Realism for scientific ontologies

Michel Dumontier [a,1] and Robert Hoehndorf [b,c]
[a] *Department of Biology, Institute of Biochemistry and School of Computer Science, Carleton University*
[b] *European Bioinformatics Institute*
[c] *Institute for Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig*

Abstract. Science aims to develop an accurate understanding of reality through a variety of rigorously empirical and formal methods. Ontologies are used to formalize the meaning of terms within a domain of discourse. The Basic Formal Ontology (BFO) is an ontology of particular importance in the biomedical domains, where it provides the top-level for numerous ontologies, including those admitted as part of the OBO Foundry collection. The BFO requires that all classes in an ontology are actually instantiated in reality. Despite the fact that it is hard to show whether entities of some kind exist or do not exist in reality (especially for unobservable entities like elementary particles), this criterion fails to satisfy the need of scientists to communicate their findings and theories unambiguously. We discuss the problems that arise due to the BFO's realism criterion and suggest viable alternatives.

Keywords. Realism, biomedical ontology, philosophy of science

1. Introduction

Science aims to better understand reality through a variety of rigorously empirical (observation and experimentation) and formal methods (mathematics, statistics and logic). Reality, however, is hard to investigate due to intrinsic stochasticity in addition to unknown, inaccessible, and multiple levels of complexity. To reduce this complexity, contemporary scientists develop tractable experimental or computational models of reality that provide a slightly more controlled environment in which experiments may be carried. Indeed, scientific models of reality provide great insight into otherwise complex real world entities, and more importantly, may predict the existence of yet unobserved entities with detailed descriptions of their innate structure, qualities and roles and functions.

Increasingly in the last decade, scientists have engaged in the building of ontologies for the specific purpose of knowledge management and knowledge discovery – that through common understanding of terms, scientific data and findings could be consistently annotated thus promoting data integration and exchange across heterogeneous representations, and support advanced methods for symbolic and numeric data mining. The Basic Formal Ontology (BFO) [16] is one of several foundational ontologies that is guiding the development of dozens of biomedical domain ontologies. Formal ontology aims

[1] Corresponding Author: Michel Dumontier, michel_dumontier@carleton.ca.

to provide a specifcation of the meaning of terms within a vocabulary [18]. Ontologies that use the BFO as its upper level ontology may be considered for inclusion into the OBO Foundry [44], provided that the ontology is orthogonal to other candidate ontologies. A key criterion for BFO-compliant ontologies is that they employ a particular form of realism, termed here *BFO-realism*, which, among other things, insists that classes in the ontologies must have instances in reality.

Leaving philosophical debate aside, we must ask ourselves whether the criteria for realist ontologies used in the OBO Foundry lead to ontologies that are useful for scientific applications and for scientific communication. In particular, we believe that scientific hypotheses, categories of objects which cannot (yet) be shown to exist, and specifications of categories of objects are used in communications amongst scientists and should be represented and formally described in ontologies. Ontologies that are exclusively built upon classes that have instances in reality face the problem of representing these issues and fail to support interoperability.

Here, we discuss the kind of realism employed by the BFO and by the OBO Foundry ontologies with regard to its adequacy for scientific research, and propose an alternative to solve some of the problems we identify.

2. Realism

In its simplest form, realism corresponds to the belief that "there is a real world existing independently of us" [41], a world constituted of objective facts. These facts exist independently of our minds and independently of the representations we create of them, they exist completely independent of us.

Realism is not identical with or implies a theory of truth. In particular, realism is not identical to the correspondence theory of truth. Realism is not an epistemic theory, nor does realism prefer a form of representing reality. "Realism does not say how things are but only that there is a way that they are" [41].

Realism can be contrasted with idealism, which is the philosophical view that the nature of reality depends on human minds and perception. Idealism is often derived from the observation that the only epistemic basis we have for claims about the external world are our perceptual experiences, or that "[a]ny cognitive state occurs as part of a set of cognitive states and within a cognitive system" [41]. Various arguments for idealism in the context of formal ontology have already been discussed and rejected [41,43], and we will not add to this discussion here. Instead, we will focus the discussion on realism about universals and its role in the construction of formal ontologies.

Depending on what kind of object is assumed to exist, different kinds of realism can be distinguished. Realism about universals exerts the belief that universals (entities that may be instantiated in many things) exist. A universal is an entity which "is of such a nature as to be predicated of many subjects" [3]. They are the things to which the general terms in utterances correspond. Whether universals *exist* is an old problem in philosophy [27].

Two major forms of realism about universals can be distinguished, Platonic and Aristotelian realism. In Platonic realism, universals exist in a realm of their own, are not tied to space and time, and only some of these universals are actually instantiated in our world.

According to Aristotelian realism, universals exist *in re*, i.e., in their instances. Therefore, according to Aristotelian realism, universals are in space and time. For example, the universal *Chair* (or chair-ness) is present in all chairs which existed in the past, exist now and will exist in the future. In Aristotelian realism, universals and their instances share a symbiotic relationship; one cannot exist without the other. In particular, there are no universals which have no instance in reality.

For example, *Unicorn* is not a universal in the Aristotelian sense, because it has no instances in reality (to the best of our knowledge, because only depictions and descriptions of the unicorn can be found, and the unicorn is understood to be a creature of fantasy). As Aristotelian universals exist in their instances, and no instances exist, the universal *Unicorn* does not exist.

In its defense of realism against idealism and other anti-realist philosophies, Smith [43] proposes the principle to use Aristotelian universals as a basis for building ontologies:

> Bad ontologies are (*inter alia*) those whose general terms lack the relation to corresponding universals in reality, and thereby also to corresponding instances.
> Good ontologies are reality representations, and the fact that such representations are possible is shown by the fact that, as is documented in our scientific textbooks, very many of them have already been achieved, though of course always only at some specific level of granularity and to some specific degree of precision, detail and completeness.

In physics, the Higgs boson [21] is a fundamental particle predicted by the Standard Model of physics [35], and since no instances are known, *Higgs boson* may not be an Aristotelian universal. In chemistry, molecules that are not known to exist, but are described through chemical specifications may also not qualify as Aristotelian universals. The consequences of applying this principle to the development of ontologies is, that categories such as *Higgs boson* or chemicals which are not believed to exist (yet), must not be included in a scientific ontology [46]:

> Designer drugs are conceived, modeled, and described long before they are successfully synthesized, and the plans of pharmaceutical companies may contain putative references to the corresponding chemical universals long before there are instances in reality. But again: such descriptions and plans can be perfectly well apprehended even within terminologies and ontologies conceived as relating exclusively to what is real. Descriptions and plans do, after all, exist. On the other hand it would be an error to include in a scientific ontology of drugs terms referring to pharmaceutical products which do not yet (and may never) exist, solely on the basis of plans and descriptions. Rather, such terms should be included precisely at the point where the corresponding instances do indeed exist in reality, exactly in accordance with our proposals above.

This principle has subsequently been adopted by the OBO Foundry ontologies built on the Basic Formal Ontology (BFO) [16], and is being followed, to the extent possible, by ontologies still under development such as the Ontology of Biomedical Investigations (OBI) [7] and the Information Artifact Ontology (IAO) [14]. These ontologies must only contain classes and use relations that are based on Aristotelian universals. BFO-based ontologies cannot have classes which have no instances in reality.

3. Applications of ontologies in science

To understand whether BFO's criterion to use Aristotelian universals for building ontologies indeed leads to *good* ontologies, we must answer the question whether the use of this principle leads to ontologies that are better suited to solve the problems that ontologies are designed to solve.

Biomedical ontologies have been pursued to increase the accuracy of capturing biological knowledge, whether through the semantic annotation of data, or by representing the outcomes of biological experiments. Over 150 controlled vocabularies and ontologies are now available through the NCBO BioPortal [6] and include ontologies to describe scientific processes such as experimentation, the formulation and testing of hypothesis, and the reporting of scientific results in publications or scientific databases. Ultimately, ontologies serve as a means for knowledge management and knowledge discovery, exemplified by formalized domain-specific languages that are in use in various disciplines, in particular chemistry [9] or molecular biology [23], and sophisticated knowledge representation for model organism biology [47], or in capturing the specific response to drugs based on genetic background [11].

3.1. Database interoperability

Ontologies such as the Gene Ontology (GO) [4] have been designed to facilitate interoperability between different research communities, each of which used a slightly different terminology. This required tedious and potentially erroneous mappings to be created and maintained. Established in 1998, the Gene Ontology Consortium has pursued the development of an ontology to formally and consistently specify the meaning of terms used to describe the processes, functions and cellular locations of gene products. Teams of curators in different model organism projects link gene products to their corresponding function, localization and biological process. Having a common terminological basis has certainly helped improve the progress of bioinformatics in function prediction by identifying putative relations across new or integrated experimental data [37].

Where GO terms refer to categories whose existence has been demonstrated in scientific reports, the criterion of using Aristotelian universals is adequate. However, with more than 25,000 categories in the GO, it is unclear whether all GO categories have been demonstrated to have instances. Even if this were the case, the Gene Ontology and other OBO Foundry [44] ontologies are increasingly being used for other purposes, including hypothesis generation [38] or literature retrieval [10]. As discussed below, these tasks put new and unexpected demands on the ontologies.

3.2. Scientific discourse and hypotheses

Scientific discourse deals with the written or spoken communication of scientific investigation including hypothesis, evidence, physical and computational experiments, materials and methods, data and results including charts and figures. Consider a hypothesis a clever idea or a proposed explanation for some observable phenomenon which may be expressed as a mathematical model or as a logical proposition. But what if the entities described within the hypothesis are not known to exist? This is, after all, a major motivation for, in the very least, biological research. The BFO-realists would balk at including such unknown or debatable entities into a scientific ontology, yet it is obviously critical in not only accurately describing scientific models and theories, but also in being truly innovative and thinking outside already established and accepted scientific knowledge.

For example, the elementary particle *Higgs boson* [21] is predicted by the Standard Model in particle physics [35]. However, so far no instances of this predicted entity have been found, despite considerable efforts [48]. Consequently, within BFO's realist framework, there is no space for an ontological category of *Higgs boson*s, because there may not be any instances of such a category. Yet clearly there is a need amongst scientists to communicate information about Higgs bosons, their properties and especially their potential to enter into causal relations. Furthermore, there is a need to specify what a Higgs boson is (or would be if there is one) in information systems that store, process or retrieve information about elementary particles. So we must answer the question how the nature of hypothetical entities such as *Higgs boson*s can be specified through BFO-realism.

3.3. Information artifacts and the aboutness problem

Hypotheses (as written propositions), theories (as collections of propositions) and written specifications (as documents describing how to represent information) are Information Content Entities (ICEs) in the domain of the Information Artifact Ontology (IAO) [14]. In the IAO, ICEs are generically dependent continuants (GDCs) – these depend on the existence of some material object of a certain kind. They are always *about* [40] some entity in reality. For example, a datum of temperature measurement is generically dependent on a physical representation of the datum, and it is *about* a temperature quality of some entity.

The IAO is represented using OWL-DL [31,36], the description logic variant of the Web Ontology Language (OWL). OWL-DL is a decidable fragment of first order logic, which extends the semantics of the *SROIQ* description logic [25] with datatypes and punning.

The axiom corresponding to the aboutness of information content entities (ICE) is the following description logic statement:

$$ICE \sqsubseteq \exists is\_about.Entity \tag{1}$$

The class *Entity* refers to BFO's top-level class. As BFO is an ontology of particulars, instances of *Entity* must be particulars, too; either continuants or occurrents. *Entity* does not include ontological categories in its extension.

With exceptions (such as a hypothesis about the existence of a planet or asteroid), hypotheses in empirical sciences are about *categories* of particulars. Kinds of hypothetical particles in physics are a consequence of the Standard Model [35] (or other theories). Yet no particular particles are predicted, but rather categories of particles, characterized by their mass, spin and other properties. Limited by the BFO, which includes only particulars in their domain of discourse, the IAO currently fails to represent *about* what information entities referring to categories are.

At least three options exist how to represent theories and hypotheses, or, more generally, information entities, about categories in a BFO compliant way:

1. Using higher-order relations in OWL permits to relate particulars (GDCs) to OWL classes (which represent categories of entities).
2. Adding a new top-level distinction between *Universal*s and *Particular*s in BFO. Instances of the OWL class *Universal* are universals, while the current BFO taxonomy can be kept as a sub-category of the *Particular* class. This approach is illustrated in Figure 1.
3. Representing categories as GDCs that are existentially dependent on a universal requires no change to the BFO.

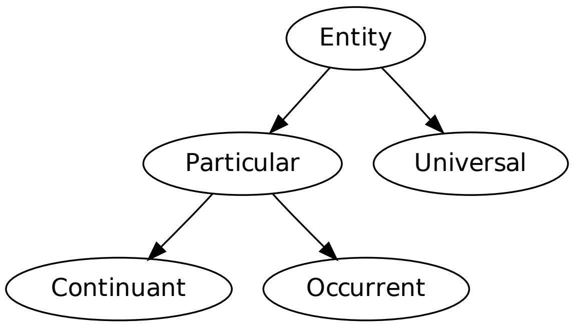

Figure 1. Modified top-level of BFO.

The first option would require no changes to the BFO, yet necessitates the use of an undecidable logic, OWL-Full. The second option is already implemented in the OBO Relationship Ontology (RO) [45] where instances of both categories are related using an explicit **instance-of** relation, and the third option reflects the classification of universals as a classification of GDCs.

However, whichever option is selected to represent information about universals in the BFO, all approaches fail to represent the meaning of terms which refer to categories of beings for which we are not certain that instances exist in this world, because these categories may not be universals, and, consequently, no class corresponding to the category will be included in a BFO-based ontology. Although the BFO has the capability to represent the features of theories or hypotheses, it fails to capture what these theories are about. However, these hypothetical categories are of high importance for the progress of the sciences, and the failure to include them in formal scientific ontologies hinders the ability of scientists to communicate about their theories, and may potentially hinder the progress of science. So far, we have demonstrated this using an example from physics, discussing theories of *Higgs boson*s. Yet the restriction employed by the BFO does not only limit current scientific progress, but fails to serve already established disciplines such as chemistry, where borderline cases and impossibility results are vital for the understanding of particular phenomena and domains.

3.4. Molecules and molecular specifications

Molecules are material entities in that they are maximally connected, spatially extended, retain their identity in time and are the bearer of qualities. Molecules themselves are composed of atoms that are connected through stable covalent bonds, and the three-dimensional arrangement of the atoms that constitute a molecule is called molecular structure. Molecular structure determines several properties including reactivity, polarity, color, magnetism, and biological activity.

Over 80 chemical specification languages exist to capture and represent knowledge about molecules. Molecular graphs (atoms, connectivity and stereochemistry) can be represented by the Simplified Molecular Input Line Entry Specification (SMILES) [1], the IUPAC International Chemical Identifier (InCHI) [32], and extended with other knowledge representation languages such as the XML-based Chemical Markup Language (CML) [34]. More recently, an OWL-based chemical structure ontology [28] has been developed. Since multiple chemical specification languages can be used to specify the same molecule, it is important to provide an ontological foundation for these languages. Furthermore, there is a need to describe information about molecules, particularly the values obtained by a calculation or a measurement protocol. This goal is being addressed in a collaborative effort to develop an ontology of chemical information [13]. The important aspect of encoding this knowledge is that the information content is *about* a molecule or a class of molecules. This viewpoint is critical for the OBO Foundry ontologies IAO and the chemical information ontology: that there exists a relation between the information entities and the real world entities they describe.

Thus, a major challenge for the BFO-realists is how to represent molecular structures that have been designed, but for which no instances are yet known to exist. It has been argued [15] that molecular structure is in fact a "non-effective specification", in contrast to an "effective specification" which details how one might actually execute the procedure by which one realizes the objective. These specifications should *not* hold an aboutness relation to real world entities until, of course, the real world entity has been demonstrated to exist. It is further argued that any properties that we might derive from the structure are simply predictions, and don't really describe the real world entity.

Yet, it is entirely unsatisfying that we cannot examine a glyph, diagram or sequence of characters, conforming to one of the molecular structure specification languages and acknowledge that this description in fact corresponds to a specific category of molecules, independent of whether we know it to exist. Moreover, whether the specified molecules exist or not, the properties we compute are also information about the specific molecules. Instances of the category of molecules would indeed have all the normal properties and characteristics of molecules, and they are studied under that aspect.

It is routine science for chemists to investigate chemicals that do not yet exist. During the course of their investigation, they specify the target molecule's structure and put forward biosynthetic strategies. Increased due diligence could include a computational study to assess the compound's stability, so as to better understand under which conditions it may be synthesized. For example, de Silva and Goodman [8] pursue a computational chemistry approach to answer the question of which saturated acyclic alkanes can never be made under normal conditions. They identify two molecules, illustrated in Figure 2, which have never been synthesized and are unlikely to exist except perhaps briefly under extremely low temperatures. It is clear that their study provides information about

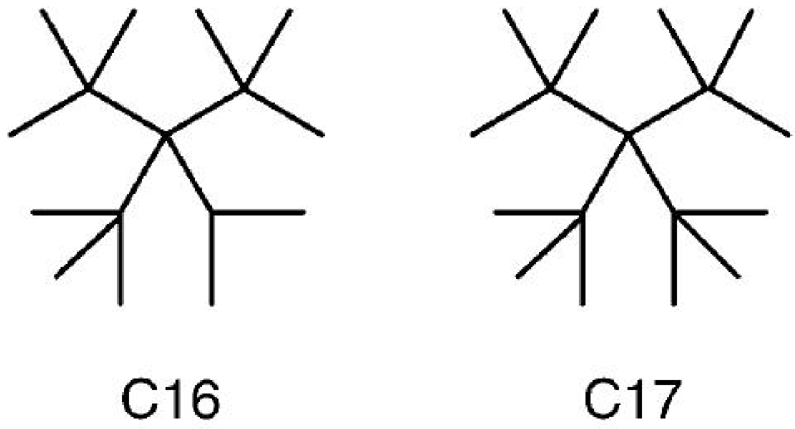

Figure 2. The smallest acyclic alkanes that could not naturally exist.

a class of chemical compounds for which there is no *known* physical evidence of their existence, and the scientific contribution is exactly the impossibility result, as outlined in the study's conclusion [8]:

> The strain energy analysis and the DFT calculations suggest that the limit of stability lies either at C15 or C16, and so the smallest alkane that can never be made will be either C16 or C17, except perhaps at very low temperatures for a short time. We are confident that C17 could never be made and suggest that C16 is the smallest saturated alkane that cannot be made.

While BFO-based ontologies can represent the properties and features of the *specification* of a class of molecules, it fails to represent the properties and features of the class of specified objects – independent from whether or not instances of the specified class of objects exist.

3.5. Perpetual motion

Perpetual motion and perpetual motion machines are discussed in science for centuries. They have served to test the boundaries of scientific theories and hypotheses [2] and continue to do so [20].

A perpetual motion machine is a machine that runs forever. Perpetual motion machines violate either the first or the second law of thermodynamics [12]. Perpetual motion machines of the first kind create energy from nothing (and violate the first law of thermodynamics about the conservation of energy). Perpetual motion machines of the second kind are able to extract energy at some point, use it for work, and have everything return to an original state at the end of the cycle (violating the second law about increasing entropy). Using other criteria, several other classes of perpetual motion machines can be distinguished [26].

The notion of perpetual motion and perpetual motion machines continues to be useful to test the limits of scientific theories and challenge new discoveries. Hence, a scientific classification (or ontology) of perpetual motion machines has been developed. It is precisely the fact that these machines are considered to be impossible that makes them useful for science [5]. Consequently, they are useful in scientific communication and information systems in scientific domains, and therefore an important notion for scientific ontologies.

3.6. Alternatives to BFO's realism

So, are there alternatives to the use of Aristotelian universals for building ontologies? There are, and they are quite common in the formal ontology community outside the biomedical domain [18,17]. It is the view of ontology as a specification of a conceptualization of a domain, as a formal theory specifying the *meaning* of terms in a vocabulary – a vocabulary containing all the relevant terms within the domain to be covered. In this view, an ontology is not concerned with the question whether one of its classes has an instance *in this world* (or in "reality"), but rather with specifying the properties and characteristics of the instances *if there are any*.

The importance of *impossibility* results in science such as the apparent impossiblity of perpetual motion or certain chemical structures has some important consequences for constructing scientific ontologies. When formal ontologies for science specify the necessary properties of a category's instances (i.e., the properties of the instances in every possible world in which the category does have instances), the laws of nature must be seen as contingent, because these laws are the subject matter of the domain modeled. Scientific ontologies are intended to support the communication between scientists and must, by their very nature, be able to accommodate different scientific views. Restricting the categories in scientific ontologies by the natural laws themselves would make it impossible to use these ontologies for the discovery, discussion or communication of novel results that may contradict any of the established scientific theories, because the choice of the ontology would already preclude the possibility of such results.

It may still be desirable to know whether a majority of scientists *believe* some theory to be true or false, or some category to have instances or not. For this purpose, OWL meta-modelling [33] can be used to add epistemic information to an ontological class, e.g., with attributes such as *predicted*, *hypothetical* or *confirmed*.

4. Discussion

The criterion employed by the BFO to base ontologies on universals that are instantiated fails to satisfy the needs of scientists in various illustrative cases that are representative of a larger set of more general problems. It must be emphasized that, within the BFO, no definition or axiomatic characterization of "universal" is currently available. The closest to a definition of "universal" in the BFO is the statement [43]:

> A universal is defined as anything that is instantiated, and an instance as anything that instantiates some universal. The relation of instantiation is hereby taken as primitive, and it is specified axiomatically that it holds exclusively between instances and universals (in that order).

However, this does not help us to determine if an entity is a universal or an instance of a universal, or even whether instances and universals are disjoint or may overlap.

Yet it has been argued that BFO's criterion of using only Aristotelian universals for classes in an ontology leads to better ontologies, because it helps to prevent various mistakes. In particular, false **is-a** assertions are claimed to be prevented through the use of the criterion.

However, it is hard to see how the use of universals contributes to avoiding the identified mistakes [43]. The argumentation in [43] seems entirely focused on the benefits of using universals over *concepts*. The notion of concept is, it is claimed [43,42], not well-defined, and should be replaced by the notion of "universal". Yet, within the BFO, "universal" is only specified as something that has instances, making it a weak notion rarely more adequate than "concept". It is therefore unclear why **is-a** relations such as

- dog **is-a** dog or apple,
- dog and mammal **is-a** dog or apple,
- dog **is-a** non-cat

are apparently excluded with the use of universals instead of concepts. Since axioms or definitions for "universal" are not provided in the BFO, no option remain to determine whether "dog or apple" is a universal or not – except by reference to an ontological authority.

We focus on the BFO here due to its importance in constructing ontologies in the biomedical and related domains. However, other ontologies are used within these domains as well, in particular DOLCE [29] in the BioTop project [39] and the GFO [19] in the GFO-Bio project [24].

Neither DOLCE nor the GFO have a restriction similar to the BFO in that their classes must correspond to Aristotelian universals. As a consequence, neither faces the challenges in representing scientific theories, hypotheses or specifications we discussed so far. Furthermore, each explicitly includes some form of representing ontological categories. The DOLCE contains, in an extension [30], a way of constructing *concepts*.

The GFO is designed from the beginning as an ontology not only of individuals, but also of ontological categories. GFO permits the inclusion of multiple kinds of categories, including mind-independent categories (universals) and mind-dependent categories (concepts).

5. Conclusions

We discussed a form of realism that is promoted by the Basic Formal Ontology and is being employed throughout biomedical ontologies that are candidates for inclusion in the OBO Foundry. This form of realism states that the classes and relations in these ontologies must be instantiated in reality. In particular, categories of entities which are not assumed or *known* to exist cannot be included in ontologies exemplifying this form of realism. This includes a number of entities predicted by current theories in physics such as *Higgs boson*, and also the referents of chemical specifications, in particular molecules.

The form of realism employed by the BFO and the OBO Foundry ontologies is inadequate for numerous tasks that ontologies are being used for, in particular for the capture and communication of scientific discourse, the representation of the referents of specifications and, in general, the contents of scientific theories. While BFO's realism criterion is intended to facilitate the construction of "good" ontologies, it hinders at the same time the development of more useful ontologies for science. We show that other criteria which are more common in the formal ontology community, such as the use of the axiomatic method [22] and the specification of the meaning of every term necessary within a domain or application, suffice to avoid the problems BFO's realism criterion tries to address, and apply it to uninstantiated classes such as the referents of scientific theories, chemical specifications and impossibility results.

Acknowledgements

We thank Nico Adams for interesting and helpful discussions about realism and chemicals that cannot exist. We thank two anonymous reviewers for valuable comments on an earlier draft of this manuscript.

References

[1] E. Anderson, G.D. Veith, and D. Weininger. Smiles: A line notation and computerized interpreter for chemical structures. Technical report.

[2] Stanley Angrist. Perpetual motion machines. *Scientific American*, 218(1):115–122, 1968.

[3] Aristotle. *De Interpretatione*, chapter 7. Electronic Text Center, University of Virginia Library, 2001.

[4] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, Issel L. Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 25(1):25–29, May 2000.

[5] John D. Barrow. *Impossibility : The Limits of Science and the Science of Limits*. Oxford University Press, 1998.

[6] Ncbo bioportal. `http://bioportal.bioontology.org/`.

[7] Mélanie Courtot et al. The owl of biomedical investigations. In Catherine Dolbear, Alan Ruttenberg, and Ulrike Sattler, editors, *OWLED*, volume 432 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2008.

[8] K. M. Nalin de Silva and Jonathan M. Goodman. What is the smallest saturated acyclic alkane that cannot be made? *J. Chem. Inf. Model.*, 45(1):81–87, 2005.

[9] K. Degtyarenko, P. Matos, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcantara, M. Darsow, M. Guedj, and M. Ashburner. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Research*, 2007.

[10] A. Doms and M. Schroeder. GoPubMed: exploring PubMed with the Gene Ontology. *Nucleic Acids Res*, 33(Web Server issue):783–786, Jul 2005.

[11] Michel Dumontier and Natalia Villanueva-Rosales. Towards pharmacogenomics knowledge discovery with the semantic web. *Briefings in bioinformatics*, February 2009.

[12] Enrico Fermi. *Thermodynamics*. Dover Publications, New York, 1956 [1936].

[13] `http://code.google.com/p/semanticchemistry/`.

[14] `http://code.google.com/p/information-artifact-ontology/`, 2009.

[15] `http://groups.google.com/group/bfo-discuss/browse_thread/thread/93d960b1f3deee39`.

[16] Pierre Grenon. Bfo in a nutshell: A bi-categorial axiomatization of BFO and comparison with DOLCE. Technical report, University of Leipzig, Leipzig, 2003.

[17] Thomas R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43(5-6), 1995.

[18] Nicola Guarino. Formal ontology and information systems. In Nicola Guarino, editor, *Formal Ontology in Information Systems: Proceedings of the First International Conference (FOIS'98), Trento, Italy, 6-8 June 1998*, volume 46 of *Frontiers in Artificial Intelligence and Applications*, Amsterdam, June 1998. IOS Press.

[19] Heinrich Herre, Barbara Heller, Patryk Burek, Robert Hoehndorf, Frank Loebe, and Hannes Michalek. General Formal Ontology (GFO) – A foundational ontology integrating objects and processes [Ver-

sion 1.0]. Onto-med report, Research Group Ontologies in Medicine, Institute of Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, 2006.

[20] Paul R. Heyl. Perpetual motion in the twentieth century. *The Scientific Monthly*, 22(2):143–145, 1926.

[21] P. W. Higgs. Broken symmetries, massless particles and gauge fields. *Physics Letters*, 12:132–133, September 1964.

[22] David Hilbert. Axiomatisches denken. *Mathematische Annalen*, 78:405–415, 1918.

[23] Robert Hoehndorf, Janet Kelso, and Heinrich Herre. The ontology of biological sequences. *BMC Bioinformatics*, 10(1):377+, 2009.

[24] Robert Hoehndorf, Frank Loebe, Roberto Poli, Janet Kelso, and Heinrich Herre. Gfo-bio: A biological core ontology. *Applied Ontology*, 2008. in press.

[25] Ian Horrocks, Oliver Kutz, and Ulrike Sattler. The even more irresistible sroiq. In *In KR*, pages 57–67. AAAI Press, 2006.

[26] Kevin T. Kilty. Perpetual motion. `http://www.kilty.com/pmotion.htm`, 1992.

[27] Gyula Klima. The medieval problem of universals. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Winter 2008 edition, 2008.

[28] Mykola Konyk, Alexander De Leon, and Michel Dumontier. Chemical knowledge for the semantic web. In *Proceedings of Data Integration in the Life Sciences (DILS2008)*, page Lecture Notes in Bioinformatics (LNBI). Springern, 2008.

[29] Claudio Masolo, Stefano Borgo, Aldo Gangemi, Nicola Guarino, and Alessandro Oltramari. WonderWeb Deliverable D18: Ontology library (final). Technical report, Laboratory for Applied Ontology, ISTC-CNR, Trento, Italy, 2003.

[30] Claudio Masolo, Laure Vieu, Emanuele Bottazzi, Carola Catenacci, Roberta Ferrario, Aldo Gangemi, and Nicola Guarino. Social roles and their descriptions. In *Proceedings of the Ninth International Conference on the Principles of Knowledge Representation and Reasoning (KR2004), Whistler, Canada, June 2-5, 2004*, 2004.

[31] Deborah L. Mcguinness and Frank van Harmelen. OWL Web Ontology Language overview. W3C recommendation, World Wide Web Consortium (W3C), February 2004.

[32] Alan McNaught. The iupac international chemical identifier:inchl. *Chemistry International*, 28(6), 2006.

[33] Boris Motik, Peter F. Patel-Schneider, and Bernardo Cuenca Grau. Owl 2 web ontology language: Direct semantics. Technical report, W3C, 2009.

[34] P. Murray-Rust and H. S. Rzepa. Chemical markup, xml and the worldwide web. part 4. cml schema. *Journal of Chemical Information and Modeling*, pages 757–772, 2003.

[35] Cheng Ta Pei and Li Ling Fong. *Gauge Theory of Elementary Particle Physics*. Oxford, 1984.

[36] Peter, Patrick Hayes, and Ian Horrocks. Owl web ontology language – semantics and abstract syntax. urlhttp://www.w3.org/TR/owl-semantics/, February 2004.

[37] Kay Prufer, Bjoern Muetzel, Hong-Hai Do, Gunter Weiss, Philipp Khaitovich, Erhard Rahm, Svante Paabo, Michael Lachmann, and Wolfgang Enard. Func: a package for detecting significant associations between gene sets and ontological annotations. *BMC Bioinformatics*, 8:41+, February 2007.

[38] S. A. Racunas, N. H. Shah, I. Albert, and N. V. Fedoroff. Hybrow: a prototype system for computer-aided hypothesis evaluation. *Bioinformatics*, 20 Suppl 1, August 2004.

[39] Stefan Schulz, Elena Beisswanger, Joachim Wermter, and Udo Hahn. Towards an upper-level ontology for molecular biology. *AMIA Annu Symp Proc*, 2006, 2006.

[40] John R. Searle. *Intentionality: An essay in the philosophy of mind.* Cambridge University Press, 1983.

[41] John R. Searle. *The Construction of Social Reality*. Free Press, January 1997.

[42] B. Smith. From concepts to clinical reality: an essay on the benchmarking of biomedical terminologies. *J Biomed Inform*, 39(3):288–298, June 2006.

[43] Barry Smith. Beyond concepts: Ontology as reality representation. In Achille Varzi and Laure Vieu, editors, *Proceedings of FOIS 2004*, 2004.

[44] 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. *Nat Biotech*, 25(11):1251–1255, November 2007.

[45] 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), 2005.

[46] Barry Smith, Waclaw Kusnierczyk, Daniel Schober, and Werner Ceusters. Towards a reference terminology for ontology research and development in the biomedical domain. In *Proceedings of KR-MED 2006*, 2006.

[47] Natalia Villanueva-Rosales and Michel Dumontier. yowl: An ontology-driven knowledge base for yeast biologists. *J. of Biomedical Informatics*, 41(5):779–789, 2008.

[48] W. M. Yao et al. Searches for higgs bosons. *Journal of Physics G: Nuclear and Particle Physics*, 33(1), 2006.