Foundational Process Relations in Bio-Ontologies

Atalay ÖZGÖVDE [a] and Michael GRÜNINGER [b]
[a] *Department of Computer Science, University of Toronto*
[b] *Department of Mechanical and Industrial Engineering*

Abstract. This paper aims to provide rigor to the process relations in the OBO relations ontology. We present a formal axiomatization of these relations in first order logic as a definitional extension of the Process Specification Language (PSL) Ontology. In particular, we formalize the process parthood, process specialization, process instantiation and instance level parthood relations of OBO. We also present a formal characterization of the models of our axiomatization and provide representation theorems showing the correctness of our characterization.

1. Introduction

Although a wide variety of biological and medical ontologies have been proposed, their axiomatizations in a logical language have often been insufficient for automated reasoning and semantic integration. In fact, many are not specified in a logical language, and even those which are axiomatized have not been evaluated with respect to their intended models.

Open Biomedical Ontologies(OBO)([1], [7]) is an ontology repository shared across several biological and medical domains. OBO Foundry [10] is an initiative within the OBO community to develop principles that enhance compatibility and consistency of OBO ontologies. One of these principles is that ontologies use relations in ways consistent with their definitions as outlined in the *Relations in biomedical ontologies* paper([11]). The authors refer to the set of core relations described in the paper as OBO relation ontology(RO).

OBO Relation Ontology however provides definitions for only class level relations which connect classes of entities such as the class of all genes, the class of all instances of DNA replication and so forth. Although the definitions of the class level relations rely on instance level relations, a formal characterization of the instance level relations is not provided. In ([2]), axiomatic specifications are given for logical properties of the instance level mereogeometrical relations in the Relation Ontology. In this paper we take a similar approach to characterize the process relations in the Relation Ontology. The primary motivation for this work is the recognition that the distinction between classes and instances alone is inadequate for characterizing the intended semantics of the relations for a process ontology.

Table 1 shows the process relations in the OBO Relation Ontology; since the same relation name is often used for both class-level and instance-level relations, we introduce some new relation names to allow us to distinguish between these two cases. The resulting nonlogical lexicon is also shown in Table 1, and we refer to the axiomatization of the intended semantics of this lexicon as the OBO Process Ontology. In this paper, we show that the OBO Process Ontology can be axiomatized as a definitional extension of the Process Specification Language (PSL) Ontology ([3], [4]) [1]. We specify translation definitions ([5]) for all of the relations in the nonlogical lexicon of OBO Process Ontology, which provides a first-order axiomatization. The OBO Process Ontology does not have an explicit axiomatization in a logical language. This paper is therefore not an exercise in ontology mapping, which presumes that both ontologies are axiomatized, but rather illustrates the methodology of augmenting an informally specified ontology with the axiomatization of a first-order ontology.

In addition to axiomatizing the intended semantics of the process terminology within the OBO Relation Ontology, we provide a formal specification of the models of the axiomatization. We also illustrate several propositions that are consequences of the axiomatization, demonstrating that the OBO Process Ontology is capable of supporting automated first-order reasoning with process descriptions.

RO Relation	1.arg	2.arg	3.arg	Nonlogical Lexicon
p instance_of P	instance	process	-	instance_of
p part_of p_1	instance	instance	-	subinstance
t earlier t_1	timepoint	timepoint	-	earlier
p has_participant c at t	process	continuant	timepoint	has_participant_at
p has_agent c at t	process	continuant	timepoint	has_agent_at
P is_a P_1	process	process	-	is_a
P part_of P_1	process	process	-	subprocess
p occurring_at t	instance	timepoint	-	occurring_at
p preceded_by p_1	instance	instance	-	preceded_by
t first_instant p	timepoint	instance	-	first_instant
t last_instant p	timepoint	instance	-	last_instant
p immediately_preceded_by p_1	instance	instance	-	immediately_preceded_by
P preceded_by P_1	process	process	-	preceded_by

Table 1. Process relations in the relation ontology and the corresponding lexicon used in this text. Parameter types: P - process class, C - continuant class, p - process instance, c - continuant instance, t - timepoint.

2. An example biological process

We will use the cell division process as the running example in order to clarify the axiomatic approach and its benefits. Cell division is a process that results in the division of a parent cell into two or more daughter cells. There are two types of cell division. Mitotic cell division is the process by which the parent nucleus divides to produce two new nuclei with the same number of chromosomes, identical to the parent nucleus; the purpose of mitotic cell division is the growth, repair and replacement of cells. Meiotic cell division results in four daughter cells, in which the number of chromosomes in the

[1]The PSL Ontology has been published as an international standard ISO 18629. The complete set of axioms for the PSL Ontology can be found at `http://www.mel.nist.gov/psl/psl-ontology/`.

new cells are reduced by half; the purpose of meiotic cell division is sexual reproduction. Figure 1 shows the parthood and specialization structure of the cell division process; this is a representation of the structures which will be axiomatized by the process ontology.

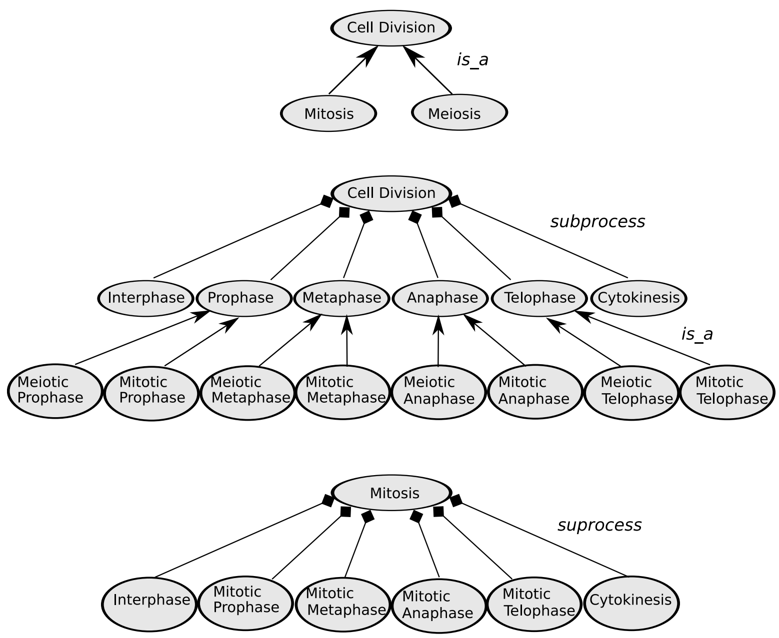

Figure 1. Diagrammatic view of the cell division process. Diamond arrows denote parthood, arrows denote specialization.

Cell division has six phases. During Interphase, cells carry out their normal function. At the later stages of interphase, cells prepare for mitosis or meiosis. During prophase, metaphase, anaphase and telophase the cytoplasm and nucleus of the dividing cell replicate and regroup into two sets in the opposite poles. Cytokinesis is the final stage where the cytoplasm and cell membrane splits into two. The example is a much simplified representation of cell division and provided for illustrative purposes; we will not attempt to fully axiomatize the cell division process.

3. Processes and Occurrences

Processes are repeatable behaviors whose occurrences cause continuants to undergo changes. It is important to distinguish between processes and process occurrences. Processes are neither endurants (continuants) nor perdurants (occurrents), since they do not change and they do not have temporal parts. Process occurrences are perdurants – they may have temporal parts (i.e. sub-occurrences such as changing the coffee filter while making coffee), and they have beginning and end timepoints.

The intuitive semantics of the RO process relations as described in ([11]) correspond very naturally to the ontological commitments made by the core theories in the PSL Ontology. The PSL Ontology axiomatizes the relationship between processes and process occurrences (instantiation), processes and subprocesses (process parthood) and process occurrences and subprocess occurrences (instance level parthood). In fact, one of the reasons for choosing the PSL Ontology as the basis for the axiomatization in this paper is that it is one of the few first-order ontologies in which both complex processes and their occurrences are elements of the domain.

The current version of the OBO relation ontology treats processes as classes and the *instance_of* relation as a counter part of the set-theoretic membership relation. In the PSL Ontology, processes and process instances are reified i.e. they are first class entities. In fact, most ontologies of action and change (see [8], [9] or [12]) favor reified represention for actions which allows actions to be arguments to relations and functions and expression of formulas that quantify over actions. This is also the approach taken in Thomas Bittner's Isabelle axiomatization of the mereogeometrical relations in the OBO Relations Ontology (See `http://isabelle.in.tum.de`).

Within the PSL Ontology, the *occurrence_of* relation associates processes with their occurrences. Each process occurrence is associated with a unique process, although processes can have multiple occurrences. We can characterize the *instance_of* relation of OBO in terms the *occurrence_of* relation of PSL. The axiomatization of the $instanceOf$ relation is closely tied to the formalization of activity specialization, which is addressed in Section 4.

3.1. Subactivities

Primitive processes are at the lowest level of granularity that a domain designer chooses to represent, and they are responsible for all the physical change in the world. A primitive process is specified in terms of the conditions that are necessary for its occurrence (preconditions) and the impact of its occurrences on the world (effects). Occurrences of complex processes correspond to sequences of occurrences of primitive processes; consequently, a complex process is specified in terms of dependencies and constraints among occurrences of its subprocesses.

Complex activity occurrences are also first class objects in the PSL ontology. The PSL ontology characterizes complex processes in terms of the relationship between occurrences of complex processes and occurrences of their subprocesses. Subprocesses occur during the occurrence of the process, however subprocesses can also occur externally. In other words, an occurrence of a subprocess during occurrence of a process may be coincidental. The PSL Ontology uses the *subactivity_occurrence* relation to associate complex activity occurrences with the occurrences of its subprocesses.

The OBO Process Ontology includes a composition relation that is defined on processes. The PSL Ontology uses the *subactivity* relation to capture the intuitions for process parthood. The $subactivity$ relation itself is a discrete partial ordering. An activity may have subactivities that do not occur and an activity may be a subactivity of multiple activities. Different subactivities may occur in different occurrences of an activity, with the requirement that subactivity occurrences are occurrences of subactivities. The axiomatization of the $subactivity$ relation in the PSL Ontology allows models where processes that share subactivities do not co-occur.

Process parthood is a class level relation in the current version of OBO Relation Ontology. Formal definitions for class level relations are given in terms of their instance level counter parts in the ontology. Process parthood definition is also given in the same style. Based on the discussion above, one could write the definition of process parthood in the OBO Relation Ontology as follows:

$$\forall a, a_1.subProcess(a_1, a) \equiv \forall o_1.occurrence\_of(o_1, a_1) \supset \exists o.occurrence\_of(o, a) \wedge subactivity\_occurrence(o_1, o)$$

However, this definition requires every instance of a subprocess be part of an instance of the process it is part of. Consider the cytokinesis phase of the cell cycle. Cytokinesis is a subProcess of both mitosis and meiosis. The definition above amounts to saying that whenever cytokinesis occurs, both mitosis and meiosis must occur such that the cytokinesis occurrence is part of their occurrences. Cytokinesis may also be part of other processes such as cancerious cell division. Occurrence of cytokinesis indicates occurrence of one of mitosis, meiosis or a cancerious growth mutually exclusively.

Following this discussion, we can introduce the translation definitions for OBO Process Ontology relations associated with $partOf$:

$$\forall a, a_1.subProcess(a_1, a) \equiv subactivity(a_1, a) \quad (1)$$

$$\forall s, o.subinstance(s, o) \equiv subactivity\_occurrence(s, o) \quad (2)$$

Note that the constraints between the *subactivity* and *subactivity_occurrence* relations in PSL do not require that processes that share subactivities necessarily co-occur.

3.2. Orderings over Occurrences and Time

The next set of relations in the OBO Process Ontology focus on the ordering relations on process occurrences and the timepoints at which processes occur. Some approaches ([6], [9]), do not distinguish between timepoints and process occurrences, so that process occurrences form a subset of timepoints, and the ordering relations are conflated to the ordering over timepoints. Nevertheless, it is important to maintain a distinction between timepoints and process occurrences; although each occurrences is associated with a beginning and ending timepoints, occurrences have preconditions and effects, whereas timepoints do not.

In the PSL Ontology, timepoints are distinct objects from activities and activity occurrences and they form an infinite linear ordering corresponding to the *before* relation. This allows us to specify the following translation definition for the ordering over timepoints in the OBO Process Ontology:

$$\forall t_1, t_2.earlier(t_1, t_2) \equiv before(t_1, t_2) \quad (3)$$

In addition to the ordering over timepoints, the PSL Ontology axiomatizes an ordering over process occurrences known as an occurrence tree; this is a discrete partially ordered set of primitive activity occurrences containing all sequences of activity occurrences [2]. Since not all of these sequences will intuitively be physically possible within a given domain, we consider the subtree of the occurrence tree that consists only of possible

[2]Occurrence trees are similar to the situation trees that are models of Reiter's axiomatization of the situation calculus ([8]).

sequences of activity occurrences, which we refer to as the legal occurrence tree. The $legal(o)$ relation specifies that the primitive activity occurrence o is an element of the legal occurrence tree. The PSL Ontology uses the *precedes* relation to axiomatize the partial ordering of the occurrences in the legal occurence tree, leading to the following translation definition for the OBO Process Ontology ordering relation over process occurrences:

$$(\forall p, p_1)\ preceded\_by(p, p_1) \equiv (\exists o, o_1)\ root\_occ(o, p) \wedge leaf\_occ(o_1, p_1) \wedge precedes(o_1, o) \tag{4}$$

Since each branch of the occurrence tree is a discrete linear ordering of process occurrences, the PSL Ontology uses the $successor$ function to denote the next occurrence of a particular process within the tree; this captures the intended semantics of the following relation in the OBO Process Ontology:

$$(\forall p, p_1)\ immediately\_preceded\_by(p, p_1) \equiv (\exists a)\ (p_1 = successor(a, p)) \tag{5}$$

Occurrences along a branch of the occurrence are linearly ordered and the $begin\_of$ and end_of functions in map occurrences to the underlying timeline; this corresponds to the following OBO Process Ontology relations:

$$(\forall p, t)\ first\_instant(t, p) \equiv (t = beginof(p)) \tag{6}$$

$$(\forall p, t)\ last\_instant(t, p) \equiv (t = endof(p)) \tag{7}$$

Interestingly, the relations within the OBO Process Ontology that associates objects and occurrences with the different timepoints over which the process is occurring is identical in name to the corresponding relation in the PSL Ontology:

$$(\forall o, t)\ occurring\_at(o, t) \equiv occurring\_at(o, t) \tag{8}$$

$$(\forall a, x, t)\ has\_participant(a, x, t) \equiv (\exists o)\ occurrence\_of(o, a) \wedge participates\_in(x, o, t) \tag{9}$$

Occurrences of complex processes are not part of the occurrence tree. Figure 2, shows a possible model of occurrences of mitosis and meiosis processes, $o_1^{mitosis}$ and $o_2^{meiosis}$ respectively. In the figure occurrences are enumerated to distinguish distinct occurrence objects, and superscripts are the associated processes; for example, $s_1^{interphase}$ denotes the particular occurrence, s_1, in the model that is an occurrence of the interphase process. The dotted arrows indicate the *subactivity_occurrence* relation. The occurrences s_1, s_9, s_{10}, s_{378}, s_{380} and s_{381} are subactivity occurrences of $o_1^{mitosis}$ and they are part of the occurrence tree.

The OBO Relation Ontology includes an ordering relation $preceded\_by$ over processes, in addition to process occurrences. Unfortunately, partial ordering relations over processes are not well-defined in cases where there are multiple occurrences of the same processes.

4. Process specialization

A specialization of an activity is one way in which the activity can occur. The axiomatization of the OBO Process Ontology therefore treats specialization as a special kind

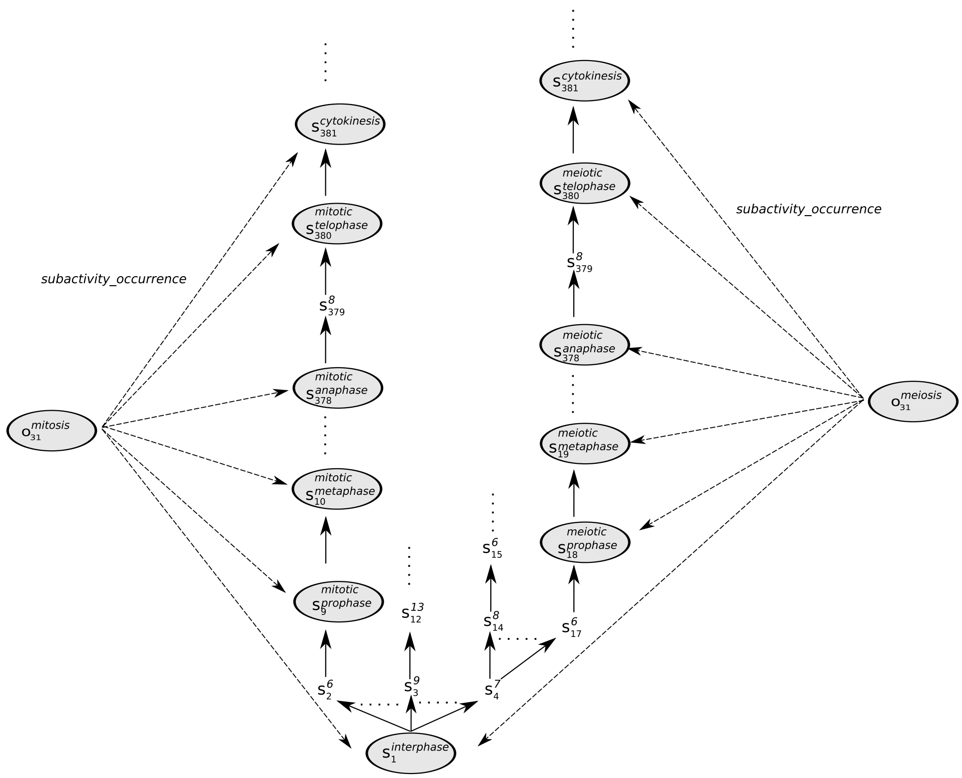

Figure 2. A fragment of a possible PSL model of occurrences of mitosis and meiosis. Note that s^y_x denotes an occurrence of process y.

of nondeterminism, since complex processes capture the different ways of aggregating primitive process occurrences. Intuitively, the more general the process, the more nondeterministic it is, while processes become more deterministic as they get more specialized. In this approach, a specialization of a process is a nondeterministic subprocess, so that occurrences of a specialized process correspond to a subset of the occurrences of the more general process. Consider the cell division example. An occurrence of cell division corresponds to the occurrence of one of mitosis or meiosis; either of these is a specialization of cell division. In addition, an occurrence of cell division involves an occurrence of prophase, which can be mitotic or meiotic depending on which specialization of cell division occurs.

4.1. Axiomatization

Recall that the notion of process occurrence in PSL is distinguished from the notion of a process instance. For example, when mitosis occurs there is also an occurrence of cell division; however these two occurrences are distinct even though one process is a specialization of the other. What is common between these two occurrences is that they share the same primitive subprocess occurrences. We call such occurrences *coextensive*, and introduce the following conservative definition using the PSL Ontology.

$$\forall o_1, o_2.coextensive(o_1, o_2) \equiv activity\_occurrence(o_1) \wedge activity\_occurrence(o_2) \wedge$$
$$(\forall s.legal(s) \supset (subactivity\_occurrence(s, o_1) \equiv subactivity\_occurrence(s, o_2))) \tag{10}$$

Using this notion, we can axiomatize the isA relation in terms of coextensive process occurrences, with the additional property that specializations of a process do not occur outside occurrences of the more general processes:

$$\begin{aligned}
&\forall a_1, a_2.isA(a_1, a_2) \equiv \\
&\ activity(a_1) \wedge activity(a_2) \wedge subactivity(a_1, a_2) \wedge \\
&\ (\forall o_1.occurrence_of(o_1, a_1) \supset \exists o_2.occurrence_of(o_2, a_2) \wedge \\
&\ coextensive(o_1, o_2) \wedge subactivity_occurrence(o_1, o_2))
\end{aligned} \tag{11}$$

The isA relation can therefore be seen as a restriction of the $subactivity$ relation in the PSL Ontology; we can define a corresponding restriction to the $occurrence\_of$ relation:

$$(\forall o, a)\ instanceOf(o, a) \equiv \exists a_1.isA(a_1, a) \wedge occurrence\_of(o, a_1) \tag{12}$$

4.2. Consequences

Let T_{obo} be the set of axioms 1 - 12 and T_{psl} be the set of axioms in the PSL Ontology. the following results are the consequences of the axioms [3].

Theorem 1

$$T_{obo} \cup T_{psl} \models [(\forall a, b) isA(a, b) \equiv ((\forall o_1)\ instanceOf(o_1, a) \supset$$
$$((\exists o_2)\ instanceOf(o_2, b) \wedge ((\forall s)\ partOf(s, o_1) \equiv partOf(s, o_2))))]$$

Informally, the next theorem follows from the *subactivity_occurrence* relation being reflexive, transitive and antisymmetric and the *coextensive* relation being an equivalence relation.

Theorem 2 *The isA relation is reflexive, transitive and antisymmetric:*

$$T_{obo} \cup T_{psl} \models (\forall a)\ isA(a, a)$$
$$T_{obo} \cup T_{psl} \models (\forall a_1, a_2, a_3)\ isA(a_1, a_2) \wedge isA(a_2, a_3) \supset isA(a_1, a_3)$$
$$T_{obo} \cup T_{psl} \models (\forall a_1, a_2)\ isA(a_1, a_2) \wedge isA(a_2, a_1) \supset (a_1 = a_2)$$

We can next show that our characterization of the *isA* relation respects the way the *isA* relation defined in the OBO Relation Ontology.

Theorem 3

$$T_{obo} \cup T_{psl} \models (\forall a, b)\ isA(a, b) \equiv ((\forall o)\ instanceOf(o, a) \supset instanceOf(o, b))$$

---

[3] Due to space limitations, we do not include proofs of these two theorems in this paper, although proofs have been generated automatically using the Prover9 theorem prover.

Proof $\Rightarrow$: Let a, b be such that $isA(a,b)$ holds also let o be such that $instanceOf(o,a)$ holds. Then there exists a_1 such that $isA(a_1,a)$ and $occurrenceOf(o,a_1)$. By theorem 2, $isA(a_1,b)$ also holds. Then by the definition of the instanceOf relation, $instanceOf(o,b)$ follows.

$\Leftarrow$: Let $\forall o.instanceOf(o,a) \supset instanceOf(o,b)$ and let o, a be such that $occurrenceOf(o,a)$. By the definition of the *instanceOf* relation and theorem 2, $instanceOf(o,a)$ holds. Then we also have that $instanceOf(o,b)$. Since every occurrence is an occurrence of a unique activity, using the definition of the instanceOf relation again we conclude that $isA(a,b)$. □

Finally, we can prove that the translation definitions for several of the OBO Process Ontology relations are logically equivalent to straightforward transcriptions of their intended semantics as discussed in [11]:

Theorem 4 $T_{obo} \cup T_{psl} \models$

$$\begin{aligned}
occurring_at(o,t) \equiv & \exists t_1, t_2, s_1, s_2.root_occurrence(s_1,o) \wedge \\
& leaf_occurrence(s_2,o) \wedge begin_of(s_1) = t_1 \wedge end_of(s_2) = t_2 \wedge \\
& (earlier(t_1,t) \vee t_1 = t \vee earlier(t,t_2)) \\
preceded_by(p,p_1) \equiv & \forall t_1, t_2, occurring_at(p,t) \wedge occurring_at(p_1,t_1) \supset earlier(t_1,t) \\
first_instant(t,p) \equiv & occurring_at(p,t) \wedge \forall t_1.earlier(t_1,t) \supset \neg occurring_at(p,t_1) \\
last_instant(t,p) \equiv & occurring_at(p,t) \wedge \forall t_1.earlier(t,t_1) \supset \neg occurring_at(p,t_1)
\end{aligned}$$

5. Reasoning with processes

Subprocesses and associated specialization hierarchy do not completely characterize a process. Two processes that share the same set of subprocesses and parents may be distinct in terms of dependencies and dynamics among their subprocess occurrences. In order to fully represent a process, a description must be given that specifies the subprocesses that occur as part of its occurrences as well as constraints on the subprocess occurrences. For example, the description of the mitosis process can be specified as:

Process Description $\Sigma_{mitosis}$ (The mitosis process):

$$\begin{aligned}
& \forall o, instance_of(o, mitosis) \supset \exists s_1, s_2, s_3, s_4, s_5, s_6. \\
& \quad instance_of(s_1, interphase) \wedge instance_of(s_2, mitoticProphase) \wedge \\
& \quad instance_of(s_3, mitoticMetaphase) \wedge instance_of(s_4, mitioticAnaphase) \wedge \\
& \quad instance_of(s_5, mitoticTelophase) \wedge instance_of(s_6, cytokinesis) \wedge \\
& \quad subinstance(s_1,o) \wedge subinstance(s_2,o) \wedge subinstance(s_3,o) \wedge \\
& \quad subinstance(s_4,o) \wedge subinstance(s_5,o) \wedge subinstance(s_6,o) \wedge \\
& \quad preceded_by(s_2,s_1) \wedge preceded_by(s_3,s_2) \wedge preceded_by(s_4,s_3) \wedge \\
& \quad preceded_by(s_5,s_4) \wedge preceded_by(s_6,s_5)
\end{aligned}$$

One of the advantages of designing a first-order logic based theory is that one can use an off-the-shelf generic first-order theorem prover such as Prover9 or Vampire to validate the theorems of the theory or to do automated reasoning at the domain level. For example, we can express the property that whenever *mitosis* occurs there are occurrences of *mitotic_metaphase* and *mitotic_anaphase* such that the occurrence of *mitotic_anaphase* is preceded by the occurrence of *mitotic_metaphase*; moreover, this is a consequence of Process Description for mitosis, together with the axioms of the ontology:

$$
\begin{aligned}
T_{obo} \cup T_{psl} \cup \Sigma_{mitosis} \models (\forall o) instance\_of(o, mitosis) \supset \\
(\exists s_1, s_2)\, instance\_of(s_1, mitotic\_metaphase) \\
\wedge instance\_of(s_2, mitotic\_anaphase) \wedge preceded\_by(s_2, s_1)
\end{aligned}
$$

Suppose we included a description for the *meiosis* process similar to the one given for the *mitosis* process. Furthermore, suppose we also added constraints for the *subprocess* and *is_a* relations to capture the cell division process as depicted in Figure 1, $is\_a(mitosis, cell\_division)$, $subprocess(mitotic\_prophase, mitosis)$, $is\_a(meiotic\_telophase, telophase)$ and so forth. Then we could also prove the following property:

$$
\begin{aligned}
T_{obo} \cup T_{psl} \cup \Sigma_{mitosis} \models (\forall o)\, instance\_of(o, cell\_division) \supset \\
(\exists s_1, s_2)\, instance\_of(s_1, metaphase) \wedge \\
instance\_of(s_2, anaphase) \wedge preceded\_by(s_2, s_1)
\end{aligned}
$$

Note that a description for the cell division process is not required to prove the property above. Also note that *metaphase* and *anaphase* are abstract processes that occur only through their mitotic or meiotic specializations.

Reasoning with the OBO Process Ontology is not limited to reasoning about the static structure of complex processes. Making predictions about the future and explanations about the past often involve the state of the world and how processes change the state. PSL includes a sub-theory that is intended to capture the basic intuitions about states and their connections to processes.

A state is intuitively a set of fluents and fluents are parameters of a domain that are subject to change through occurrences processes. The effect axioms of processes describe how their occurrences change fluents. Temporal mereogeometrical relations axiomatized in ([2]), such parthood, location and connection relations would be fluents in an integrated theory. For example, the effect axioms for the *cytokinesis* process would describe how the spacial properties of the participating cell change as a result of the occurrence of the process.

6. Verification of the Ontology

In this section, we provide a formal characterization of the models of the axiomatization of the OBO Process Ontology, and propose theorems that demonstrate that the models of the axioms are equivalent to the models in the characterization. Although a complete characterization has been done for all of the axioms in the ontology, space restrictions only allow us to specify the models of the axioms for the isA and $instanceOf$ relations in this paper.

The ontology is verified by providing a complete characterization of all models of the axioms up to isomorphism. One approach to this problem is to use representation theorems – we evaluate the adequacy of the ontology with respect to some well-understood class of mathematical structures (such as partial orderings, graph theory, and geometry) that capture the intended interpretations of the ontology's terms. Given the definition of some class of structures $\mathfrak{M}$, we prove that the class exists and is nonempty, which also provides a characterization of the structures in the class up to isomorphism. We prove that every structure in the class is a model of the ontology and that every countable model of the ontology is isomorphic to some structure in the class.

6.1. Models of the PSL Ontology

The models of the axioms of the PSL Ontology have been characterized up to isomorphism [3]. We first review some of the underlying intuitions before defining the fundamental structures that constitute the models of the PSL Ontology.

6.1.1. Intuitions

The basic structure that characterizes occurrences of complex activities within models of the ontology is the activity tree, which is a subtree of the legal occurrence tree that consists of all possible sequences of primitive subactivity occurrences; the relation $root(s, a)$ denotes that the subactivity occurrence s is the root of an activity tree for a, and relation $leaf(s, a)$ denotes that the subactivity occurrence s is the leaf of an activity tree for a. The $min\_precedes(s_1, s_2, a)$ relation is used to denote that subactivity occurrence s_1 precedes the subactivity occurrence s_2 in occurrences of the complex activity a. Elements of an activity tree are ordered by the $min\_precedes$ relation; each branch of the activity tree is a linearly ordered set of occurrences of subactivities of the complex activity.

In a sense, an activity tree is a microcosm of the occurrence tree, in which we consider all of the ways in which the world unfolds in the context of an occurrence of the complex activity. Different subactivities may occur on different branches of the activity tree – different occurrences of an activity may have different subactivity occurrences or different orderings on the same subactivity occurrences.

6.1.2. Basic Structures for Models of the PSL Ontology

To formalize the intuitions in the previous section, we provide the following definitions for the structures which are used to specify models of the PSL Ontology; they will also be used to specify the models of the OBO Process Ontology in the next section.

The structure that corresponds to the $subactivity$ relation is the following partial ordering:

Definition 1 *A subactivity ordered set $\mathcal{A} = \langle A, \prec \rangle$ is a discrete partial ordering over the set of activities A.*

The subactivity ordered set specifies how complex activities can be decomposed into subactivities; the activity tree specifies how occurrences of a complex activity correspond to occurrences of its subactivities.

Definition 2 *An activity tree τ for an activity $\mathbf{a}$ is a subtree of a legal occurrence tree such that*

- *all elements of τ are occurrences of primitive subactivities of $\mathbf{a}$;*
- *τ contains a unique element if $\mathbf{a}$ is primitive.*

Definition 3 *A complex activity structure $\mathcal{C}$ is the union of all activity trees for all activities in the subactivity ordered set $\mathcal{A}$.*

In addition, there is a one-to-one correspondence between occurrences of complex activities and branches of the associated activity trees, which is captured by the following mapping:

Definition 4 *Let $\mathcal{C}$ be a complex activity structure. Let $\mathcal{B}$ be the set of branches in activity trees in $\mathcal{C}$, and let O be the set of activity occurrences of complex activities.*

A complex mapping $\beta : \mathcal{B} \rightarrow O$ is a bijection such that if $\mathbb{B}$ is a branch of an activity tree for a complex activity $\mathbf{a}$, then $\langle \beta(\mathbb{B}), \mathbf{a} \rangle \in \mathbf{occurrence\_of}$

The axioms for $subactivity\_occurrence$ relation guarantee that the branches of the activity trees for a subactivity are contained in the branches of the activity tree for the complex activity.

6.2. Models of OBO Process Ontology

6.2.1. Mappings on Activity Trees

Before we define the class of structures that formalizes the notion of activity specialization, we first need to specify a mapping that captures the relationships between activity trees, and consequently the relationships between the occurrences of the activities. Given the definition of activity trees, any mapping that preserves the activity tree needs to preserve the orderings over the elements of the activity trees and it must also preserve the relationship between elements of the activity trees and occurrences of subactivities of $\mathbf{a}$.

Definition 5 *Let τ_1, τ_2 be activity trees for the activities $\mathbf{a_1}$, $\mathbf{a_2}$, respectively.*

A mapping $\varphi : \tau_1 \rightarrow \tau_2$ is a full embedding iff

1. $\langle \mathbf{s_1}, \mathbf{s_2}, \mathbf{a_1} \rangle \in \mathbf{min\_precedes} \Leftrightarrow \langle \varphi(\mathbf{s_1}), \varphi(\mathbf{s_2}), \mathbf{a_2} \rangle \in \mathbf{min\_precedes}$*;*
2. $\langle \mathbf{s_1}, \mathbf{s_2}, \mathbf{a_1} \rangle \in \mathbf{leaf} \Leftrightarrow \langle \varphi(\mathbf{s_1}), \varphi(\mathbf{s_2}), \mathbf{a_2} \rangle \in \mathbf{leaf}$.

6.2.2. Classes of Structures

We begin by defining a partial ordering on the sets of activity trees.

Definition 6 *Let $\mathcal{C}$ be a complex activity structure.*

$\langle \mathcal{C}^s, \subseteq \rangle$ is a partial ordering on the set of activity trees in $\mathcal{C}$ such that $\tau_i \subseteq \tau_j$ iff there is a full embedding of τ_i into τ_j.

One intuition behind activity specialization is that an activity does not occur outside of an occurrence of any of its generalizations. This is formalized as the requirement that all activity trees for the activity are embedded within the activity trees of the generalized activities. A second intuition is that an activity's specializations correspond to the different ways in which an activity can occur, so that specialized activities are subactivities of the general activities. These two considerations lead us to the following definition of the structure which will ultimately be used to characterize the extension of the $\mathbf{isA}$ relation:

Definition 7 *Let $\mathcal{A}$ be a subactivity ordered set, let $\mathcal{C}$ be a complex activity structure, and let $\varphi : \mathcal{C} \rightarrow \mathcal{A}$ be a surjective mapping such that $\varphi(\tau) = \mathbf{a}$ iff τ is an activity tree for the activity $\mathbf{a}$.*

A substructure $\mathbb{S} \subseteq \mathcal{A}$ is specialized subactivity ordered set iff φ maps ideals in $\langle \mathcal{C}, \subseteq \rangle$ to ideals in $\mathbb{S}$.

The ideal of an activity in the partial ordering is equal to the set of all activities that are generalizations of the activity. If $\mathbf{a}$ is an activity in the specialized subactivity ordered set and τ is an activity tree for $\mathbf{a}$, then the set of activity trees containing τ (i.e. the ideal in $\langle \mathcal{C}, \subseteq \rangle$) is mapped to the set of activities containing $\mathbf{a}$ as a subactivity (i.e. the ideal in $\mathcal{A}$). This guarantees that there are no activity trees for an activity that are not subtrees of the activity trees for its generalizations.

The second class of structures represents the extension of the $\mathbf{instanceOf}$ relation.

Definition 8 *Let (A, O, E) be a bipartite graph consisting of activities A and activity occurrences O.*

Suppose $N(\mathbf{a_i}) = \{\mathbf{o} : (\mathbf{o}, \mathbf{a}) \in E\}$.

$\mathbb{I} = \langle A, O, E \rangle$ is an instance graph iff $(\bigcup_i N(\mathbf{a_i}), \subseteq) \cong \mathbb{S}$ where $\mathbb{S}$ is the specialized subactivity ordered set.

6.2.3. Introducing $\mathfrak{M}^{obo}$

Definition 9 *Let $\mathfrak{M}^{obo}$ be the class of structures such that $\mathcal{M} \in \mathfrak{M}^{obo}$ we have*

1. *$\mathcal{M} = \mathcal{N} \cup \mathbb{S}$ such that*
   - *(a) $\mathcal{N}$ is a model of T_{psl};*
   - *(b) $\mathbb{S} = \langle A, \sqsubseteq \rangle$ is a specialized subactivity ordered set.*
   - *(c) $\mathbb{I} = \langle A, O, E \rangle$ is an instance graph.*
2. *$\langle \mathbf{o_1}, \mathbf{o_2} \rangle \in \mathbf{coextensive}$ iff $\beta^{-1}(\mathbf{o_1}) = \beta^{-1}(\mathbf{o_2})$.*
3. *$\langle \mathbf{a_1}, \mathbf{a_2} \rangle \in \mathbf{isA}$ iff $\mathbf{a_1} \sqsubseteq \mathbf{a_2}$.*
4. *$\langle \mathbf{o}, \mathbf{a} \rangle \in \mathbf{instanceOf}$ iff $(\mathbf{o}, \mathbf{a}) \in E$.*

If $T_{special}$ is the subtheory of T_{obo} containing the axioms for isA and $instanceOf$, then we can then prove the following characterization theorems:

Theorem 5 *Any structure $\mathcal{M} \in \mathfrak{M}^{obo}$ is a model of $T_{special} \cup T_{psl}$ and any countable model of $T_{special} \cup T_{psl}$ is isomorphic to a structure $\mathcal{M} \in \mathfrak{M}^{obo}$.*

7. Summary

In this paper, we have provided a first-order axiomatization of the intended semantics of the process relations within the OBO Relations Ontology. We did this by specifying translation definitions using the PSL Ontology, so that the axioms which constitute the OBO Process Ontology form a definitional extension of the PSL Ontology. We also characterized the models of the resulting axiomatization, thereby demonstrating the formal verification of the ontology.

With the axioms of the OBO Process Ontology and the PSL Ontology, we can specify process descriptions for biological processes, which cannot be done with existing ontologies. For example, the Biological Process Ontology of the Gene Ontology cannot represent biological pathways since biological processes in the ontology are specified only in terms of the associated molecular functions.

The OBO Process Ontology can also be used to define classes of queries associated with more traditional reasoning problems within knowledge representation, such as temporal projection, planning, (e.g. biochemical synthesis) and plan verification. This opens up new avenues for research in the application of automated reasoning to biomedical ontologies.

References

[1] Obo: Open biomedical ontologies. `www.obo.sourceforge.net/`.

[2] Thomas Bittner. Logical properties of foundational mereogeometrical relations in bio-ontologies. *Appl. Ontol.*, 4(2):109–138, 2009.

[3] M Grüninger. Guide to the ontology of the process specification language. In S Staab, editor, *Handbook of Ontologies in Information Systems*. Springer-Verlag, 2003.

[4] M Grüninger and C Menzel. Process Specification Language: Principles and applications. *AI Magazine*, 24(3):63–74, 2003.

[5] Michael Gruninger and Joseph B. Kopena. Semantic integration through invariants. *AI Mag.*, 26(1):11–20, 2005.

[6] P Hayes. A catalog of temporal theories. *Artificial Intelligence Technical Report UIUC-BI-AI-96-01*, 1996.

[7] Christopher J. Mungall. Obol: integrating language and meaning in bio-ontologies: Conference papers. *Comp. Funct. Genomics*, 5(6-7):509–520, 2004.

[8] Raymond Reiter. *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*. MIT Press, 2001.

[9] Murray Shanahan. The event calculus explained. *Lecture Notes in Computer Science*, 1600:409–430, 1999.

[10] Barry Smith, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis J. Goldberg, Karen Eilbeck, Amelia Ireland, Christopher J. Mungall, Neocles Leontis, Philippe Rocca-Serra, Alan Ruttenberg, Susanna-Assunta Sansone, Richard H. Scheuermann, Nigam Shah, Patricia L. Whetzel, and Suzanna Lewis. The obo foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25(11):1251–1255, November 2007.

[11] Barry Smith, Werner Ceusters, Bert Klagges, Jacob Köhler, Anand Kumar, Jane Lomax, Chris Mungall, Fabian Neuhaus, Alan L. Rector, and Cornelius Rosse. Relations in biomedical ontologies. *Genome biology*, 6(5):R46+, 2005.

[12] Heinrich Wansing. Agency and deontic logic, j.f. horty. *J. of Logic, Lang. and Inf.*, 13(3):379–381, 2004.