

The Bridge between Philosophy and Information-Driven Science

Barry SMITH*

Department of Philosophy, University at Buffalo

Journal of Knowledge Structures & Systems

December 2021

Vol.: 2 Issue: 2 Pages: 47-55

Abstract

This essay is a response to Luis Augusto's intriguing paper (this issue) on the rift between mainstream and formal ontology. I will show that there are in fact two questions at issue here: 1. concerning the links between mainstream and formal approaches within philosophy, and 2. concerning the application of philosophy (and especially philosophical ontology) in support of information-driven research for example in the life sciences.

Key words: Applied ontology; Formal philosophy; Gene Ontology; Philosophy of science; Basic Formal Ontology

WHY I AM NOT A PHILOSOPHER. In October 2006 I presented a talk with this title to my colleagues in the University at Buffalo.¹ My thesis was that philosophy, as on previous occasions over the centuries (for example in the case of psychology at the end of the 19th century), was on the point of shedding a new discipline—in this case the discipline of ontology. Gary Merrill, another philosopher like myself working on the practical applications of ontology in life science research, responded to this idea some time later by arguing that it is part of the philosopher's job precisely to build ontologies, or in other words to construct "systems of categories to be used in describing and understanding the world around us" (Merrill, 2011). He points out that already Leibniz, from a distance of 300 years, had

characterized much of the work being done today in the domain of informatics, and especially medical informatics—mentioning explicitly the value of systematic classification in domains that we would now think of as

*✉ phismith@buffalo.edu

¹A version of this talk from 2018 is available at [1].

knowledge representation, knowledge management, information retrieval, and inferencing—and Leibniz sees it as the work of the philosopher. (*Op. cit.*)

Perhaps taking the title of my talk a mite too seriously, Merrill argues against the idea, which he calls the *evolution model*, of a spinning off of ontology from its philosophical mothership. He promotes instead what he calls the *strong participation model*, according to which philosophers should involve themselves *qua* philosophers, and so still working within departments of philosophy as standardly conceived, in addressing the tasks of systematic classification in the information-driven sciences. Their role, then, is to reach out with friendly offers of assistance to their eventual scientist counterparts in helping the latter to build properly scientific ontologies. On the evolution model defended in my talk, in contrast, ontology should be viewed as being a discipline which began its life as a part of philosophy but which is now breaking free, with some inevitable birthing pains, to become a discipline in its own right, and with results of its own which belong properly to science and not to philosophy (Smith, 2008).

Luis Augusto’s paper (Augusto, 2021), too, defends a participation model. It proposes, indeed, that work for example by hitherto neglected scholastic philosophers may have something to teach us as concerns how we build ontologies for application purposes today. His paper contains valuable reflections on the general issue of collaboration—or lack of collaboration—between what he calls “mainstream” and “formal ontology”, and he notes that there is a parallel issue also in the field of epistemology, and indeed in philosophy as a whole. This concerns not the question of philosophy in relation to its application in the sciences, but rather to the relation within philosophy between philosophers in the mainstream and those who use formal methods. There are of course individuals, starting with Frege and extending through in our own day to Kit Fine, who have contributed to what we might think of as the creative core of mainstream philosophy through their use of formal methods. Perhaps we are now, as Augusto suggests, entering a period in which work at the interface between philosophical treatments of knowledge and work on the application of knowledge in machines will also yield contributions to this creative core. There are reasons to be skeptical also here, however. Work by philosophers in areas such as AI (for example in AI ethics) is typically not formal in the mathematical or logical sense. And it has to be said that individuals like Frege, or Gödel, or Tarski are exceptions, and that high-quality formal philosophy, for example as practiced by Ed Zalta in the Metaphysics Research Lab [2], still largely fails to affect the priorities of mainstream philosophers.

FORMAL ONTOLOGY AND ITS APPLICATIONS. What, now, of the relations between philosophy (both formal and mainstream) and ontological work in the domain of scientific applications? There are now many applied ontologies which are being used in information-driven scientific disciplines to improve the degree to which empirical data can be captured in forms that make it more readily understandable, reusable, integratable and analyzable.

What should be the relation between philosophical ontology and this applied ontology? Issues here are complex, and they have as much to do with the nature of applied ontology as an “applied” discipline as with the nature of ontology simpliciter. The protectors of the rights of university departments have a well-entrenched tendency to look down on “applied” versions of their respective disciplines, seeing them

as being in some respects watered down versions of what can be achieved by theoreticians uncontaminated by applications. Creating watered down versions of theories is of course often an indispensable step in building useful applications. At the same time, simplifications will sometimes bring negative consequences, not least in mission critical for example in hospitals or nuclear power stations. But then unsimplified (“pure”) theories which never yield any applications or practical consequences are for scientific purposes uninteresting.

These matters can be illustrated by pointing to the fact that universities took a long time finding a place for biomedical informatics—a form of applied computer science—in their organizational structures, since neither the biomedical departments nor the departments of computer science saw work in this not-yet-discipline as properly scientific. A similar story can be told about applied physics, and the same story applies even more poignantly to the not-yet-discipline of applied ontology.

This may provide one good reason why applied ontology should want to leave its philosophical mother discipline, but a move of this sort also has its problems, and the idea that (applied) ontology will evolve from being a branch of philosophy to being a discipline in its own right is at best a rather simplified portrayal of what will happen in the future. One strategy is already being implemented in the University at Buffalo, where the Medical School has created within its very large Department of Biomedical Informatics a Division of Biomedical Ontology. This is, as far as we know, the first academic unit in any university devoted explicitly to the discipline of applied ontology.

Formal ontology as Augusto conceives it “requires an ontology to be machine-interpretable, or at least machine-readable” (Augusto, 2021); its constructs must therefore be “expressible in a formal language, preferably a logical one, that not only can represent knowledge as machine-readable constructs, but also allows for automated deduction over those constructs” (*ibid.*). But of course not just any formal ontology which meets these requirements will serve the purposes of aiding scientists.

THE BIRTH OF BFO. The founding of the UB Division of Biomedical Ontology, like my 2006 talk, grew out of the work of the Institute for Formal Ontology and Medical Information Science (IFOMIS) which I founded in 2002. It was clear almost from the beginning that the work of this Institute would consist primarily in collaborating with molecular biologists, clinical scientists, and others to find sustainable ways to support the useful exploitation of the huge amounts of data (then, and still today) being produced in the life sciences in the wake of the successful decoding of the human and other model organism genomes.

The question at hand was, how are we to link the new sequence data, which look like this:

```
MTQLQISLLLTATISLLHLVWATPYEAYPIGKQYPPVARVNESFTFQISNDTYKS
SVDKTAQITYNCFDLPSWLSFDSSSRFTSGEPSSDLLSDANTTLYFNVILEGTDS
ADSTSLNNTYQFVVVNRPSISLSSDFNLLALLKNYGYTNGKNALKLDPNEVFVNT
FDRSMFTNEESIVSYGRSGLYNAPLPNWLFFDSGELKFTGTAPVINSIAIAPETS
YSFVIIATDIEGFSAVEVEFELVIGAHQLTTSIQNSLIINVTDTGNVSYDLPLNY
VYLDDDPISSDKLGSINLLDAPDWVALDNATISGSVPDELLGKNSNPANFSVSIY
DTYGDVIYFNFEVSTTDLFAISSLPNINATRGEWFSYFLPSQFTDYVNTNVSL
EFTNSSQDHDW (Source: [3])
```

to biological data of the more traditional sort, relating for example to genetic diseases such as cystic fibrosis or Marfan syndrome? One initial guiding hypothesis was that, if we could identify the genes or proteins responsible for the counterparts of such diseases in mouse or fly or zebrafish, then we could perform experiments on these model organisms using drugs or treatments targeted to those genes or proteins—experiments which we could not perform on human beings—and thereby draw conclusions as to what might be the effects of these drugs or treatments for the parallel conditions in human beings.

How, then, to link together the two types of data to make possible inferences of this kind? The solution to this problem was *data annotation*, which means, roughly, taking expressions from old biology and from medicine, and associating these expressions with specific sequences in the newly burgeoning (genomic and proteomic and now many other sorts of -omic) databases. This will work, however, only if mouse and fly and zebrafish (and rat and worm and yeast and bonobo) biologists use the same terms with the same meanings when annotating their respective genomes—which to a large extent, in 1998, they did not. In that year, however, there was born the GO, for “Gene Ontology”, described by its founders as “a tool for the unification of biology” (Ashburner et al., 2000). The GO was designed to provide a species- and discipline-neutral *controlled vocabulary* for the annotation of biological data, in which each term would have a definition; in which the terms would be arranged in graph-theoretic hierarchies of greater and lesser generality; and in which lateral relations—especially parthood—would be defined between the terms. The ontology was built not merely as a digital artifact, but also as a service to the community—thus it incorporated feedback strategies whereby users could report gaps and errors and thereby ensure that the ontology would evolve in tandem with advances in scientific knowledge (Smith, 2019).

The GO was an immediate success, and its value to biomedical researchers continues to grow from year to year. But initially it also had certain problems:

1. Its scope was limited to just those three sorts of data which are of immediate importance when a new gene or gene product has been identified, namely: where is it located in the cell? what is its function on the molecular level? and how does the realization of this function contribute to what happens downstream biologically? The GO is, correspondingly, divided into three submodules of: Cellular Component, Molecular Function, and Biological Process.
2. The existence of data annotated with a structured controlled vocabulary like the GO should in principle enable automated reasoning over the data; but these possibilities could hardly be realized by the original versions of the GO, since there were problems with the formal structure of the ontology, especially concerning the definitions of and the treatment of the relations between the terms.

The definitions in the ontology and the structural principles used in building its three term hierarchies were underspecified, so that terms and other syntactic elements were used inconsistently from one part of the ontology to the next.

An example of a problem of this sort is the lack of a coherent definition of the relation of parthood in early versions of the GO. Indeed, according to the *GO User Guide* of 2003, “A part of B”, where “A” and “B” are terms in the GO, can mean one or more of:

- A can be part of B,
- vocabulary A is included within vocabulary B, or
- A is sometimes part of B in the sense that there is for each instance of A some time at which it is part of an instance of B.²

There were also structural gaps, for example because no ontological relations are defined which would connect the three hierarchies. This meant that the GO could not, for example, record the fact that a given function is the function of a given component, or that a given process is the exercise of a given function. There is, for example, no linkage in the ontology between the term “taste”, which is a Biological Process term, and the term “taste receptor activity”, which is a Molecular Function term.

To address 1. the GO community encouraged the development of sister ontologies (Ashburner & Lewis, 2003), including anatomical ontologies; molecule, cell and tissue type ontologies; a disease ontology, and so forth. Unfortunately, this expansion of the realm of GO-associated ontologies led to even further difficulties under 2., since the new ontologies were being built following different structural principles, which made it difficult to build the bridges between them that would be needed to do justice to the interrelations between for example molecules, cells, and diseases in the real world.

It was as part of an effort to solve such problems that I was invited to join the GO community following a meeting in Leipzig in May 2004 entitled *The Formal Architecture of the Gene Ontology*, where I gave a talk entitled “Stop! or: Smart Terminologies through Ontological Principles”.

An outcome of this meeting was the decision to create a common set of principles, not only for the GO but also for all sister ontologies that would be used in tandem with the GO, within the framework of what we called the Open Biomedical Ontologies (OBO) Foundry (see [4]).

One principle, initially uncodified, was that all ontologies in the OBO Foundry should use the same top-level ontology, which should be modelled on the top-level ontology used by the GO when it divided the biological reality that is studied in molecular biology and genetics experiments into Cellular Components, Molecular Functions, and Biological Processes. In philosophical terms, this implied a top-level ontology constructed around categories of independent continuants (such as cells and their components), dependent continuants (such as qualities and functions), and occurments (such as processes), respectively.

The top-level ontology created to address this requirement was called “Basic Formal Ontology” (see [5]), and BFO is now used not only in the biological domain but also in areas such as the spatial and geosciences, in engineering, and in defense and intelligence. On the basis of this evidence of reuse across a wide scope of disciplines BFO was approved in 2021 as standard (ISO/IEC 21838-2) by the Joint Technical Committee (JTC 1) of the International Standards Organization and the International Electrotechnical Commission.

BFO: A BRIDGE BETWEEN COMPUTATIONAL AND PHILOSOPHICAL ONTOLOGY. The content of BFO consists of terms, relational expressions, definitions,

²On the initial steps to improve matters in this respect see Smith, Köhler & Kumar (2004); Smith & Rosse (2004); Bittner et al. (2004).

and examples of use. To enable use by human beings this content is made available, first, in natural language and in the standard first-order logic format. But it is available also both in the W3C Web Ontology Language format (OWL) widely used in applied ontology circles, and in the Common Logic format, which is an enhanced version of first-order logic optimized for computational purposes.

The ontology itself is a combination of elements deriving from Aristotle's metaphysics (Smith, 2000) and Husserl's *Logical Investigations* (which contains the first use of the term "formal ontology"). It contains a formal treatment of functions (derived from Millikan (1989)), which it applies systematically to both biological and artifact functions (Smith et al., 2016), and Smith (2022) shows how this treatment can elucidate terminological confusions regarding "function" in the Gene Ontology.

BFO also serves as top-level architecture for a number of philosophically relevant ontologies at lower levels, including, from the OBO Foundry [4]:

IAO (Information Artifact Ontology) – covers information entities, which are defined in terms of the relation of aboutness

MFO (Mental Functioning Ontology) – covers cognitive processes and associated traits, such as intelligence; includes extension ontologies for mental disease (MFO-MD) and for emotions (MFO-EM)

OBI (Ontology for Biomedical Investigations) – covers much of the ground of philosophy of science

ECO (Evidence Ontology) – covers experimental and other evidence statements

SWO (Software Ontology) – covers software tools, their types, tasks, versions, provenance and associated data

All of these provide, like BFO itself, formal definitions, available both in natural language and in logical form of terms used in much-discussed areas of philosophy. They enable computer reasoning with these terms (including automatic consistency checking for example when new logical definitions are proposed for existing terms or when new terms or new axioms are proposed for inclusion in an ontology). But with very few exceptions none of these ontologies has been studied from a philosophical point of view.³ As Augusto (2021) puts it, modern computational ontologies "are direct heirs to the ontological investigations in the long Western philosophical tradition, but it is not easy to make bridges between them."

ONE FORMAL ONTOLOGY TO RULE OVER THEM ALL. I will conclude with some remarks on another feature of Merrill's *strong participation model*, which he describes as follows:

the role of the philosophical ontologist in science . . . is not to serve up a complex metaphysical theory as the starting point for the construction of a scientific ontology, but rather to work together with scientists in applying effective methods and principles (based, perhaps, on one or more such theories, even if these are of an incomplete or fragmentary nature) to specific problems facing the scientist. (*Op. cit.*)

³These exceptions include, apart from Augusto and Merrill, Adrien Barton (Bihan & Barton, 2021), Fumiaki Toyoshima (2020), and Ludger Jansen (2018).

It is a commonplace especially among philosophers of science that philosophers should not attempt to tell scientists how to go about their business. However, almost 20 years of interacting with scientists and others in the job of creating practically useful ontologies have taught me that, at least in the domain of applied ontology, this piece of conventional wisdom is misguided. When one of the axioms in the early versions of the GO is *molecular process is_a Gene Ontology*, there is something going wrong philosophically with the way scientists are building their ontologies.

Efforts to rectify such problems across multiple ontologies can succeed only if ontologists do their work on the basis of some single, logically well-articulated framework (compare Haller, 2020). A grounding of this sort is, I believe, a pre-condition of doing meaningful and successful ontology in a scientific context.

METAPHYSICS AS AN EMPIRICAL DISCIPLINE. The principal thesis of my 2006 talk was that the work I had then been doing for just 3 years in collaboration with scientists on the building of *formal* or *computational* or *applied* ontologies was not properly a part of philosophy. Rather, it was a combination of philosophical, logical, and computational components wrapped around a set of primary empirical axes deriving from the empirical disciplines to which the ontologies were being applied. Naively, I had hoped at that stage that this new discipline of what is now in many circles called “applied ontology” would be of interest to mainstream philosophers, and indeed that it might provide an empirical test for the validity of mainstream approaches to metaphysics.

The idea would be that distinct metaphysical frameworks might compete to determine which could bring most benefit when applied to assist the work of scientists in enhancing their data. Interestingly, like all the upper-level ontologies which have been successfully applied in different disciplines—which means Cyc (Lenat et al., 1986), DOLCE (Masolo et al., 2003), and SUMO (Niles & Pease, 2001)—BFO rests on a three-dimensionalist (3D) metaphysical theory, which comprises (like the GO) three main branches of independent and dependent continuants and occurrents. This observation would seem to have empirically testable implications for the proponents of four-dimensionalist ontologies. Are they able to demonstrate that a 4D ontology could be at least be of equal utility to scientists (or indeed be of any utility at all)?

References

- Ashburner, M., Ball, C. A., Blake, J. A. et al. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, 25(1), 25-29.
- Ashburner, M. & Lewis, S. E. (2003). On ontologies for biologists: The Gene Ontology – untangling the web. In G. R. Bock & J. A. Goode (eds.), *In Silico simulation of biological processes*. New York: John Wiley & Sons.
- Augusto, L. M. (2021). Bridging mainstream and formal ontology: A causality-based upper ontology in Dietrich of Freiberg. *Journal of Knowledge Structures & Systems*, 2(2), 1-35.

- Bittner, T., Donnelly, M., & Smith, B. (2004). Individuals, universals, collections: On the foundational relations of ontology. In A. Varzi & L. Vieu (eds.), *Formal ontology in information systems. Proceedings of the Third International Conference* (pp. 37–48). Amsterdam: IOS Press.
- Haller, A., & Polleres, A. (2020). Are we better off with just one ontology on the Web? *Semantic Web*, 11(1), 87-99.
- Husserl, E. (1913/2000). *Logical investigations*. 2 vols. Transl by J. N. Findlay. 2nd ed. London: Routledge.
- Jansen, L. (2018). Functions, malfunctioning, and negative causation. In A. Christian et al. (eds.), *Philosophy of science. Between the natural sciences, the social sciences, and the humanities* (pp. 117-135). Cham: Springer.
- Le Bihan, B. & Barton, A. (2021). Analytic metaphysics versus naturalized metaphysics: The relevance of applied ontology. *Erkenntnis*, 86, 21–37.
- Lenat, D. B., Prakash, M., & Shepherd, M. (1985). CYC: Using common sense knowledge to overcome brittleness and knowledge acquisition bottlenecks. *AI magazine*, 6(4), 65-65.
- Masolo, C., Borgo, S., Gangemi, A., Guarino, N., & Oltramari, A. (2003). Wonder-Web Deliverable D18. Ontology Library (Final). Technical Report. Laboratory for Applied Ontology. ISTC-CNR, Trento.
- Merrill, G. H. (2011). Ontology, ontologies, and science. *Topoi*, 30(1), 71-83.
- Millikan, R. G. (1989). In defense of proper functions. *Philosophy of Science*, 56, 288–302.
- Niles, I. & Pease, A. (2001). Towards a standard upper ontology. In C. Welty & B. Smith (eds.), *Formal ontology in information systems* (pp. 2-9). Ogunquit, Maine.
- Smith, B. (2000). Objects and their environments: From Aristotle to ecological ontology. In A. Frank et al. (eds.), *Life and motion of socio-economic units* (pp. 84-102). Boca Raton: CRC Press.
- Smith, B. (2008). Ontology (Science). In C. Eschenbach & M. Grüninger (eds.), *Formal Ontology in Information Systems (FOIS)* (pp. 21-35). Amsterdam: IOS Press.
- Smith, B. (2019). Ontology as product-service system: Lessons learned from GO, BFO and DOLCE. In *Proceedings of the 10th International Conference on Biomedical Ontology (ICBO)*, Buffalo, NY.
- Smith, B. (2022). Biomedical Ontologies. In P. Elkin (ed.), *Terminology, ontology and their Implementations*. Springer Nature Switzerland AG. (In press.)

- Smith, B., Köhler, J., & Kumar, A. (2004). On the application of formal principles to life science data: A case study in the Gene Ontology. In E. Rahm (ed.), *Proceedings of DILS 2004 (Data Integration in the Life Sciences)* (pp. 79–94). (Lecture Notes in Bioinformatics 2994). Berlin: Springer.
- Smith, B. & Kumar, A. (2004). Controlled vocabularies in bioinformatics: A case study in the Gene Ontology. *Drug Discovery Today: BIOSILICO*, 2(6), 246-252.
- Smith, B. & Rosse, C. (2004). The role of foundational relations in the alignment of biomedical ontologies. *Studies in Health Technology and Informatics*, 107, 444-448.
- Spear, A.D., Ceusters, W., & Smith, B. (2016). Functions in Basic Formal Ontology. *Applied Ontology*, 11(2), 103-28.
- Toyoshima, F. (2020). Natural necessity: An introductory guide for ontologists. *Applied Ontology*, 15(1), 61-89.

Online Resources

- [1] <https://www.youtube.com/watch?v=e5zjZnbi-ZA>
- [2] <https://mally.stanford.edu/>
- [3] <https://www.ncbi.nlm.nih.gov/genbank/samplerecord/>
- [4] <https://obofoundry.org>
- [5] <http://basic-formal-ontology.org>

Cite this article as:

Smith, B. (2021). The bridge between philosophy and information-driven science. *Journal of Knowledge Structures & Systems*, 2(2), 47-55.

EDITORIAL INFORMATION

Editor-in-chief:^a Luis M. Augusto

^aCommentary articles are reviewed only by the editor-in-chief for appropriateness of content and style.