The Role of Foundational Relations in the Alignment of Biomedical Ontologies

Barry Smith[a], Cornelius Rosse[b]

[a]*IFOMIS, University of Leipzig, Germany and Department of Philosophy, University at Buffalo, NY, USA*

[b]*Structural Informatics Group, Department of Biological Structure, University of Washington, Seattle, WA, USA*

Abstract

The Foundational Model of Anatomy (FMA) symbolically represents the structural organization of the human body from the macromolecular to the macroscopic levels, with the goal of providing a robust and consistent scheme for classifying anatomical entities that is designed to serve as a reference ontology in biomedical informatics. Here we articulate the need for formally clarifying the is-a *and* part-of *relations in the FMA and similar ontology and terminology systems. We diagnose certain characteristic errors in the treatment of these relations and show how these errors can be avoided through adoption of the formalism we describe. We then illustrate how a consistently applied formal treatment of taxonomy and partonomy can support the alignment of ontologies.*

Keywords:

Bioinformatics, Ontology, Logic, Mereology, Taxonomy, Gene Ontology, Foundational Model of Anatomy, UMLS.

Introduction

The growth of bioinformatics has led to an increasing number of evolving ontologies which must be correlated with the existing terminology systems developed for clinical medicine. A critical requirement for such correlations is the alignment of the fundamental ontological relations used in such systems, above all the relations of class subsumption (*is_a*) and partonomic inclusion (*part_of*). To achieve this end, however, existing clinical and evolving bioinformatics terminologies need to call upon formalisms whose significance was not evident at the time these resources were originally conceived.

Both *is_a* and *part_of* are ubiquitous in bioinformatics ontologies and terminologies. Yet their treatment is inconsistent and problematic, and in some cases the two relations are not clearly distinguished at all. SNOMED-RT, for example, has: *both testes is_a testis*. UMLS has: *plant leaves is_a plant*.

In this communication we argue that a coherent treatment of *is_a* and *part_of* must be based on explicit formal definitions which take into account not only the classes involved as terms of these relations but also the instances of these classes. We base our arguments on the lessons we learned during the evolution of the Digital Anatomist Foundational Model of Anatomy (FMA, for short [1]) in which we have refined the treatment of these relations over time and distinguished between classes and instances in terms of canonical and instantiated anatomy. [2]

Our objectives are to define canonical and instantiated anatomy before giving formal definitions of *is_a* and *part_of* in terms of a theory of instantiation. We then discuss in this light issues of universal relevance to ontologies, such as classes vs. wholes and sets, granularity, idealization, and the roles of time and change. After illustrating problematic usage of *is_a* and *part_of* we draw conclusions for ontology alignment, pointing to the need for supplementing Description Logic-based reasoning implementations with rigorous manual auditing of underlying data-sources based on formal analyses in terms of instance-level relations and on clear and intuitive principles of curation.

Canonical and Instantiated Anatomy

Canonical anatomy is a field of anatomy (science) that comprises the synthesis of generalizations based on anatomical observations that describe idealized anatomy (structure). These generalizations have been implicitly sanctioned by their usage in anatomical discourse. *Instantiated anatomy* is a field of anatomy (science) that comprises anatomical data pertaining to individual instances of organisms and their parts. Instantiated anatomy is needed to support the application of biomedical knowledge in clinical care and in fields such as image analysis. The corresponding instance-data is not incorporated into the FMA, which deals with idealizations at a higher level of abstraction. In introducing the relation between canonical and instantiated anatomy, however, the FMA provides the key to an adequate formal treatment of *is_a* and *part_of*, for the latter can be defined and formally interrelated only when the relation of *instantiation* between individuals and classes is taken into account.

Formal Theory of *Is_a* and *Part_of*

We use the term *entity* as a universal ontological term of art embracing objects, processes, functions, structures, times and places, and we distinguish among entities in general two special subtotalities, called *instances* and *classes*, respectively.

Instances are individuals (particulars, tokens) of special sorts. Thus each is a simply located entity, bound to a specific (normally topologically connected) location in space and time. [3] Classes (also called universals, kinds, types) are multiply located; they exist in their respective instances.

To formalize these notions we use standard first-order logic with variables x, y, x_1, etc. ranging over instances, and A, B, A_1, etc. ranging over classes. Our system rests on two primitive relations of *inst* and *part*.

Inst is the relation of instantiation between an instance and a class, illustrated by: Jane is an instance of human being. We define a class as anything that is instantiated; an instance as anything (any individual) that instantiates some class. The principal axioms governing *inst* are: (1) that it holds in every case between an instance and a class, in that order; and (2) that nothing can be both an instance and a class. [4]

Part is the relation of parthood among individuals, illustrated by: Jane's heart is part of Jane's body. The axioms governing *part* (also called 'proper part') can be specified as follows [5]. It is (1) *irreflexive* (no entity is part of itself), (2) *asymmetric* (if *part*(x, y) then not-*part*(y, x)), and (3) *transitive* (if *part*(x, y) and *part*(y, z), then *part*(x, z)). In addition, it satisfies: (4) principles governing the formation of sums of parts (for example that binary sums $x+y$ exist), and (5) a remainder axiom, to the effect that if *part*(x, y) then there is some part z of y which does not share parts in common with x.

We use the standard quantifiers of first-order logic: $\exists$, abbreviating *for some value of*, and $\forall$, abbreviating *for all values of*. The device of quantification allows us to take account of instantiation in generic fashion, i.e. without the need to take specific instances into account. The full formalism requires general axioms specifying the properties of classes as natural kinds (rather than arbitrary collections) [6,7], together with more specific axioms dealing with the different sorts of classes (of objects, functions, processes, pathways, sites, etc.) in the different domains of biomedical ontology. It also requires an axiom of extensionality, to the effect that classes which share identical instances are themselves identical.

We can now define *is_a*, the relation of class subsumption:

D1 A *is_a* B $=_{\text{def}}$ $\forall x$ (*inst*(x, A) $\rightarrow$ *inst*(x, B))

where '$\rightarrow$' abbreviates: *if ... then ...*. To say that A *is_a* B is to say that every instance of A is an instance of B.

To define *part_of* we start, following [8], by defining:

D2 A *part_for* B $=_{\text{def}}$

$\forall x$ (*inst*(x, A) $\rightarrow$ $\exists y$ (*inst*(y, B) & *part*(x, y))).

D2 provides information primarily about As; it tells us that As do not exist except as instance-level parts of Bs. Conversely:

D3 B *has_part* A $=_{\text{def}}$

$\forall y$ (*inst*(y, B) $\rightarrow$ $\exists x$ (*inst*(x, A) & *part*(x, y)))

provides information primarily about Bs; it tells us that Bs do not exist except with As as instance-level parts.

Because there are female as well as male human beings, we can state: *human testis part_for human being*, but we cannot state: *human being has_part human testis*. Because non-human vertebrates also have hearts, we can state *human being has_part heart*, but not: *heart part_for human being*.

We now define the relation *part_of* by combining D2 and D3:

D4 A *part_of* B $=_{\text{def}}$ A *part_for* B & B *has_part* A

Thus A *part_of* B if and only if: for any instance x of A there is some instance y of B which is such that x stands to y in the instance-level part relation, *and vice versa*. This implies a strong structural tie between the classes A and B. It guarantees that As exist only as parts of Bs and that Bs are structurally organized in such a way that As must appear in them as parts.

We can, given the above, prove that *is_a* is reflexive (for every class A we have: A *is_a* A), and antisymmetric (if A *is_a* B and B *is_a* A, then A and B are identical).

We need to add as axiom that *part_of* is irreflexive (that no class is *part_of* itself). From this we can prove that *part_of* is also asymmetric (if A *part_of* B then not-B *part_of* A).

We can prove also that both *is_a* and *part_of* are transitive: thus if A *is_a* B and B *is_a* C, then A *is_a* C, and if A *part_of* B and B *part_of* C then A *part_of* C.

That partonomies like those associated with the FMA are structured by the full *part_of* relation defined in D4 is ensured by the fact that here all terms for body parts are assumed to have an implicit prefix designating the type of organism involved.

Sometimes we need to capture mereological relations involving specific numbers of instances. Thus in a case like *human being has_part brain*, we need to express that each instance of *human being* has exactly one instance of *brain* as part:

$$\text{\textit{inst}}(x, human) \rightarrow \exists y \forall z((\text{\textit{inst}}(z, brain) \,\&\, \text{\textbf{part}}(z, x)) \leftrightarrow z = y)$$

with generalizations to represent a human being's canonical organization as having two lungs, ten fingers, and so on.

Both *is_a* and *part_of* are standardly treated as relations between classes. The formal structure of D4 makes it clear, however, that the latter does not signify that classes stand in some special class-level mereological inclusion relation. Rather, it expresses the more fundamental part-relations between the underlying instances captured in D2 and D3.

A distinction analogous to that between D2, D3 and D4 is indispensable to the formal definition of many other foundational relations of biomedical ontologies – including the majority of the 54 relations contained in the UMLS Semantic Network (UMLS-SN, Version 2003AB) [9]. In particular, reference to instances is a necessary first step in the rigorous implementation, in system like the FMA, of mereotopological relations such as spatial occupation and spatial adjacency, as also of concepts such as *junction*, *boundary*, *cluster*, and the like. [1,10,11]

Classes vs. Wholes: Granularity and Idealization

A rigorous system of formal definitions to support biomedical ontology alignment must clarify also the relations between the concepts of *class*, *mereological whole* and *set*. Here, too, the reference to instances is indispensable. For classes are distinguished by the fact that they capture their instances in a way which involves the factor of granularity, which means: in such a way as to divide up the corresponding domain into whole units or members, whose interior parts and structure are traced over. [12] A mereological sum is not granular in this sense. The mereological sum of human beings comprehends also all instance-

level parts (including organs, cells, and so on, all the way down to sub-molecular parts). The *class* of human beings is instantiated only by human beings as single, whole units.

Which classes exist in a given domain is a matter for empirical research. Hence a good first clue to the existence of a class is provided by the fact that there exists a corresponding term that has either been sanctioned by (in our case anatomical) science or can be inferred from terms so sanctioned by the need to fill gaps in the taxonomy or partonomy (for example terms for higher-level classes and for not previously named classes instantiated by macroscopic parts of the body). [13] In anatomy and related disciplines a supplementary clue may be provided through the association of given classes with the structural genes whose coordinated expression gives rise to the corresponding instances.

The instances (units, members) in a class are marked by the fact that, in the Aristotelian terms used by the FMA, they share a common essence. [14,15] Each class-definition in the FMA specifies the essence shared by the corresponding instances via the specification of (i) a *genus*, which is some wider class to which the given class belongs, together with (ii) the *differentiae* which mark out its instances within this wider class.

Biological classes are marked always by an opposition between standard or prototypical instances and a surrounding penumbra of non-standard instances (not all instances of the class *human being* are marked by the presence of amputation stumps or pituitary tumors). To do justice to these matters FMA introduces the factor of *idealization*, which means (in first approximation) that the classes of the FMA's Anatomy Taxonomy **AT** include only those instances to which canonical anatomy applies.

This means that we need to revise definitions D1–D4 by restricting the range of variables x, y, ... to the realm of individuals which satisfy the generalizations of canonical anatomy, so that the same abstraction of anatomy (structure) will be represented in all the instances of any given **AT**-class. This device of specifying different ranges of variables gives us the means also to represent the generalizations belonging to the different branches of canonical anatomy, for example to canonical anatomy for male vs. female human beings, for human beings at various developmental stages, and for organisms in other species. It can allow us also to represent the generalizations governing the anatomical variants yielded by the presence of, for example, coronary arteries or bronchopulmonary segments, which deviate from canonical anatomical patterns of organization.

Classes vs. Sets: Granularity and Time

Sets in the mathematical sense, too, are marked by the factor of granularity, which means that each set comprehends its members as single, whole units. A class or set is laid across reality like a grid consisting (1) of a number of slots or pigeonholes each (2) occupied by some member. (This informal talk of grids and slots is formalized in [16] in terms of a theory of 'granular parti-

tions.') Classes are distinguished from sets, however, by the fact that a set is *determined by its members*. This means that it is (1) associated with a specific number of slots, each of which (2) must be occupied by some specific member. A set is thus specified in a double sense. A class, in contrast, survives the turnover in its instances, and so it is specified in neither of these senses, since both (1) the number of associated slots and (2) the individuals occupying these slots *may vary with time*.

Sets are distinguished from classes also in this: a set with n members has in every case exactly 2^n subsets, constituted by all the combinations of these members. The subclasses of a class, on the other hand, are limited in number, and which classes are subsumed by a larger class is a matter for empirical science to determine. Leaves (lowest nodes) in the taxonomy are (changing) collections of instances. As we move up the taxonomy we encounter in succession collections of such collections of instances, collections of collections of such collections, etc., organized in a nested hierarchy reaching up to the maximal class or 'root'. We can visualize the classes at different levels as being analogous to geopolitical entities (towns, counties, states) as represented on a map. Instances correspond in this analogy to the corresponding populations: a class is not determined by its instances as a state is not determined by its citizens.

Classes are distinguished from sets also by their relation to time. A set is an abstract structure, existing outside time and space, and this is so even when its members are parts of concrete reality. Since each set is determined by its members, the set of human beings existing at t is (timelessly) a different entity from the set of human beings existing at t' because of births and deaths.

Matters are different with regard to classes. The class *human being* can survive the change in the stock of its instances which occurs when John and Jane die, because classes *exist in time*. John and Jane themselves can similarly survive changes in the stock of cells or molecules by which they are constituted.

To do justice to the fact that classes in the biological domain endure even when their extensions change, a full definition of the *is_a* relation must involve a temporally indexed reading of ***inst*** (with variables t, t', etc., ranging over times):

D1* A *is_a* $B =_{\text{def}} \forall t \, \forall x \, (\, inst(x, A, t) \rightarrow inst(x, B, t) \,)$,

so that A *is_a* B means: at all times t, if x is an instance of A at t then x is an instance of B at t. D1* will also take care of false positives such as *adult is_a child*, which an untensed reading of D1 would otherwise allow. In general, all statements of *inst* and *part* relations involving objects in biomedical ontologies, like all the data of instantiated anatomy, are indexed by times.

Taxonomy and Partonomy

A taxonomy such as **AT** is formally speaking a tree in the mathematical sense. It satisfies axioms to the effect that (1) it has a root or unique maximal genus (here: *anatomical entity*) and (2) all other classes are connected to this root via finite chains of *is_a* relations satisfying a principle of single inheritance. A partonomy, in contrast, is a partial order in the mathematical sense, with top (here: *organism* – the class instantiated by mereologically maximal entities), to which all other classes are connected via chains of *part_of* relations.

We can define the concepts of root and leaf of a taxonomy and top and bottom of a partonomy as follows.

D5 *root* $(A) =_{\text{def}} \forall B$ (B *is_a* A)

D6 *leaf* $(A) =_{\text{def}} \forall B$ (B *is_a* $A \rightarrow A = B$)

D7 *top*$(A) =_{\text{def}}$

$\forall B$ ($A = B$ or B *part_of* A) & not-$\exists B$ (A *part_of* B)

D8 *bottom*$(A) =_{\text{def}}$ not-$\exists B$ (B *part_of* A).

We can then postulate axioms to the effect that every class includes some leaf as subclass, and that every instance of every class instantiates some leaf:

$\forall A \exists B$ (*leaf* (B) & B *is_a* A)

$\forall A \forall x$ (*inst* $(x, A) \rightarrow \exists B$ (*leaf* (B) & *inst* (x, B)))

The taxonomical union $A \cup B$ of classes A and B is defined as the minimal class satisfying the condition that it contains both A and B as subclasses. Such a class always exists, since A and B are in any case subclasses of the root. The taxonomic union of *femur* and *liver*, for example, is *organ.* The partonomic union (or 'sum') of two classes $A+B$ is the class, if it exists, whose instances are sums $x+y$ of instances of classes A and B respectively. While every pair of classes has a taxonomic union, only some classes have a partonomic union, since entities of the form $x+y$ are instances of classes only in some highly restricted cases, for example: *left lung = upper-lobe-of-left-lung + lower-lobe-of-left-lung.* Such examples characteristically involve the phenomenon of fiat boundaries. [17,18]

As concerns taxonomic intersection, a class is never immediately subordinated to more than one higher class within a tree. This means that if two classes overlap in sharing some common subclass, then this is because one is a subclass of the other. $A \cap B$, the taxonomic intersection of A and B, if it exists, is then simply the smaller of these two classes. We can add further an axiom to the effect that, if two classes are such as to overlap in sharing some common instances, then this, too, is because one is a subclass of the other:

$$\exists x \; (\textbf{\textit{inst}}(x, A) \wedge \textbf{\textit{inst}}(x, B)) \rightarrow A \; \textit{is\_a} \; B \; \text{or} \; B \; \textit{is\_a} \; A.$$

Classes can overlap *partonomically*, on the other hand, in such a way that there is a class which stands in the *part_of* relation to both, though neither stands in this relation to the other:

D9 A_1 *partonomic_overlap* $A_2 =_{\text{def}}$

$\exists A$ (A *part_of* A_1 & A *part_of* A_2).

For example: *pelvis* and *vertebral column* overlap in the *sacrum* and *coccyx.* Most classes in the biomedical domain do not overlap partonomically in this sense, yet it is this difference in behavior between taxonomic and partonomic overlap which captures the essential difference between the tree structure of taxonomies and the partial order structure of partonomies.

Problematic Representations of *Is_a* and *Part_of*

To see what such clarifications are good for, we consider the example of the Gene Ontology™ (GO) [19]. GO, too, is centered around the two foundational relations 'is a' and 'part of'. The former, also sometimes called 'is a subtype of', it uses in a way which is broadly in accord with our definition D1 above. [20] The relation 'part of', in contrast, is used by GO in at least three ways. First, in assertions like 'Cell Component part of Gene Ontology', to represent inclusion relations between vocabularies. Second, to represent a time-dependent mereological relation between classes that is specified only informally:

"*part of* means *can be a part of*, **not** *is always a part of*: the parent need not always encompass the child. For example, in the component ontology, **replication fork** is a part of the **nucleoplasm**; however, it is only a part of the nucleoplasm at particular times during the cell cycle" [19].

Third, in examples such as *flagellum part of cell*, to stand for a variant of our *part_for* defined in D2 above.

GO's first usage represents a simple inclusion relation between lists. Its second usage would seem to correspond formally to:

D10 A *sometimes_part_of*$_{\text{GO}}$ $B =_{\text{def}} \exists t \; \exists x \; \exists y$

(*inst*(x, A, t) & *inst*(y, B, t) & *part*(x, y, t)).

A is part of B, on this reading, if there is some time at which an instance of A stands in the **part** relation to an instance of B.

Its third usage suggests a definition along the lines of:

D11 A *part_of*$_{\text{GO}}$ $B =_{\text{def}} \exists C$ (C *is_a* B & A *part_of* C),

according to which A is part of B means: there is some subclass C of B for which A *part_of* C.

Against this background GO's Usage Guide [19] lists four 'logical relationships' between its 'is a' and 'part of' relations, which can be summarized in our terms as follows:

(1)   (A *part_of*$_{\text{GO}}$ B & C *is_a* B) $\rightarrow A$ *part_of*$_{\text{GO}}$ C

(2)   *is_a* is transitive

(3)   *part_of*$_{\text{GO}}$ is transitive

(4)   (A *is_a* B & C *part_of*$_{\text{GO}}$ A) $\rightarrow C$ *part_of*$_{\text{GO}}$ B.

(1) – (3) are accepted by GO; (4), on the other hand, is rejected. Yet there are many cases for which (1) fails. For example:

hydrogenosome part_of$_{\text{GO}}$ *cytoplasm*

sarcoplasm is_a cytoplasm

But not: *hydrogenosome part_of*$_{\text{GO}}$ *sarcoplasm.*

Some of GO's curators accordingly advocate removing (1) from the GO Usage Guide.

As concerns (3), consider:

plastid part_of$_{\text{GO}}$ *cytoplasm*

cytoplasm part_of$_{\text{GO}}$ *cell (sensu Animalia)*

But not: *plastid part_of*$_{\text{GO}}$ *cell (sensu Animalia).*

While 'cell (sensu Animalia)' is not a term in GO, it does conform to GO's rules for term formation, and this suggests reason for some uncertainty also as to the validity of (3).

GO justifies its rejection of (4) with the following example:

meiotic chromosome is_a chromosome

synaptonemal complex part_of$_{\text{GO}}$ *meiotic chromosome*

But not necessarily:

synaptonemal complex part_of$_{GO}$ chromosome.

On the reading of GO's 'part of' as meaning 'can be part of', however, it seems that *synaptonemal complex* **is** 'part of' *chromosome*. And if the reading of GO's 'part of' given in D11 is correct, then (4) can indeed be proved as a matter of logic.

We suggest that it is only by appeal to formal definitions that these and related uncertainties can be resolved. [21] Such definitions would help also to ensure that, when the terms of controlled vocabularies like GO are mapped into other systems, then this can be done in ways that support the drawing of reliable inferences across ontologies.

Conclusion

Practitioners in the biomedical sciences move easily between the realm of classes and the realm of instances existing in time and space. For historical reasons, however, work on biomedical ontologies and terminologies – which grew out of work on medical dictionaries and nomenclatures – has focused almost exclusively on classes (or 'concepts') atemporally conceived. This class-orientation is common in work on knowledge representation, and its predominance has led to the entrenchment of an assumption according to which all that need be said about classes can be said without appeal to formal features of instantiation of the sorts described above. This, however, has fostered an impoverished regime of definitions, in which the use of identical terms in different systems has been allowed to mask underlying incompatibilities. Matters have not been helped by the fact that work on Description Logics, the prevalent framework for terminology-based reasoning systems, was for a long time oriented primarily around reasoning with classes.

Certainly if we are to produce information systems with the requisite computational properties, then this entails recourse to a logical framework like that of description logic. At the same time, however, we must ensure that the data that serves as input to such systems is organized formally in a way that sustains rather than hinders successful alignment with other systems. The way forward is to recognize, as does the FMA, that these are two distinct tasks, both of which are equally important to the construction of biomedical ontologies and terminologies.

Acknowledgements

References

[1] Rosse C, Mejino JLV. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform*. 2003;36:478-500.

[2] Rosse C, Mejino JL, Modayur BR, Jakobovits R, Hinshaw KP, Brinkley JF. Motivation and organizational principles for anatomical knowledge representation: the Digital Anatomist Symbolic Knowledge Base. *J Am Med Inform Assoc* 1998; 5: 17-40.

[3] Grenon P, Smith B. SNAP and SPAN: Towards dynamic spatial ontology. Spatial Cognition and Comput. In press.

[4] Smith B. The logic of biological classification. D Westerstahl (ed.): *Proc 10th Int Conf Logic Methodology Philosophy of Science*, Oviedo, Spain, 2003 (in press).

[5] Simons PM. Parts. *An essay in ontology.* Oxford: Clarendon Press, 1987.

[6] Cocchiarella N. On the logic of natural kinds. *Philosophy of Science*, 1976; 43: 202-222.

[7] Buck RC, Hull DL. The logical structure of the Linnaean hierarchy. *Sys Zoology*, 1966; 15: 97-111.

[8] Schulz S, Hahn U. Necessary parts and wholes in bio-ontologies. *Proc KR 2002*; 387-394.

[9] http://www.nlm.nih.gov/research/umls/META3.HTML.

[10] Smith B. Mereotopology: a theory of parts and boundaries. *Data and Knowledge Engineerin*g, 1996;20:287-303.

[11] Smith B, Varzi AC. Surrounding space: the ontology of organism-environment relations, *Theory in Biosciences*, 2002; 121: 139-162.

[12] Smith B, Brogaard B. Quantum mereotopology. *Ann Math Artificial Intelligence* 2002; 35: 153-175.

[13] Agoncillo A, Mejino JLV, Rosse C. Influence of Digital Anatomist Foundational model on traditional representations of anatomical concepts. *Proc AMIA Symp* 1999;2-6.

[14] Aristotle. *The categories.* Cambridge, Mass: Harvard University Press, 1973.

[15] Michael J, Mejino JLV, Rosse C. The role of definitions in biomedical concept representation. *Proc AMIA Symp 2001*; 463-467.

[16] Bittner T, Smith B. *A theory of granular partitions. Foundations of Geographic Information Science,* M Duckham et al. (eds.). London: Taylor & Francis, 2003: 117-151.

[17] Smith B. Fiat objects. *Topoi*, 2001; 20: 131-148.

[18] Smith B, Varzi AC. Fiat and bona fide boundaries. *Phil Phenomenol Res,* 2000; 60: 401-420.

[19] http://www.geneontology.org (Version as at 1.9.2003)

[20] Smith B, Köhler J, Kumar A. On application of formal principles to life science data. *Proc DILS 2*004 (in press).

[21] Smith B, Williams J, Schulze-Kremer S. The ontology of the Gene Ontology. *Proc AMIA Symp* 2003, 609–613.

Address for Correspondence

Barry Smith, Department of Philosophy, University at Buffalo, NY 14260, USA. phismith@buffalo.edu, http://ontology.buffalo.edu.