eHealth Beyond the Horizon – Get IT There
S.K. Andersen et al. (Eds.)
IOS Press, 2008

How Granularity Issues Concern Biomedical Ontology Integration

Stefan SCHULZ [a,1], Martin BOEKER [a] and Holger STENZHORN [a,b]
[a] *Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Germany*
[b] *Digital Enterprise Research Institute (DERI), Galway, Ireland*

Abstract. The application of upper ontologies has been repeatedly advocated for supporting interoperability between domain ontologies in order to facilitate shared data use both within and across disciplines. We have developed BioTop as a top-domain ontology to integrate more specialized ontologies in the biomolecular and biomedical domain. In this paper, we report on concrete integration problems of this ontology with the domain-independent Basic Formal Ontology (BFO) concerning the issue of fiat and aggregated objects in the context of different granularity levels. We conclude that the third BFO level must be ignored in order not to obviate cross-granularity integration.

Keywords. Formal ontology, domain ontology

Introduction

The application of upper ontologies (also known as top-level ontologies) has been advocated for supporting interoperability between domain ontologies in order to facilitate the shared use of data both within and across disciplinary boundaries. Ideally, any upper ontology should be restricted to only a small set of domain-independent and highly general categories (types), relations, and constraints. Currently, several upper ontologies co-exist with each one differing greatly in terms of their respective user community, topical focus and logico-philosophical foundations [7].

In biomedicine, the UMLS Semantic Network [6], the GALEN High Level Ontology [9], the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [1], the General Formal Ontology (GFO) [3] as well as the Basic Formal Ontology (BFO) [2], all have been applied for this purpose. The need for bridging the gap between domain-independent upper ontologies and the application-oriented domain ontologies has led to the proposal of a new kind of ontologies, viz. the top-domain ontologies, like the Simple Bio Upper Ontology [10], GFO-Bio [4], or BioTop [15]. Their purpose is to define the most general categories relevant for the entire biomedical domain, such as Protein, Tissue, DNA, or BiologicalFunction to be used as a common reference for more specialized domain ontologies.

[1] Corresponding Author: Stefan Schulz, Universitatsklinikum - Institut für Medizinische Biometrie und Medizinische Informatik, Stefan-Meier-Strasse 26, 79104 Freiburg, Germany, E-mail: stschulz@uni-freiburg.de

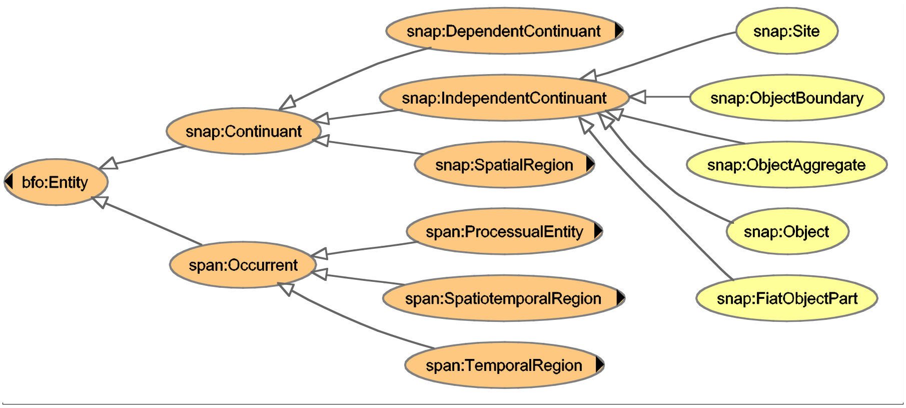

Figure 1. Excerpt from the BFO class tree

In the following we concentrate on the BioTop ontology[2] and its interface to BFO. The goal of BioTop is to provide classes and classificatory criteria to categorize the foundational kinds of biology, without any restriction to granularity, species, developmental stages or states of structural well- or ill-formedness [11]. In Bio Top's initial development, no definitive commitment existed towards any upper ontology, except for the distinction between continuants ("entities that endure through time while maintaining their identity and that are fully present at one single point in time", e.g., objects) and occurrents ("entities that are never fully present at any one given moment in time, but unfold themselves in successive phases, or temporal parts", e.g., processes), a principle on which most upper ontologies roughly agree. The primary focus at this point was primarily set on representing continuants from the area of interest. In the continued development the focus was broadened to include the representation of biological processes and BioTop was aligned with BFO. An analysis of the resulting implications of this alignment is at the center of the following deliberations.

1. Material and Methods

1.1. Upper Ontology Grounding

The choice of BFO was initiated by the following basic considerations:

- BFO's theoretical background had been extensively studied and scrutinized in logico-philosophical research on ontology for several years.
- An OWL-DL implementation of BFO was developed in a peer reviewed process and is freely available (with documentation) from its website[3].

Figure 1 depicts a smaller subtree of BFO, highlighting the strict subdivision of the class IndependentContinuant with Table 1 summarizing the definitions of the specific classes which form the basis for the subsequent discussion. It is important to point out that BFO classes on the same level are supposed to be mutually disjoint (i.e., a given instance cannot instantiate sibling classes such as Object and FiatObjectPart at the same time). Further, from the definitions it becomes obvious that the notion of

[2] BioTop website: http://purl.org/biotop
[3] BFO website: http://www.ifomis.org/bfo

"fiatness" [14] (e.g., the delimitation of a material entity without any physical discontinuity) is one of the most important differentiation criteria for the subclasses of IndependentContinuant.

BFO Class	Definition	Examples
Continuant	An entity that exists in full at any time in which it exists at all, persists through time while maintaining its identity and has no temporal parts.	
Independent-Continuant	A continuant that is a bearer of quality and realizable entities, in which other entities inhere and which itself cannot inhere in anything.	
Object	An independent continuant that is spatially extended, maximally self-connected and self-contained (the parts of a substance are not separated from each other by spatial gaps) and possesses an internal unity.	a cell, a body
Object-Aggregate	An independent continuant that is a mereological sum of separate object entities and possesses non-connected boundaries.	a heap of stones, a collection of bacteria
FiatObjectPart	An independent continuant that is part of an object but is not demarcated by any physical discontinuities.	the upper and lower lobes of the left lung, the dorsal and ventral surfaces of the body

Table 1. Textual definitions of the BFO classes used in this study

1.2. Ontology Modeling

The following principles have guided the development of BioTop: Firstly, full textual and logical definition of as many classes as possible to make the meaning of the various classes clear both to the human user as well as to automatic reasoning devices, and secondly, the avoidance of multiple is-a (inheritance) hierarchies because such hierarchies are intricately prone for modeling errors[4].

BioTop is modeled in OWL-DL (Ontology Web Language - Description Logics) using the Protégé editor [5] together with the Pellet reasoner [13] for terminological reasoning and consistency checking. Language, editor, and reasoner have been established as standards in the Semantic Web community. During its development, the evolving ontology was repetitively assessed in terms of consistency (i.e., the classifier detected inconsistent classes which were then manually analyzed and corrected), and adequacy (i.e., domain experts checked the inferred hierarchy for correctness after each classification and corrected it, if necessary).

Once the ontology had reached the targeted coverage and both a consistent and adequate state, its topmost classes (e.g., EntireMolecularEntity,, HomogenousCollection, HeterogenousCollection, Organism) were mapped to BFO by importing its OWL-DL implementation and attaching BioTop classes as children to the respective BFO classes (e.g., MolecularFunction as subclass of BFO's Function). In light of this integration, we encountered some phenomena which led us to critically appraise BFO's differentiating criteria.

[4] Given complete definitions, important taxonomic parents are later inferred by the terminological classifier anyway

2. Results

At a first glance, the inconsistencies found were mainly due to overlaps between the following subclasses of IndependentContinuant: Object, FiatObjectPart, ObjectAggregate (cf. Figure 1 and Table 1 for their respective BFO definition). As these classes are defined to be mutually disjoint, classification errors occurred wherever a BioTop class was found to fulfill the criteria to be subsumed by any two of these BFO classes (i.e., we did not assert multiple parents for a given BioTop class but rather let the reasoner infer such setting from the given set of logical restrictions on a class). An in-depth analysis yielded the following results:

2.1. Object vs. ObjectAggregate

BFO's alleged assumption that there is a clear ontological division between Object or ObjectAggregate is already challenged by the fact that any self-connected physical object can be also be described as a mereological sum of molecules, atoms, or elementary particles. A crucial point to stress here is self-connectedness: According to BFO, this property means that all the constituent parts are not separated from each other by spatial gaps. Albeit this is fully consistent with the restriction on ObjectAggregate, *viz.* that they are mereological sums of separate objects, the self-connectedness property requires a theory of connectedness in practice. For the ontology engineer there are two possibilities out of the dilemma:

Firstly, one could pursue an intuitive approach. On a macroscopic level spatial gaps can (possibly) be identified with the naked eye or by simple physical manipulation of the involved objects. But if shifting to a microscopic level and in case molecular entities should be represented, the situation might be quite different. Then, two cells could well appear disconnected under a light microscope but tiny interweaving molecular bridges could be revealed by a stronger electron microscope. As a consequence, modeling decisions are granularity-dependent.

Alternatively, one could resort to chemistry and the theory of the chemical bond in order to find a scientific criterion for connectedness or disconnectedness. Here continuous transitions [12] exist not only between the different kinds of bonds (i.e., ionic, covalent, metallic) but also in terms of strength of a bond [8]. For instance, large numbers of hydrogen bonds (which are in a strict sense not considered as molecule-forming chemical bonds) can form stable molecular complexes such as antigen/antibody aggregations. But on the other hand, the same kind of chemical bond can sporadically appear between separated biological membranes without any functional implications. Here, the case of liquids may be seen as borderline where the single molecules are not stationary but nevertheless cohesive due to connecting forces (e.g., water dipoles).

In BioTop we interpreted the connectedness requirement in a strict way only where no physical connecting forces can be assumed at all (e.g., in the instances of the classes PortionOfGas or PopulationOfHumans). This however entailed the necessity to create highly general BioTop classes (e.g., Population) as children of the union class Object OR ObjectAggregate, since there are populations of organisms in which the individuals are connected, such as in many aggregations of unicellular organisms (e.g., bacteria or yeast). But since Object and ObjectAggregate are disjoint, an instance of this union can therefore actually belong to either one of the classes but not to both at the same time.

2.2. *Object vs. FiatObjectPart*

With this distinction BFO postulates that everything can be classified according to the existence of a physical discontinuity (see again the definitions in Table 1), which is debatable, at least for biological entities. Beyond any doubt, many biologically relevant entities fulfill the strict criteria for Object, such as cells in the blood, molecules in a gas, or independent organisms. But many other classes cannot be unambiguously attributed to be either Object or FiatObjectPart. An emblematic example for this is a Siamese twin: Whereas the conjoined double-organism fulfills the criteria for Object, either twin is a FiatObjectPart because there is no clear-cut discontinuity between them. But after (surgical) separation, however, each twin becomes an independent Object. This discussion can be extended for many symbiotic organisms, to telophase cells in the mitotic cycle, or to atoms in some molecular associations. Finally, in the realm of gross anatomy, the conception of bodily organs kinds as Object would require to neglect their transitory structures (e.g., ducts, vessels, nerves) connecting them with their neighboring structures without any clear discontinuities.

In BioTop the emerging inconsistencies of classes subsumed both by Object and FiatObjectPart could only be fixed by introducing yet another union of BFO classes, *viz.* Object OR FiatObjectPart. Most important biological classes are subsumed by this union class (e.g., Cell, Organism, or Atom). But as with the union mentioned in the preceding section, the disjointness between Object and FiatObjectPart makes it impossible to declare e.g., the same given instance of a Siamese twin as a proper object or a fiat object part using the same ontology to just express different points of view.

3. Conclusion

Biomedical ontologies address different levels of granularity, reaching from populations down to single atoms. Neutrality towards granularity issues was a major architectural principle when creating BioTop, a top-domain ontology for the linking and interfacing of domain ontologies in the biomedical and biomolecular sciences.

The alignment of the BioTop classes with the Basic Formal Ontology (BFO) re-introduces some problems concerning granularity, since the BFO definitions that address aggregation and connection are obviously committed to a mesoscopic viewpoint and can therefore not be generalized to the whole range of granularity.

We here defend the opinion that upper ontologies should be neutral to granularity because addressing granularity issues in an upper ontology is like opening Pandora's Box: On the one hand this would presuppose a consensus on how to divide the granularity continuum, which is hard if not impossible to achieve. On the other hand, each level of granularity would then require its own set of axioms that could possibly clash with axioms from another granularity level. Also, to divide such cross-granular ontologies into separate ontologies tailored for granularity levels is likewise unfeasible since the inherently cross-granular nature of the biomedical domain. Furthermore, as the Siamese twin example demonstrates, the prominence of the notion of fiat boundaries complicates the classification of entities even at the same granularity level.

The conceptualization of connectedness is still an intricate philosophical problem [16]. Certainly, such problems may be addressed in some cases where a restricted domain description is committed to as well defined level of granularity. In such cases, the subdivisions of IndependentContinuant could indeed be quite useful. But for

BioTop, as a deliberately granularity independent ontology, both distinctions between Object vs. ObjectAggregate and Object vs. FiatObjectPart are rather inexpedient.

In accordance with several postings to the BFO discussion group[5], where our report on the described problems was starting point of a lively (and ongoing) discussion, a tentative solution is emerging as follows:

1. For an extension of BFO in terms of a comprehensive domain top level (such as plugging BioTop to BFO) only the first two BFO levels are used.
2. In contrast, the mereotopological subclasses of IndependentContinuant should only be used once certain parameters (i.e., time, granularities) are fixed for a restricted subdomain.

Acknowledgments:

We thank all participants of the BFO Discussion Group for the lively exchange of ideas related to the discussed topics. Our work has been supported by the European Commission project BOOTStrep[6] (FP6 - 028099).

References

[1] A. Gangemi, N. Guarino, C. Masolo, and A. Oltramari. Sweetening Ontologies with DOLCE. In Proc. of the Internat.. Conference on Knowledge Engineering and Knowledge Management (EKAW), 2002.
[2] P. Grenon, B. Smith, and L. Goldberg. Biodynamic Ontology: Applying BFO in the Biomedical Domain. In Proceedings of the Workshop on Medical Ontologies, pages 20-38, 2004.
[3] B. Heller and H. Herre. Ontological Categories in GOL. Axiomathes, 14(1):57-76, 2004.
[4] R. Hohndorf. GFO-Bio: A Biomedical Core Ontology, http://onto.eva.mpg.de/gfo-bio.html, last accessed: October, 15th, 2007.
[5] H. Knublauch, O. Dameron, and M. Musen. Weaving the Biomedical Semantic Web with the Protege OWL Plugin. In U. Hahn, S. Schulz, and R. Cornet, editors, Proceedings of the 1st International Workshop on Formal Biomedical Knowledge Representation (KR-MED 2004), pages 39-47, 2004.
[6] A. McCray, Y. Yang, and R. Patil. The UMLS Semantic Network. In Proc. of the 13th Annual Symposium on Computer Applications in Medical Care, pages 503-507, 1989.
[7] L. Obrst, P. Cassidy, S. Ray, B. Smith, D. Soergel, M. West, and P. Yim. The 2006 Upper Ontology Summit Joint Communique. Applied Ontology, 1(2):203-211, 2006.
[8] L. Pauling, editor. The Nature of the Chemical Bond. Cornell University Press, 1960.
[9] A. Rector, J. Rogers, and P. Pole. The GALEN High Level Ontology. In Proc. of the 14th International Congress of the European Federation for Medical Informatics, pages 174—178, 1996.
[10] A. Rector, R. Stephens, and J. Rogers. Simple Bio Upper Ontology, http://www.cs.man.ac.uk/~rector/ontologies/sample-top-bio, last accessed: October, 15th, 2007.
[11] S. Schulz and U. Hahn. Towards the Ontological Foundations of Symbolic Biological Theories. Artificial Intelligence in Medicine, 39(3):237-250, 2007.
[12] S. Schulz and I. Johansson. Continua in Biological Systems. The Monist, 90(4), 2007.
[13] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, and Y Katz. Pellet: A Practical OWL-DL Reasoner. Journal of Web Semantics, 5(2), 2007.
[14] B. Smith and A. Varzi. Fiat and Bona Fide Boundaries. Philosophy and Phenomenological Research, 60(2):401-420, 2000.
[15] H. Stenzhorn, E. Beisswanger, and S. Schulz. Towards a Top-Domain Ontology for Linking Biomedical Ontologies. in Proc. of the 12th World Congress on Medical Informatics, pages 1225-1229, 2007.
[16] A. Varzi. Mereological Commitments. Dialectica, 54:283-305, 2000.

---

[5] BFO Discussion Group website: http://www.ifomis.org/bfo/discussion
[6] BOOTStrep website: http://www.bootstrep.eu