles while still supporting rich inferencing.

Rendering class level relations in RDF

The two languages used widely for bio-ontologies, OBOF and OWL, differ markedly in the way they express the semantics of relations between classes.^{11,12} OWL expresses such relations by defining properties (relations between instances) in a property hierarchy. Relations between classes are created by adding extra fillers on the properties, which allows for number restrictions and grouping properties as necessary and sufficient conditions for defining classes. These fillers make the links between classes indirect in OWL/RDF, the RDF representation of OWL. OBOF, on the other hand, assumes all classes as defined by definition tags and relations are never considered as sufficient conditions. This approach has allowed to make direct links for relations between classes (see Figure 1).

The modeling with direct links in RDF, illustrated in Figure 1A, has a number of advantages over OWL/RDF: it is less verbose, it requires less computational power for loading and querying, and it is more intuitive. Moreover, the number of instances documented in biological ontologies is very small compared to the number of classes, which makes the treatment of relations between classes especially important in this domain. The direct links for relations between classes in OBOF can be readily modelled in RDF by putting them in the central place of an RDF triple (the predicate).

Interestingly, these relations (like *is_a*, *part_of*, etc.) have their own URIs (Unique Resource Identifiers) and they can appear in the first or the last place of a triple (resp. subject or object) as well. They can connect with any other URIs by using metarelations

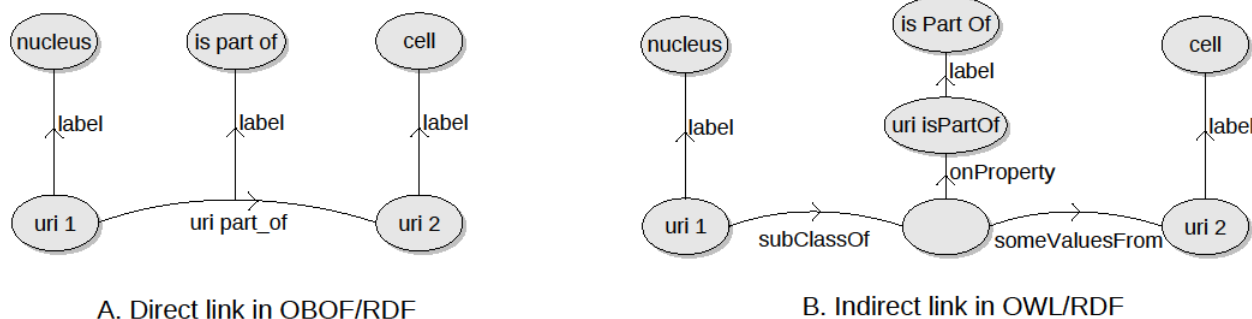


Figure 1. OBO/RDF (1A) relies on direct links between classes. OWL/RDF (1B) uses anonymous classes, as blank nodes, to link classes

as predicates. To exploit the opportunities this opens, we created Metarel, a meta-ontology for relations that can support RO, the Relationship Ontology of the OBO Foundry.^{13,14} With Metarel, OBO ontologies can be translated to RDF in a format with direct links, without any loss of expressivity.

Metarel was manually created with OBO-Edit and it has an export to RDF, created with ONTO-PERL.¹⁵ We list some of its most important features: 1. It allows to create meta-relations between relations; 2. It distinguishes ‘all-some’ relations from other types of relations; 3. It can indicate unambiguously which pairs of relations are each others inverses; 4. It has a place in its hierarchy where formally defined instance level relations can be attached; 5. It classifies reflexive and transitive relations in meta-classes; 6. It contains constructs for composites of relations; 7. It contains a meta-class for relations that are relevant towards inferencing (e.g. exclude *anatomically_related_to*, but not *dorsal_to*).

Metarel is used in the RDF knowledge base BioGateway, which has OBO relations.^{16,17} By linking all the relations in BioGateway to Metarel we obtain BioMetarel, essentially a bioscience specification of Metarel.

Inferencing with OBO Foundry relations

To investigate the efficacy of BioMetarel to facilitate inferencing in BioGateway we started from the semantics of its OBO relations. First of all, we emphasize that only relations at the class level are defined in RO. Their definitions refer to relations at the instance level, but those have neither unique identifiers nor definitions on their own. Therefore we will only infer new relations at the class level. As discussed in [14], the relations in OBOF have an ‘all-

some’ semantics. This means that if e.g. *A part_of B*, then for all the instances *a* of class *A* there is some instance *b* of class *B* for which *a* is part of *b*. The validity of any inferences from this semantics depends on the extent to which annotators have applied this rule correctly in producing knowledge statements for OBO. We found five sound mechanisms to infer new relations with OBO semantics:

1. A reflexive closure creates a relation link *A R A* for every class *A* and for every reflexive relation *R*. A query that asks for all the parts of the Golgi apparatus, will also return ‘Golgi apparatus’, because *part_of* is reflexive.
2. A transitive closure creates a relation link *A R C*, from any class *A* to any class *C*, for every transitive relation *R*, if the relation links *A R B* and *B R C* exist already. E.g. ‘nucleolus *part_of* nuclear lumen’ and ‘nuclear lumen *part_of* nucleus’, creates ‘nucleolus *part_of* nucleus’.
3. The inferencing of relations that have priority over the subsumption relation ‘*is_a*’ creates a relation link *A R C* if the links *A R B* and *B is_a C* exist, as well as for *A is_a B* and *B R C*, whenever *R* has an all-some semantics. E.g. ‘BRAF1_HUMAN *has_function* diacylglycerol binding’ and ‘diacylglycerol binding *is_a* lipid binding’, creates ‘BRAF1_HUMAN *has_function* lipid binding’.
4. The inferencing from the relation hierarchy creates a relation link *A R B* if the link *A S B* exists, and if *S* is a subrelation of *R*. E.g. ‘AKIP_HUMAN *negatively_regulates* mitosis’ creates ‘AKIP_HUMAN *regulates* mitosis’.
5. A compositional closure creates a relation link *A R C* if the links *A S B* and *B T C* exist and if

```

BASE    <http://www.semantic-systems-biology.org/>
PREFIX  ssb:<http://www.semantic-systems-biology.org/SSB#>
INSERT INTO GRAPH <25.H_sapiens_tc> {
    ?class1 ?resulting_relation ?class3.
}
WHERE {
    GRAPH <25.H_sapiens_tc> {
        ?class1 ?first_relation ?class2.
    }
    GRAPH <gene_ontology_edit_tc> {
        ?class2 ?second_relation ?class3.
    }
    GRAPH <biometarel> {
        ?composite ssb:first_relation ?first_relation.
        ?composite ssb:second_relation ?second_relation.
        ?composite ssb:resulting_relation ?resulting_relation.
    }
}

```

Figure 2. A SPARUL update query that computes the compositional closure of GOA annotations for Homo Sapiens. In this update, a composition like ‘located_in plus part_of results in located_in’ is extracted from BioMetarel and operated over the Gene Ontology and the GOA data

the relations S , T and R , not necessarily all different, form a composite (role chain). E.g. ‘NARF_HUMAN located_in nuclear lumen’ and ‘nuclear lumen part_of nucleus’ creates ‘NARF_HUMAN located_in nucleus’.

Somewhat counterintuitively, relations with an all-some semantics can not have inverses. Consequently, they can not be symmetric either, as this would imply they are their own inverses. Consider for example the statement ‘feather part_of animal’. As every feather is part of a bird, and every bird is an animal, this can be considered a sound statement. The inverse statement ‘animal has_part feather’, meaning that every animal has some feather as part, is clearly nonsense. To indicate e.g. that every feather is part of some bird, and also every bird has some feather as part, annotators should use two statements.

Creating the closures

As an application, we created a relational closure over the relations in the Gene Ontology Annotations (GOA) and the Gene Ontology (GO) in BioGateway.^{18,19} By this we mean the explicit creation of all the relations that are relevant in queries from users, and that can be inferred from the documented relations in BioGateway and from the semantics of relations. We used the RDF update language SPARUL for computing and adding these relations as RDF triples.²⁰ The method consists of four steps:

1. *Creating Biorel.obo.* This file expands RO.obo with all the relations that are used in the OBO Foundry, and with some extra tags for transitivity that were missing in RO.obo.
2. *Creating BioMetarel.* We merged the exports Biorel.rdf and Metarel.rdf to the BioMetarel RDF graph, with SPARUL. The enhancement of the semantic content for inferencing occurred in this step.
3. *Creating the closure of the Gene Ontology.* This was done by recursively running the SPARUL queries, (effectively applying the inferencing mechanisms in the previous section) on the BioMetarel and the GO graphs, until no further inferences could be made. The closure graph of GO contains 1.2 million triples, whereas the original GO graph contained only 0.57 million triples.
4. *Creating the closure of the Gene Ontology Annotations.* We created the compositional closure and the priority over *is_a*, for all the GOA graphs in BioGateway. The closure graph for *Homo sapiens*, for example, has 4.0 million triples, compared to 3.3 million triples for the normal graph.

The preconstructed closures allow many useful queries with SPARQL, the RDF query language, that are otherwise not possible.²¹ Examples include finding the proteins that are located in the same

protein complex and finding all the proteins with a given function or involved in a given process. All the original sources are kept separately, to allow querying

Conclusion

We have shown how the relation ontology Metarel can be used to perform extensive inferencing in biomedical ontologies represented in RDF, a Semantic Web standard. For this, we integrated the OBO Foundry relations in the hierarchy of Metarel, and the ensuing biological relationship ontology BioMetarel was used to recursively inference in the RDF store BioGateway. Triples constructed by inferencing were propagated by operating SPARUL update queries over BioMetarel and the relevant biomedical ontologies. Such inferences allow more powerful queries, and essentially increase the value of RDF for Knowledge Management significantly.

Acknowledgements

We thank Peter Dawyndt for discussions about Metarel, Nirmala Seethappan for maintenance of BioGateway, and Kenny Billiau, Bjørn Lindi, Dany Cuyt and Luc Van Wiemeersch for IT support.

References

1. Shadboldt N, Hall W, Berners-Lee T. The semantic web revisited, IEEE Intelligent Systems, 2006, 21:96–101.
2. RDF resource description framework, 2004. <http://www.w3.org/RDF/>.
3. OWL web ontology language guide, 2004. <http://www.w3.org/TR/owl-guide/>.
4. Ruttenberg A, Clark T, Bug W et al. Advancing translational research with the semantic web. BMC Bioinformatics, 2007, 8 Suppl 3, p. S2.
5. Antezana E, Kuiper M, Mironov V. Biological knowledge management: the emerging role of the Semantic Web technologies. Brief Bioinform, 2009, p.24
6. Lemoine F, Labedan B, Froidevaux C. GenoQuery: a new querying module for functional annotation in a genomic warehouse. Bioinformatics, 2008, 24:322–329.
7. Smith A, Cheung K, Yip K et al. LinkHub: a semantic web system that facilitates cross-database queries and information retrieval in proteomics. BMC Bioinformatics, 2007, 8 Suppl 3, p. S5.
8. Pasquier C. Biological data integration using semantic web technologies. Biochimie, 2008, 90:584–594.
9. Neumann E, Quan D. Biodash: a semantic web dashboard for drug development. In Pacific Symposium on Biocomputing. World Scientific, 2006, pp 176–187.
10. Weithöner T, Liebig T, Luther M, Böhm S. What's wrong with OWL benchmarks? In Proc. of the 2nd Int. Workshop on Scalable Semantic Web Knowledge Base Systems, 2006, pp. 101–114.
11. Smith B, Ashburner M, Rosse C et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotech, 2007, 25(11):1251–1255.
12. Golbreich C, Horrocks I. The OBO to OWL mapping, GO to OWL 1.1! In OWLED, 2007, p35.
13. Metarel. <http://www.metarel.org>.
14. Smith B, Ceusters W, Klagges B et al. Relations in biomedical ontologies. Genome Biology, 2005, 6:R46.
15. Antezana E, Egaña M, De Baets B, Kuiper M, Mironov V. ONTO-PERL: an api supporting the development and analysis of bio-ontologies. Bioinformatics, 2008, pp. 885–887.
16. BioGateway. <http://www.semantic-systems-biology.org/biogateway>.
17. Antezana E, Blondé W, Egaña M, Rutherford A, Stevens R, De Baets B, Mironov V, Kuiper M. Structuring the life science resourceome for semantic systems biology: lessons from the BioGateway project. SWAT4LS, 2008, CEUR-WS 435.
18. Camon E, Magrane M, Barrell D et al. The Gene Ontology Annotation (GOA) database: sharing knowledge in Uniprot with Gene Ontology. Nucleic Acids Research, 2004, 32:262–266.
19. The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. Nature Genet., 2000, 25:25–29.
20. SPARUL. <http://jena.hpl.hp.com/~afs/SPARQL-Update.html>.
21. SPARQL. <http://www.w3.org/TR/rdf-sparql-query/>