

The SemanticScience Integrated Ontology (SIO) for Biomedical Research and Knowledge Discovery

Michel Dumontier*

Department of Biology, Institute of Biochemistry, School of Computer Science, Carleton University.
michel.dumontier@gmail.com

Christopher J. O. Baker

Department of Computer Science and Applied Statistics, University of New Brunswick, Saint John, Canada
bakerc@unb.ca

Joachim Baran

Ontario Institute for Cancer Research, 101 College Street, Suite 800, Toronto, Ontario, M5G 0A3, Canada
joachim.baran@gmail.com

Alison Callahan

Department of Biology, Carleton University
alison.callahan@carleton.ca

Leonid Chepelev

Department of Biology, Carleton University
leonid.chepelev@gmail.com

José Cruz-Toledo

Department of Biology, Carleton University
jose.cruztoledo@carleton.ca

Geraint Duck

School of Computer Science, University of Manchester, Oxford Road, Manchester, M13 9PL, UK
duckg@cs.man.ac.uk

Dana Klassen

Data Intensive Infrastructure Unit, Digital Enterprise Research Institute, National University of Ireland
klassen.dana@gmail.com

James P. McCusker

Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY
mccusj@rpi.edu

Matthias Samwald

Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria
matthias.samwald@meduniwien.ac.at

Natalia Villanueva-Rosales

Cyber-ShARE Center of Excellence, University of Texas at El Paso

nvillanuevarosales@utep.edu

Mark Wilkinson
Centro de Biotecnología y Genómica de Plantas
UPM-INIA (CBGP)
mark.wilkinson@upm.es

Robert Hoehndorf
Department of Physiology, Development and Neuroscience
University of Cambridge, United Kingdom
rh497@cam.ac.uk

* To whom the correspondence should be addressed.

KEYWORDS: ontology, design patterns, OWL, data integration

Abstract

The Semanticscience Integrated Ontology (SIO) is an integrative ontology to facilitate semantics-powered data integration and biomedical knowledge discovery. SIO features a simple upper level comprised of essential types and relations for the rich description of arbitrary (real, hypothesized, virtual, fictional) objects, processes and their attributes. SIO specifies basic design patterns to describe and associate qualities, capabilities, functions, quantities, and informational entities including textual, geometrical, and mathematical entities, and provides specific extensions in the domains of chemistry, biology, biochemistry, and bioinformatics. SIO provides the ontological foundation for dozens of datasets that are part of the Bio2RDF linked data for the life sciences project and is used for semantic integration and discovery of over 700 SADI-based bioinformatic semantic web services. SIO is freely available to all users under a creative commons by attribution license. See website for further information: <http://sio.semanticscience.org>.

Introduction

Biomedical research is poised to enter an era of unprecedented large scale data analysis powered by hundreds of public biological databases and hundreds of millions of patient records. However, there is a real and urgent need to explore effective methods for biomedical data integration and knowledge management [1, 2]. Semantic-based technologies, such as ontologies, offer a proven method to exploit expert-based knowledge in the analysis of large datasets through terminological reasoning such as correspondence, classification, query answering and consistency checking.

The Semantic Web effort, as developed through the World Wide Web Consortium (W3C), provides a set of standards to facilitate the representation, publication, linking, querying and discovery of heterogeneous knowledge using web infrastructure [3]. In particular, the Resource Description Framework (RDF) offers simple data assertions and hierarchical organization, while the Web Ontology Language (OWL) allows for more sophisticated logic-based expressions to capture the meaning of terms

and the relations between the entities they describe. Biomedical researchers have made use of Semantic Web technologies to uncover curation errors in systems biology models[4], find putative disease-causing genes [5], identify aberrant pathways [6], and uncover alternative drug therapies based on mechanism of action [7], among others [8].

Expressive, logic-based languages such as OWL offer enormous flexibility towards the accurate capture of biomedical knowledge and markedly distinct models may emerge from different users and communities. With the goal of facilitating knowledge discovery through simple, but effective data integration, we developed the Semanticscience Integrated Ontology (SIO) with a set of basic types and relations by which arbitrary entities can be described through reusable design patterns. SIO provides classes and relations to describe and relate objects, processes and their attributes with specific extensions in the domains of chemistry, biology, biochemistry, and bioinformatics. Its relations cover aspects of spatial and temporal qualitative reasoning including location, containment, overlap, parthood and topology; participation and agency, linguistic and symbolic representation, as well as comparative relations. Using straightforward mappings, we report on the substantial benefits afforded by SIO in the retrieval of RDF-based linked data and automatic composition of OWL-described semantic web services. Although the SIO development is driven by needs in the biomedical domain, there is a reasonable expectation that SIO and its design patterns can be generally applied to a broader set of domains.

Foundation

SIO follows a three-dimensional worldview which differentiates objects from processes: objects are entities that occupy space and are fully identifiable at any moment in time in which they exist, whereas any non-exactly overlapping temporal interval of a process only identifies some part of the process. While all SIO entities *exist at* and are *located in* some space and time, these are not necessarily real space or real time, but may instead occur in a hypothetical (propositional), virtual (electronic), or fictional (creative work) setting. These and other qualities (intrinsic attributes) or realizable entities such as capabilities (action specifications) and roles (behaviors, rights and obligations) exist concurrently with the entity that bears them. In order to indicate when a capability or role plays a part in some process, SIO specifies that the capability or role is *realized in* that process. Thus, it becomes possible to delineate the set of actions that an entity makes without having to specifically enumerate them. Another significant SIO design pattern is that information about an entity is treated as an *attribute of* the entity, and literal values such as strings, numbers (integer, float, double), booleans and dates, of these attributes are directly captured using SIO's sole datatype property - `sio:has-value`.

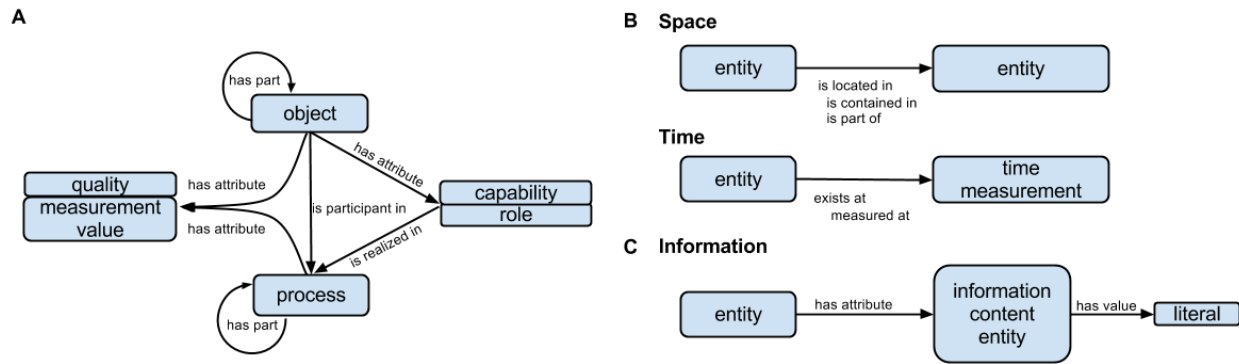


Figure 1 (A) Key SIO entities are objects, processes and their attributes (qualities, capabilities, roles, measurement values). Processes have objects as participants and may realize specific roles and capabilities. (B) Spatial and temporal qualification of SIO entities is captured through a set of relations (is located in, exists at), while (C) information in the form of literals (string, numbers, dates) are captured as instances of information content entities which are then associated with their specific objects or processes.

As of June 2013, SIO is an OWL-DL ontology (SRIQ(D) expressivity) that defines 1395 classes, 202 object properties, 1 datatype property and 8 annotation properties. English labels are provided using the `rdfs:label` annotation property while human readable, English language definitions are provided using the Dublin Core (dc:) Metadata property `dc:description`. Classes, which describe (object, process, attribute) types, and object properties, which denote a relation between entities, are organized in their own subsumption hierarchies with 1746 `subClassOf` axioms and 207 `subPropertyOf` axioms (Figure 2). In SIO, 'entity' forms the parent class while 'is related to' forms the parent object property.



Figure 2 Selected portions of (A) class and (B) property hierarchies in SIO.

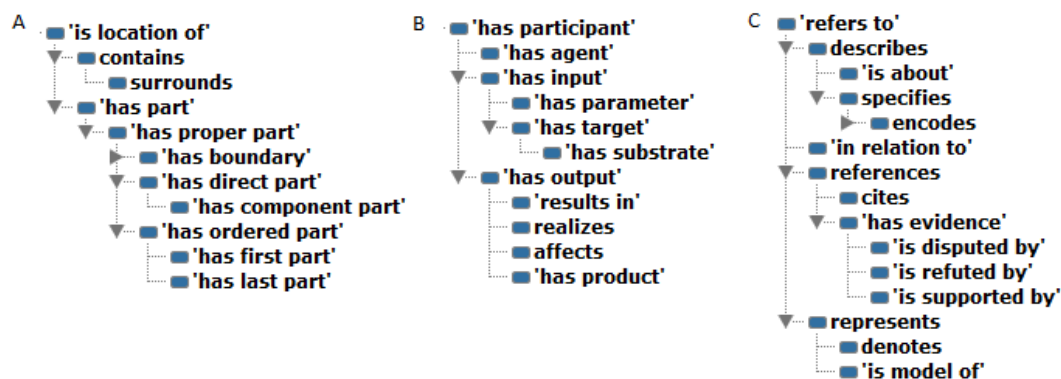


Figure 3 Relation hierarchies for (A) mereological relations (B) participatory relations and (C) referential relations

Mereotopological relations: Describing an entity in terms of its parts and how these parts are connected together is essential for building rich entity descriptions. SIO offers a number of relations to formulate mereotopological descriptions including: 'is located in', 'contains', 'surrounds', 'has part', 'has proper part', 'has direct part', 'has component part', 'is connected to', 'is directly connected to', 'is directly before' and 'is directly after'. Each of these has a specific meaning and is associated with a formal semantic when possible in OWL. At the highest level, 'is located in' is a transitive relation in which the 2D/3D spatial region occupied by one entity is a part of the 2D/3D spatial region occupied by another entity. 'contains' is a transitive relation in which the 3D spatial region occupied by A has the 3D spatial region occupied by B as a part, but it is not the case that A has B as a part. A 'surrounds' B when the A 'contains' B and A 'is adjacent to' B or A 'is directly connected to' B. 'has part' is a reflexive and transitive relation that links an entity to itself (reflexive) and all of its parts (transitive). This means that a query for the parts of the whole will include itself as well as all of its parts. 'has proper part' is an irreflexive and asymmetric relation that ensures that the whole is different from and not one of its proper parts. 'has direct part' provides an avenue to quantify the number of parts (via a cardinality restriction) at a target type granularity, which is not otherwise possible in OWL over a transitive relation. 'has component part' can be used to indicate that the part is intrinsic to the whole, and that the removal of the part changes the identity of the whole.

While the 'has part' relation hierarchy is focused on identifying the parts of a whole, the next set of relations allows one to specify how the parts are positioned to one another. 'is connected to' is a symmetric, transitive relation that specifies that components either directly share a boundary (they are directly connected to each other) or that they are indirectly connected by a path of unbroken direct connections. 'is directly connected to' is a symmetric relation that indicates that two components share a boundary. Since this relation is non-transitive, we can use it in statements to quantify the number of connections from one part to other kinds of parts. 'is directly before' is a relation between entities placed on a dimensional axis in which the projection of the position of the first entity is numerically less than the projection of the position of the second entity, and the entities are adjacent to one another. This is useful for indicating the spatial positioning of residues in linear biopolymers such as proteins or nucleic acids. A domain specific relation, 'is covalently connected to' enables one to describe the atomic connectivity within a molecule such as methane:

```

'methane'
equivalentClass
'molecule'
and 'has component part' exactly 3 'methane hydrogen atom'
and 'has component part' exactly 1 'methane carbon atom'
and 'has component part' only
('part of' some 'methane carbon atom' or 'part of' some 'methane hydrogen atom')

'methane hydrogen atom'
equivalentClass
'hydrogen atom'
and 'is component part of' exactly 1 'methane'
and 'is covalently connected to' exactly 1 'methane carbon atom'

'methane carbon atom'
equivalentClass
'carbon atom'
and 'is component part of' exactly 1 'methane'
and 'is covalently connected to' exactly 3 'methane hydrogen atom'

```

Participatory relations: Processes can be described in terms of their participants in the following way. 'has participant' indicates which entities participate in a process. 'has agent' specifies entities that direct or actively participate in the process. 'has input' specifies entities at the start of the process. 'has parameter' specifies those variables (and their values) used in the process. 'has target' specifies entities that are modified during the process, but retain their identity. 'has substrate' specifies entities that are consumed (or are sufficiently changed that they lose their canonical identity). 'has product' specifies new entities formed as a result of a process. Relations such as 'has substrate', 'has target', 'has product' are examples of role-specialized relations. In SIO, more explicit role-based assertions can be formulated by stating that the role of an entity is realized in the process, for instance in the phosphorylation of an enzyme by ATP:

```

'substrate-enzyme phosphorylation by ATP' equivalentTo:
'biochemical reaction'
and 'realizes' some
('substrate role' and 'is role of' some 'ATP-substrate enzyme complex')
and 'realizes' some
('product role' and 'is role of' some 'ADP-substrate-phosphorylated-enzyme complex')

```

Using these SIO relations infers that the entities having the realized role are also participants of the process via an OWL2 role chain (realizes o is role of -> has participant).

Referential relations: Referential relations are used to indicate what an object refers to or the nature of the mention of one entity by another. At the top level, 'refers to' enables this basic mention, while 'references' is a relation where one entity mentions another, 'describes' is a relation where one entity provides a detailed account of another, and 'represents' is a relation where one entity is a sign, symbol or model for another. 'describes' can be further partitioned into 'is about' where one entity provides information about another while 'specifies' contains specific information that can be used as evaluation criteria to determine the degree of conformance. 'references' is further subdivided into 'cites' as a relation to refer to by way of example, authority or proof, and 'has evidence' which is a relation between a proposition and something that demonstrates the truth of the assertion. 'has evidence' then has three sub-properties ('is supported by', 'is disputed by', 'is refuted by') which can articulate the degree of evidence that one entity offers another. Finally, 'represents' is subdivided into 'denotes'

which is a relation between an entity and what it is a sign or indication of, or what it specifically means and 'is model of' which indicates that an artifact is a model or representation of another.

Selected Design Patterns

A. Scientific investigation

A scientific [investigation](#) is a [procedure](#) that aims to support, dispute or refute a well formulated [hypothesis](#) through the [analysis](#) of [data](#) obtained through [observation](#) and/or [measurement](#). Investigations *usually* involve:

- the development of a research [plan](#) which includes, but is not limited to:
 - the formulation of a [hypothesis](#)
 - the formulation of aims and [objectives](#)
 - the formulation of a [study design](#)
- the execution of the research plan which includes, but is not limited to:
 - the [selection, preparation or collection](#) of [sample\(s\)](#)
 - the [collection of data](#) through [observation](#), [assay](#) or [measurement](#)
 - the [analysis](#) of [data](#)
 - the preparation of an investigational [report](#)

An investigation is related to its activity parts (e.g. sample preparation, measurement, analysis) through [has proper part](#), and activities, being themselves processes, may be temporally ordered (occurs before / after) with [precedes](#).

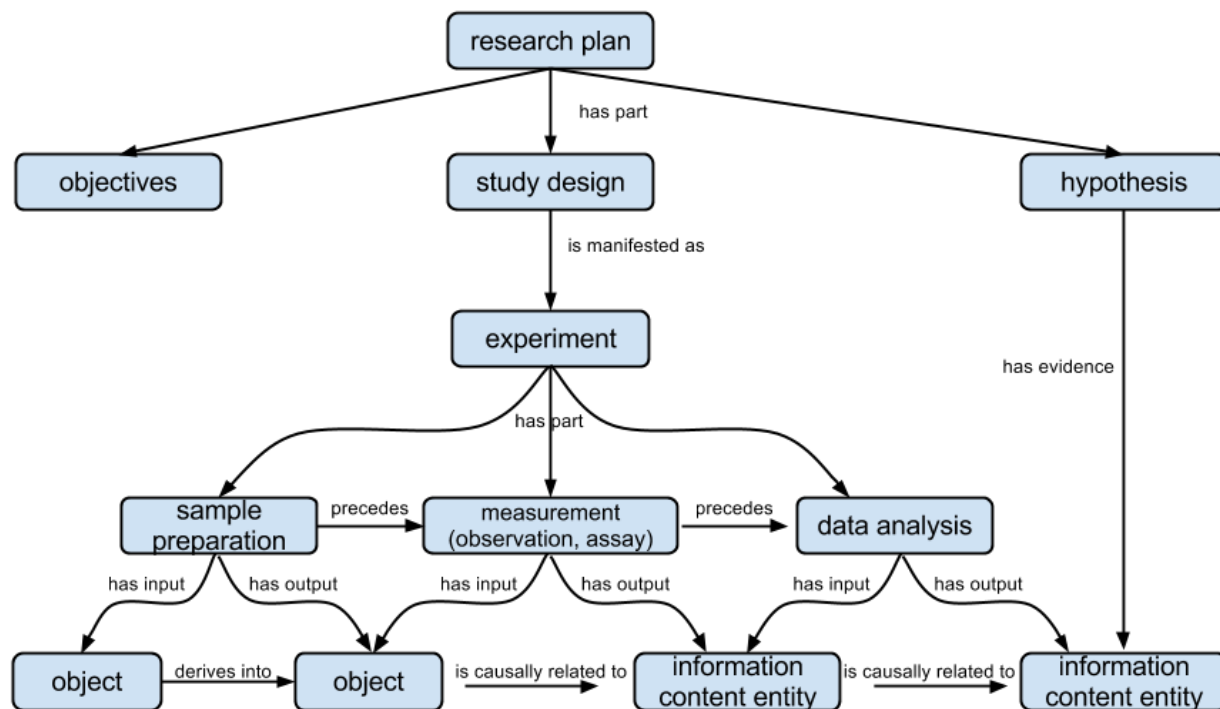


Figure 4 Diagram illustrating major entities and their relations in a scientific experiment.

Since [descriptions provide detailed information about](#) some [entity](#) ([object](#) or [process](#) or [attribute](#)), a [hypothesis](#) is a *proposed* explanation (a description) of some phenomena, while an [objective](#) is a description of a *desired* outcome. Descriptions that [specify](#) a set of actions to be executed are known as [action specifications](#), which include [plans](#), [study designs](#), recipes and [protocols](#). A plan should clearly identify ([specify](#)) one or more [objectives](#), and optionally specify a [hypothesis](#) or a [study design](#). Plans, like other action-based specifications, are [manifested](#) as processes.

[Procedures](#), such as [investigations](#), [assays](#) and [measurements](#) are [manifestations of effective specifications](#) (information content entities may be [manifested](#) as objects or processes, including [text](#), [sound](#), [physical gestures](#), etc). Thus, investigations (via the plan), hold hypotheses and objectives as [attributes](#). An objective [is realized in](#) the investigation if and only if those outcomes are fully apparent. Data generated from the investigation may also serve as [evidence for](#) the hypothesis, and more specifically found to be [supporting](#), [disputing](#), or [refuting](#).

B. Measurements and measurement values

Measurement values such as quantities or spatial positions are captured as information content entities, where the ICE is an attribute of the entity in question and the literal value is captured using SIO's only datatype property 'has value'. Units of measurements, preferably defined by the Unit Ontology (UO), are indicated with SIO's 'has unit' object property. The following RDF/N3 example (using the labels as URIs subset of SIO) shows how SIO captures Rob's mass of 74.5kg using a measurement scale on Jan 15, 2012 at 12:03pm EDT:

```
sio: <http://semanticscience.org/resource/>
taxon : <http://bio2rdf.org/taxon:>
obo: <http://purl.obolibrary.org/obo/>
```

```
:p rdfs:label "Rob" .           # Rob
:p a taxon:9606 .               # Rob is human, as per the NCBI taxonomy in Bio2RDF
:p sio:has-attribute :m1 .      # Rob has an attribute m1

:m1 rdfs:label "bob's mass on Jan 15, 2012"@en # Rob's mass
:m1 a sio:mass .               # m1 is a mass
:m1 sio:has-value "74.5"^^xsd:float .        # the value is 74.6
:m1 sio:has-unit obo:UO_0000009 .            # the unit is kilogram as per the Unit Ontology
:m1 sio:measured-at :t1 .                  # mass measured at t1

:t1 a sio:time-instant .
:t1 sio:has-value "2012-01-15T12:03:00-05:00"^^xsd:dateTime .

:mp a sio:measuring .          # measuring process
:mp sio:has-agent :bobscale .  # measuring involved bob's scale
:mp sio:has-output :m1 .       # measuring resulted in a measurement value

:bobscale rdfs:label "Rob's scale"@en # Rob's scale
:bobscale a obo:MMO_0000217 .        # body measurement scale as per the Measurement
                                      # Method Ontology
```

C. Biomedical knowledge

SIO offers a rich set of classes and relations to capture arbitrary biomedical knowledge either as more specific types and relations or facts that instantiate the ontology. SIO includes basic types and relations to characterize proteins, lipids, nucleic acids, small molecules, genotypes, phenotypes, biochemical reactions and pathways, among others. For instance, we can associate that a gene encodes a protein (via transcription and translation) and that proteins are associated with function (via experimental or computational generalization). Thus, to express that a gene encodes a protein of certain function, we state:

'gene x with protein y having function z'
subClassOf 'gene' that 'encodes' some ('protein' that 'has function' some 'capability')

In the case of genotype-phenotype associations, we express this as observing an attribute for some organism of a particular composition i.e.

'phenotype x of organism y with gene y having snp of A at position 12'
subClassOf ('phenotype' that 'is attribute of' some ('organism' that 'has part' some ('gene y' and 'has part' some ('snp' that 'has value' value 'A' and ('has attribute' some ('sequence position' that 'has value' value "12")))))

Figure 5 illustrates these and other relations pertinent to the field of molecular biology.

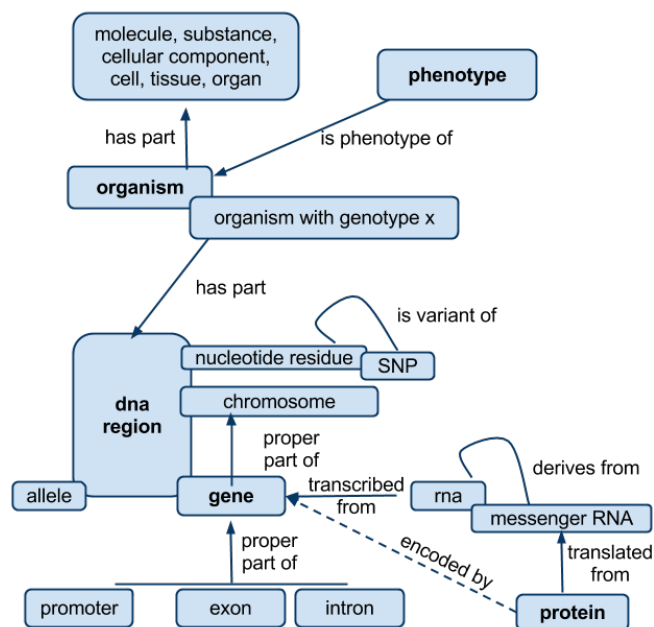


Figure 5 Conceptual map of SIO entities and their relations as it pertains to the field of molecular biology.

Applications

A. Semantic data integration and question answering

The Bio2RDF project uses Semantic Web technologies to offer the largest network of linked data for the life sciences [9]. Although the Bio2RDF approach provides minimal syntactic and referential interoperability (using RDF and a shared URI pattern), it does not address the issue of semantic interoperability across its datasets. Recent work [10] mapped SIO types and relations to Bio2RDF's dataset-specific types and relations, thereby enabling SIO-based queries over linked data. In that work, resources such as DrugBank [11], the Pharmacogenomics Knowledgebase [12] and the FDA's National Drug Code Directory (NDC) all provide drug information, and their types `http://bio2rdf.org/drugbank_vocabulary:Drug`, `http://bio2rdf.org/pharmgkb_vocabulary:Drug`, and `http://bio2rdf.org/ndc_vocabulary:human-prescription-drug` were mapped as subclasses of the SIO class 'drug'. Bio2RDF vocabulary mappings to SIO also make it possible to execute complex SPARQL queries over multiple Bio2RDF endpoints. For example, the following SPARQL query counts all the biochemical reactions in the Bio2RDF BioModels database [13] that are involved in the Gene Ontology (GO) term "protein catabolic process" or one of its subclasses:

```
SELECT ?go ?label count(distinct ?x)
WHERE {
  {
    # get all the biochemical reactions labelled with protein catabolic process
    ?go rdfs:label ?label .
    FILTER regex(?label, "^protein catabolic process")
    SERVICE <http://biomodels.bio2rdf.org/sparql> {
      ?x <http://semanticscience.org/resource/is-identical-to> ?go .
      ?x a <http://semanticscience.org/resource/biochemical-reaction>
    } # end service
  } UNION {
    # get all the biochemical reactions that are more specific than "protein catabolic process"
    ?go rdfs:label ?label .
    ?go rdfs:subClassOf+ ?tgo . # get all the subclasses of the target to term
    ?tgo rdfs:label ?tlabel .
    FILTER regex(?tlabel, "^protein catabolic process")
    SERVICE <http://biomodels.bio2rdf.org/sparql> {
      ?x < http://semanticscience.org/resource/is-identical-to> ?go .
      ?x a < http://semanticscience.org/resource/biochemical-reaction>
    } # end service
  } # end union
} # end where
```

This query is possible because the BioModels type for biochemical reaction has been mapped as a subclass of SIO '[biochemical reaction](#)'. Similarly, the BioModels predicate for 'is identical to' has been mapped as a subproperty of the SIO object property '[is identical to](#)'.

B. Semantic Web Service Interoperability

The Semantic Automated Discovery and Integration (SADI) framework [14] consists of a set of design patterns for producing stateless Web Services that natively consume and produce RDF data. The structure of the input and output data for SADI services are formally described by an input OWL class and an output OWL class respectively. SIO's rich set of types and properties make it amenable to describe Semantic Web services such as those created by Semantic Automated Discovery and Integration (SADI). These semantically described web services can then be automatically composed into a workflow to answer questions [15], classify and annotate molecules based on their structure [16], or to uncover health information regarding drug-drug interactions [17]. Over 700 SADI service have been

produced through the CANARIE C-BRASS project. Table 1 shows a selected set of classes and relations used in SADI services registered at <http://sadiframework.org>.

Table 1: Top 10 classes and relations used in SADI semantic web services registered with sadiframework.org

Class	Frequency	Property	Frequency
deoxyribonucleic acid sequence	159	has part	289
protein sequence	41	is about	71
ribonucleic acid sequence	21	has attribute	67
definition	2	is attribute of	17
name	2	is part of	9
sequence motif	2	has output	9
answer	1	is derived from	9
common name	1	derives into	9
description	1	is similar to	9
preferred name	1	overlaps with	9

Related efforts

The Basic Formal Ontology (BFO) is a small (36 class), domain-independent, upper level ontology that is used as a guiding ontology for many open biomedical ontologies (OBO). While SIO can be used to capture knowledge for any category or instance, the BFO primarily differs from SIO in that BFO is intentionally limited by its realist philosophy to only allow classes with at least one known instance and whose instances only exist in real space and time [18, 19]. Thus, SIO is broader and more applicable to situations of interest to the health care and life sciences that involve the description and characterization of imagined or virtual objects. SIO and BFO also differ in the presentation of their top level categorization (SIO: object/process/attribute) vs (BFO: continuant/occurrent/dependent continuant), but also have real differences in the description of processes and even whether they may have attributes (BFO does not allow process attributes or qualities leading to convoluted knowledge capture solutions) [20]. The BFO is often used in conjunction with the OBO Relation Ontology (RO) [21], a collection of 8 domain-independent (e.g. has part) relations developed to support the OBO Foundry effort[22]. However, unlike the BFO+RO pairing, SIO was specifically designed with both classes and the necessarily relations to facilitate the composition of a variety of meaningful logical definitions.

Nearly 160 relations have been added to RO as “RO-proposed” and include some biology-specific (e.g. develops from). To foster semantic interoperability between SIO and BFO/RO, we mapped 9 BFO classes

and 24 RO relations (<http://purl.obolibrary.org/obo/ro/bridge/sio-ro-bridge.owl>). BFO engineers are working on BFO 2.0 which promises to combine BFO classes and selected RO relations. However, the current proposal specifies new temporally and type qualified relations (e.g. 'part of continuant at all times that whole exists') that may require considerable efforts by all OBO ontology engineers to update their ontologies, therefore we encourage OBO editors to explore using SIO as a broader alternative for formulating and integrating scientific ontologies that interoperate with bioinformatic data and services.

BioTop [23] is an upper level ontology for biology and medicine that features 390 classes and 82 object properties. The class top-level is characterized by a flattened set of basic categories (material object, immaterial object, information object, process, quality, role, condition, disposition, time, value region) while the object hierarchy provides type-specific relations around physical, processual and abstract nature (e.g. has physical part, has processual part, has abstract part). BioTop includes relatively sophisticated formalization for selected terms e.g. pathological disposition is defined as "disposition that ('inheres in' some ('bearer of' some (canonicity and ('quality located' some 'noncanonical value region'))))", where SIO would simply express it as a 'biological disposition' that ('is attribute of' some ('entity' that 'has attribute' some 'pathological quality')). BioTop has been used to provide a number of ontology design patterns [24, 25] and to identify semantic type errors in the UMLS network [26].

The Ontology for Biomedical Investigations (OBI) is a domain ontology focused scientific experiments and instruments [27]. OBI features an extensive assay, material processing, and data transformation types. In general, the design patterns laid out in OBI and SIO are similar and compatible, however, given OBI's intent to join the OBO Foundry as an orthogonal ontology, it is unclear how OBI will consolidate its substantial overlap with other OBO Foundry candidates and how it will adapt its relations to those proposed in BFO 2.0.

The Translational Medicine Ontology (TMO) is a unifying ontology for chemical, genomic and proteomic data with disease, treatment, and electronic health records [7]. The TMO acted as a central schema that mapped basic types to dozens of bio-ontologies and linked open data. The utility of the TMO was demonstrated by answering a series of questions pertaining to diagnosis, prescription, drug mechanism of action, alternative therapeutics, and biomarkers. As SIO emerged from considerations in the TMO effort, SIO can be seen as the supported successor to TMO.

Conclusion

The SemanticScience Integrated Ontology (SIO) is an integrated ontology of basic types and relations to capture a wide span of knowledge through a set of emerging design patterns using RDF/OWL. SIO has emerged to support the demands of the bioinformatics community, with a special emphasis on biological knowledge representation as well as data and service interoperability. SIO is being developed to facilitate information retrieval and knowledge discovery.

Availability

The SIO homepage is <http://sio.semanticscience.org>. SIO is freely available under a creative commons by attribute license at <http://semanticscience.org/ontology/sio.owl>. The base namespace for SIO entities

(classes, properties) is <http://semanticscience.org/resource/>. SIO entities are identified using resolvable HTTP URIs, initially formulated as an alphanumeric identifier e.g. http://semanticscience.org/resource/SIO_000001, but is alternatively accessible using a label-based identifier e.g. <http://semanticscience.org/resource/is-related-to>. These and other generated subsets are available from <http://goo.gl/OLgN8>.

Acknowledgements

This work was funded, in part, by NSERC Discovery Grant to MD, Ontario Early Researcher Award to MD and CANARIE NEP-2 grant to MD, CB and MW. We would like to thank the following individuals for thoughtful discussions and contributions on and off the mailing list: Jerven Bolleman, Kevin Cohen, Melanie Courtot, Simon Jupp, Nichealla Keath, Jin-Dong Kim, James Malone, Luke McCarthy, Chris Mungall, David Osumi-Sutherland, Nuria Queralt, Alexandre Riazanov and Robert Stevens.

References

1. Gardner, S.P., *Ontologies and semantic data integration*. Drug Discovery Today, 2005. **10**(14): p. 1001-1007.
2. Goble, C. and R. Stevens, *State of the nation in data integration for bioinformatics*. J Biomed Inform, 2008. **41**(5): p. 687-93.
3. Shadbolt, N., W. Hall, and T. Berners-Lee, *The Semantic Web Revisited*. Intelligent Systems, IEEE, 2006. **21**(3): p. 96-101.
4. Hoehndorf, R., et al., *Integrating systems biology models and biomedical ontologies*. BMC Systems Biology, 2011. **in press**.
5. Hoehndorf, R., P.N. Schofield, and G.V. Gkoutos, *PhenomeNET: a whole-phenome approach to disease gene discovery*. Nucleic Acids Res, 2011.
6. Hoehndorf, R., M. Dumontier, and G.V. Gkoutos, *Identifying aberrant pathways through integrated analysis of knowledge in pharmacogenomics*. Bioinformatics, 2012. **28**(16): p. 2169-75.
7. Luciano, J.S., et al., *The Translational Medicine Ontology and Knowledge Base: driving personalized medicine by bridging the gap between bench and bedside*. Journal of Biomedical Semantics, 2011. **2 Suppl 2**: p. S1.
8. Sahoo, S.S., et al., *An ontology-driven semantic mashup of gene and biological pathway information: application to the domain of nicotine dependence*. J Biomed Inform, 2008. **41**(5): p. 752-65.
9. Belleau, F., et al., *Bio2RDF: towards a mashup to build bioinformatics knowledge systems*. J Biomed Inform, 2008. **41**(5): p. 706-16.
10. Callahan, A., J. Cruz-Toledo, and M. Dumontier, *Ontology-Based Querying with Bio2RDF's Linked Open Data*. J Biomed Semantics, 2013. **4 Suppl 1**: p. S1.
11. Knox, C., et al., *DrugBank 3.0: a comprehensive resource for 'omics' research on drugs*. Nucleic Acids Res, 2011. **39**(Database issue): p. D1035-41.
12. Whirl-Carrillo, M., et al., *Pharmacogenomics knowledge for personalized medicine*. Clin Pharmacol Ther, 2012. **92**(4): p. 414-7.
13. Chelliah, V., C. Laibe, and N. Le Novère, *BioModels Database: A Repository of Mathematical Models of Biological Processes*. Methods Mol Biol, 2013. **1021**: p. 189-99.

14. Wilkinson, M.D., B. Vandervalk, and L. McCarthy, *The Semantic Automated Discovery and Integration (SADI) Web service Design-Pattern, API and Reference Implementation*. Journal of biomedical semantics, 2011. **2**(1): p. 8.
15. Wilkinson, M.D., et al., *SADI, SHARE, and the in silico scientific method*. BMC bioinformatics, 2010. **11 Suppl 12**: p. S7.
16. Chepelev, L.L., et al., *Prototype semantic infrastructure for automated small molecule classification and annotation in lipidomics*. BMC Bioinformatics, 2011. **12**: p. 303.
17. Vandervalk, B., et al., *The SADI Personal Health Lens: A Web Browser-Based System for Identifying Personally Relevant Drug Interactions*. JMIR Res Protoc, 2013. **2**(1): p. e14.
18. Smith, B. and W. Ceusters, *Ontological realism: A methodology for coordinated evolution of scientific ontologies*. Appl Ontol, 2010. **5**(3-4): p. 139-188.
19. Dumontier, M. and R. Hoehndorf. *Realism for scientific ontologies*. in FOIS. 2010.
20. Lord, P. and R. Stevens, *Adding a little reality to building ontologies for biology*. PLoS One, 2010. **5**(9): p. e12258.
21. Smith, B., et al., *Relations in biomedical ontologies*. Genome Biol, 2005. **6**(5): p. R46.
22. Smith, B., et al., *The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration*. Nature biotechnology, 2007. **25**(11): p. 1251-5.
23. Stenzhorn, H., E. Beisswanger, and S. Schulz, *Towards a top-domain ontology for linking biomedical ontologies*. Stud Health Technol Inform, 2007. **129**(Pt 2): p. 1225-9.
24. Schulz, S., et al., *Scalable representations of diseases in biomedical ontologies*. J Biomed Semantics, 2011. **2 Suppl 2**: p. S6.
25. Seddig-Raufie, D., et al., *Proposed actions are no actions: re-modeling an ontology design pattern with a realist top-level ontology*. J Biomed Semantics, 2012. **3 Suppl 2**: p. S2.
26. Schulz, S., et al., *Alignment of the UMLS semantic network with BioTop: methodology and assessment*. Bioinformatics, 2009. **25**(12): p. i69-76.
27. Brinkman, R.R., et al., *Modeling biomedical experimental processes with OBI*. J Biomed Semantics, 2010. **1 Suppl 1**: p. S7.