Discussion of “Biomedical Ontologies: Toward Scientific Debate”

M. Brochhausen¹; A. Burgun²; W. Ceusters³; A. Hasman⁴; T. Y. Leong⁵; M. Musen⁶; J. L. Oliveira⁷; M. Peleg⁸; A. Rector⁹; S. Schulz¹⁰

¹Saarland University, Saarbrücken, Germany; ²Inserm U936, Department of Medical Informatics, Rennes 1 University, Rennes, France; ³University at Buffalo, Buffalo NY, USA; ⁴AMC-University of Amsterdam, Amsterdam, The Netherlands; ⁵National University of Singapore, Singapore; ⁶Stanford University, Stanford, USA; ⁷University of Aveiro, Aveiro, Portugal; ⁸University of Haifa, Haifa, Israel; ⁹University of Manchester, Manchester, United Kingdom; ¹⁰Medical University of Graz, Graz, Austria

With these comments on the paper “Biomedical Ontologies: Toward scientific debate”, written by Victor Maojo et al. [1], Methods of Information in Medicine wants to stimulate a discussion on advantages and challenges of biomedical ontologies. An international group of experts have been invited by the editor of Methods to comment on this paper. Each of the invited commentators forms one section of this paper.

1. Comment by M. Brochhausen

Biomedical Ontologies and the Scientific Method

The paper Biomedical Ontologies: Toward Scientific Debate [1] seeks to show that current work on biomedical ontologies is deviating from good scientific practice by borrowing insights and methods from “philosophy, linguistic and other non-sciences”. In what follows I will argue that its authors hereby apply a double standard. More precisely, I will show that, when we apply to the arguments in their own paper the same critical approach that its authors apply to recent research in biomedical ontologies, then these arguments will be shown to collapse. Indeed this paper itself falls so far short of meeting the standards of good scientific practice that it is not actually a scientific paper at all.

There is agreement that there are certain principles of good scientific practice which are the same for any domain of research, including the humanities [2]. These principles include the acknowledgement of the existing body of work, the requirement that an investigation should be coherently conceived and executed, and that there should be no flaws in methodology [3]. I will show how [1] falls short in respect of all three of these principles.

Acknowledgement of the Existing Body of Work

The authors focus their attentions on the Basic Formal Ontology (BFO) [4, 5], an upper level ontology to support integration of scientific research data through the creation of domain-specific BFO extensions. BFO is now being used for this purpose by some 75 projects almost all of them in the life sciences field, including medicine [6].

One key criticism directed by the authors towards BFO and realism deviates from BFO by downward population is the fact that it is labeled as “Aristotelian” by its developers. The argument seems to be that, because Aristotle’s works on natural phenomena do not meet the criteria of empirical science as we understand them today, and because some of the basic principles of BFO’s ontological realism have – above all in the treatment of definitions – Aristotelian roots, it follows that BFO, too, must be rejected. The level of argument here is, I believe, quite astonishing. The author’s reasoning is analogous to an argument to the effect that because, say, Newton’s works on alchemy do not meet the criteria of empirical science as we understand them today, then the principles of Newtonian mechanics must likewise be rejected.

In writing about BFO and realism Maojo et al. seem to be unaware of the contemporary debate amongst philosophers between the two camps of realists and nominalists as concerns the status of universals and their role in science [7–9]. They also seem to be unaware of the degree to which BFO’s categorization of reality differs from Aristotle’s categorization of the world given in his Categories [10]. The creators and proponents of BFO are proponents of realist doctrine that is standardly referred to in the contemporary literature as “Aristotelian” simply to mark the fact that it rests on a view that there are mind-independent categories instantiated by the particulars in reality. To this, the proponents of BFO add a number of views which have nothing to do with Aristotle, for instance a view to the effect that the way to gain knowledge about categories is to test corresponding general hypotheses by following the scientific method in carrying...
out experiments on particulars. As is clear from the experience of the multiple groups of life science researchers who are employing BFO in their work, BFO is this way promoting empirical science. Awareness of the state of the art in ontology is not to be found in the paper. The authors seem unaware of contemporary debates in philosophy; they seem to be unaware of standard usage of philosophical terms; and they seem to be unaware of the way in which BFO, and similar ontologies such as DOLCE, are in fact being used by scientists.

Maojo et al. fall short also as concerns the history of Aristotelian philosophy. They talk of the “middle ages” as an era marked by a lack of scientific understanding due to the dominance of Aristotelian philosophy—giving as evidence a reference to a popular science publication. This is not, unfortunately, the state of the art as concerns research in the history of science, but merely the expression of a common prejudice from popular culture.

This claim is refuted both in its evaluation of the standard of medieval science and the role Aristotle played in the latter. It is common knowledge that during the early medieval period only a fraction of Aristotle’s works were known in the West through translations made by Boethius and through Commentaries of Porphyry, while Aristotle’s major works in metaphysics, logic and natural philosophy were rediscovered in the 12th and 13th centuries. As has been amply documented in the recent literature on scholastic philosophy [11–14], this rediscovery spawned tremendous developments in empirical science, preparing the way for the massive development of experimental science in the 16th century in the work of, for instance, Francis Bacon, Leibniz and Locke.

Poorly Conceived and Executed Study

The authors start out from the premise that there has been a “recent introduction of philosophical assumptions” into ontology development. A thorough analysis of the development in the applied ontology field however shows clearly that this premise is false. From the very start computational ontologies came with explicit or implicit assumptions which were ontological in the philosophical sense. It was, after all, from philosophy—more specifically from the writings of Quine—that the term “ontology” was taken over by AI researchers such as John McCarthy and Patrick Hayes in the 1980s [15, 16]. In the case of upper ontologies they are explicitly representing entities that are described as “philosophical”, by IEEE’s SUO [17]. Philosophical assumptions of the ontological sort are made every day, by every scientist, computer engineer, actually by every person in the multiple contexts of life. These fundamental assumptions are referred to, following Quine, under the heading of “ontological commitment”. According to [18] ontological commitments specify the objects and relations “talked about by agents” and existing in a specified domain. [19] stresses that commitment to what exists and what is important to each given system is unavoidable in any kind of representation technique. We know of no way to carry out ontology-based research in a coherent fashion without addressing issues of this sort.

Methodological Flaws

Besides the negligent attitude to the state of the debate on realism and on the history of Aristotle’s influence, a further problem is that our authors do not take the necessary care when it comes to reading and interpreting the literature on BFO itself.

One example is their misunderstanding of the distinction between “independent entities” and “dependent entities”, as for example when the authors ask how cells or viruses can “be entirely independent entities”, protesting that this “does not capture the interrelationship essential for biophysical function and life”, BFO’s distinction between dependent and independent entities, however, as has been explained at length in multiple places [20, 21], relates specifically to the type of dependence that obtains between quality-like entities such as color, weight, or volume and the thing-like entities (such as organisms or planets) in which the former inheres. This is, in the standard terminology of philosophical ontology, a matter of existential dependence: the existence of each particular entity of the first sort presupposes the existence of a supporting entity of the second sort [4]. The color of the banana on my table exists only because the banana itself exists; your headache exists only because there is some head in which this headache inheres. The fact that some entities are existentially dependent on other entities in this precise sense does not conflict at all with the fact that, for example, cells and organisms derive from other cells and organisms, and that various sorts of causal dependence obtain between them. The goal of an ontology such as BFO is precisely to provide a system of logical definitions for the terms used in scientific discourse. To make this possible, some terms will be needed—precisely the terms of an upper level ontology—which are so general that they do not form part of the specific vocabularies of the particular sciences. A critic can object if, for example, the definitions provided are ambiguous or logically incoherent, or if the explanations are difficult to understand. He cannot, however, object because a given term is defined in a certain way, merely because he would prefer to have it be defined in some other way.

Our authors claim further that reversible and irreversible processes are “not accounted for” [1] in BFO. It is true that neither “reversible process” nor “irreversible process” is represented in the BFO hierarchy. But the reason for this is simply that “process” is one of the leaf nodes in BFO [5]. Users of BFO who feel they need “reversible process” or “irreversible process” for representing their domain are free to add these terms as children of BFO’s “process”. The two mentioned types of processes can then very easily be accounted for in an ontology using BFO. Presenting the absence of child-terms in this way as a flaw in BFO reflects once again a lack of understanding on the part of the critic, and shows only that he has not made the effort to understand BFO’s aim to serve as a small upper level ontology to support integration of scientific data precisely by employing no terms which are peculiar to the special sciences. This strategy, shared also by DOLCE, is designed to foster harmonization of existing ontologies [22]. The strategy is thus by no means philosophically motivated; rather, it follows as a necessary consequence from the perspective of data
integration between expanding numbers of user groups [23].

The underlying concern expressed in [1] is that the developers of biomedical ontologies, especially those related to BFO and the OBO Foundry, have put themselves in danger of losing touch with empirical scientific research. To see if this is so, I have provided a list of disciplines represented in the consortium memberships of a representative sample of ontologies, including one OBO Foundry ontology, namely the Gene Ontology, and three OBO Foundry candidate ontologies, namely the Ontology of General Medical Science (OGMS), the Ontology for Biomedical Investigations (OBI), and Infectious Disease Ontology (IDO) [24]. The goal is to see to what degree these ontologies are being developed by domain experts, computer scientists or philosophers:

- Biology, animal medicine and medicine 42.16 %
- Computer science, biomedical informatics 36.14 %
- Biomedical ontology 8.43 %
- Philosophy 1.2 %
- No affiliation given 12.04 %

The list shows that the cross-disciplinary work on these OBO Foundry ontologies is carried out preeminently by people in close touch with cutting edge scientific research in their domains of medicine and biology. It also illustrates the appropriately small, but still important, representation of philosophers in this enterprise.

The OBO Foundry documentation stresses that “the Foundry is an attempt to apply the scientific method to the task of ontology development, and thus it accepts that no resource will ever exist in a form that cannot be further improved” [25]. Thus, the Foundry does indeed welcome criticism of the sort provided in [1]; it would indeed be troubling, and would require immediate rectifying steps, if [1] had indeed demonstrated that the Foundry approach is in some way in conflict with the scientific method. But we hope that it is by now clear that this case has not been made.

Strikingly, the way in which the authors seek to raise doubts against BFO and the Foundry is itself not based on sound scientific practice. Rather, the authors seem to have the goal of raising doubts and uncertainties on the part of developers and users of BFO-based or OBO Foundry ontologies and they realize this goal, consciously or not, by making false claims on a multitude of issues and neglecting entirely the state of the art in ontology research. [1] is indeed convincing evidence of the necessity of the OBO Foundry approach. It shows that to make progress in ontology-driven biomedical informatics it takes a collaborative effort resting on a division of labor between domain experts familiar with the state of the art in a number of disciplines – natural scientists, informaticians, database engineers, together with a small, a very small, portion of trained philosophers.

2. Comment by A. Burgun

Formalizing Uncertainty is the Achilles Heel of Ontologies

While uncertainty is widely present in medicine, the efforts have focused on organizing biomedical hierarchies based on ontological properties of biomedical categories. For example:

- **Primary Glioblastoma** is a Glioblastoma,
- **Primary Glioblastoma Has Primary Anatomic Site Nervous System.**

Although these relations have proven useful in many tasks such as data aggregation, they do not carry all our knowledge about Glioblastoma, which is also made of many possible ‘semiological’ associations between Glioblastoma and other entities, e.g. Glioblastoma May Have Abnormal Cell Fibrillary Neoplastic Astrocyte

- **Glioblastoma** May Have Finding Seizure
- **Glioblastoma** May Have Finding Unfavorable Clinical Outcome
- **Glioblastoma** May Have Molecular Abnormality PTEN Gene Inactivation
- **Glioblastoma** May Have Cytogenetic Abnormality del(10q23)

The definition of Glioblastoma in the NCI Thesaurus illustrates features that are typical or possible, for a given disease, but are not necessary, i.e., these features often occur when the cancer is present but some instances of this cancer may not have this feature. ‘Ontological’ properties correspond to necessary and sufficient conditions for defining class membership, and they are inherited by its subclasses. Conversely, ‘semiological’ features are not necessarily present in all instances of the class, and some of the subclasses may explicitly exclude this finding. Solutions must be proposed to handle cases where a class exhibits a particular feature as a typical feature, but some of its subclasses do not have this feature; or, conversely, some subclasses may have this feature necessarily for all instances, rather than have it as a typical feature. For example, Noy and co-authors [26] proposed to have two properties, Disease_has_hX1 and Disease_May_Have_hX1, where Disease_May_Have_hX1 subsumes Disease_Has_hX1 [26].

Computable Ontologies Can Be Seen as Artefacts for Classifying Instances

Victor Maojo and co-authors [1] have used an ontology of geometrical shapes to detect shapes that satisfy certain properties. As well as the ontology of geometrical shapes may be usefully linked to systems designed to detect specific geometrical structures, biomedical ontologies are a source of computable domain knowledge that can be exploited for classifying patient cases. Although it draws on the principles of biological sciences, medicine is a practical body of knowledge brought to bear on the understanding and treatment of particular cases. To be able to classify patient data under a disease class d, the criteria that are used to determine whether a case is an instance of d must be represented. The development of knowledge-based applications in the clinical domain requires that both (i) relations such as Diabetes mellitus is a Disorder of endocrine pancreas and (ii) WHO diagnostic criteria for diabetes: fasting plasma glucose ≥7.0 mmol/l (formerly 7.8) or 2-h plasma glucose ≥11.1 mmol/l be represented in the ontology. The former
ensure logical consistency of the class hierarchy while the latter may be used for classifying real cases. Those criteria would play a major role in ensuring that different researchers or clinicians annotate data in a consistent manner. They would also be useful for semantic matching, for example for comparing eligibility criteria in clinical trials and patient data in electronic health records.

Maiojo et al. found that Modelling Data under the BFO Produced Unexpectedly Long and Complicated Expressions to Refer to the Data Themselves. From their Point of View, what was gained in terms of generality was lost in terms of usability.

In our own experience, using the assumptions specified in BFO made it difficult to represent qualities for perdurants. Perdurants are entities that “occur in time” (e.g., atrial fibrillation, cardiac rhythm) in which endurants (e.g., atrium, heart) participate. Intuitively, duration of atrial fibrillation, and heart rate are parameters or properties we can perceive or measure – they take values – that are associated with things (heart), or phenomena, or events (atrial fibrillation). In ontology parlance, they are dependent entities that are inherent in either Endurants or Perdurants. Practically, it is not clear how one can represent medical information such as duration of atrial fibrillation, and heart rate under BFO. As qualities and quality values represent most of medical data in electronic health records, modelling qualities under upper level ontologies is expected to be easy and intuitive. Moreover, choosing one upper-level ontology, e.g., DOLCE (resp. BFO) rather than BFO (resp. DOLCE), to model a biomedical ontology should not jeopardize its alignment with ontologies based on BFO (resp. DOLCE).

3. Comment by W. Ceusters

Biomedical Ontologies: Toward Sound Debate

In “Biomedical Ontologies: Toward Scientific Debate” [1] Maiojo et al. discuss a number of aspects of what they call “computational biomedical ontologies”, specifically those that are claimed to have been developed using “classical philosophical assumptions”, Aristotelian ones in particular, as applied in the Basic Formal Ontology (BFO) [27], the Relation Ontology [28], and the biomedical ontologies that are designed following the principles of Ontological Realism [4] with the goal to become accepted in the Open Biomedical Ontologies Foundry [25]. They do so on the basis of 1) an analysis of disputes in the literature, 2) the identification of phenomena such as emergence and vision, and 3) the difficulties they experienced to apply such principles in their own ontology development efforts. They finally propose a traditional concept-based ontology of shapes. Their final conclusion, in a nutshell, is that 1) ontologies following philosophical principles cannot be tested empirically, 2) many issues remain open, and 3) further scientific debate is needed.

I studied Maiojo et al.’s paper from a logical and discursive perspective, thus assessing the soundness of the arguments and the correctness of the premises, in this case the data the work is based on. Most arguments used were found to be fallacious according to Good’s Classification of Fallacious Arguments and Interpretations [30]. The proposed morphospatial ontology was analyzed using the methodology proposed in Ontological Realism, which is based on the idea that the most effective way to ensure mutual consistency of ontologies over time is to view ontologies as representations of the reality that is described by science [29]. Their ontology, despite being tiny, is unfortunately flawed as well. I summarize my findings in terms of a number of recommendations which I believe to be crucial for future scientific debate.

Be Specific

Maiojo et. al. do raise the question what ontologies are, but do not answer it. They correctly refer to the fundamentally distinct meanings in philosophy and knowledge representation but they are not precise about what sort(s) of representational artifacts they themselves denote with the term “ontology” nor provide any information on whether the word is used to denote the same sorts of artifacts by all the papers they cite. Do they accept that an ontology is whatever somebody calls an “ontology” or everything that is expressed in OWL or some other formal language? In case of the latter, any collection of mathematical formulae would constitute an ontology. Ontological Realism is more specific: representational units should denote entities that to our best scientific understanding exist in reality and the structure of an ontology should mimic the structure of that reality. These principles do not exclude, as is often misunderstood [31, 32], representational units denoting “happy thoughts” – to use a term from Feynman’s quote found in [1] on which I will elaborate later – but they require that anything which is a happy thought would indeed explicitly be classified as a happy thought. Concept-based ontologies do not make that distinction since depending on the definition used, concepts belong to the realm of (in biomedicine ‘clinical’) ideas or units of knowledge, thus assertions about reality. Since Ontological Realism also embraces the principle of fallibility, representations need of course to be updated with the advance of science. This explicit distinction between what (according to science is believed to be) the case and what is hypothetical constitutes a nice complement to mathematical formulæ where such a distinction between what variables denote is not made and which, for instance, leads to the ‘conceptual problems in quantum mechanics’ related to the interpretation of the corresponding formulæ [33]. And one surely remembers the hypothesis about the existence of the planet Vulcan on the basis of mathematical formulæ and empirical validation thereof that were found more than half a century later to be inadequate [34]. Was it here ontology that caused the interpretation of
mathematical formulae to be erroneous, or mathematics that caused an unjustified shift in ontology?

I also failed to understand what Maojo et al. mean by 'science' because a few times the words 'scientist' and 'philosopher' are used in contradistinction which under at least one interpretation is quite disturbing. Perhaps Maojo et al. use 'science' in a very narrow sense involving only those activities which follow the scientific method, a reflection I make on the basis of the utmost importance they attach to empirical validation, experimentation and prediction. However, as scholars in the Philosophy of Science confirm – and Maojo et al. do include indeed a discussion on this debate – that method has many caveats too, a conclusion which Maojo et al. however do not express.

Be Consistent

Equally important is consistent use of a vocabulary once introduced. In the morphospatial ontology that Maojo et al. propose in their paper and in the appendices provided as supplementary data both principles are violated. First, they are unclear about what their ontology is a representation of. They present a taxonomy of what they call 'shapes' but they do not give a clear and unambiguous definition of what the entities they call 'shapes' exactly are. Are they purely mathematical or geometrical constructs, thus – again quoting Feynman – 'happy thoughts which we are free to make as we wish'? After all, such constructs are not more than idealized abstractions of the shapes that real entities exhibit only by approximation. Are they what under the Basic Formal Ontology perspective would be qualities, thus dependent continuants? Or are Maojo et al. just introducing a terminology for independent continuants which they wish to categorize, via old-style classifications, on the basis of the extent to which the shape-qualities of these independent continuants correspond to one or more of the mathematical/geometrical constructs? Maojo et al. indicated to find the distinction between independent and dependent continuants not to be very useful, but in this case, it might have been very helpful to concretize their thoughts.

An example of the inconsistent use of a term that is partially defined is that of 'hexagon': On the one hand, it is classified in their taxonomy a few levels under 2-D geometrical shapes with genus 0, which, as they clarify, means there are no holes in them. On the other hand, Figure 8 of Appendix 2 has as title, I quote: 'Hexagon [note: this hexagon has one hole].' Clearly, the 'gaps that computational ontologies displayed in their early days', as Maojo et al. phrase it, are still present today.

Explain Technical Terms Sufficiently/ Ask for Explanation when in Doubt

Providing definitions for essential terms, and then using these terms consistently, is also not enough to come to mutual understanding of what opposing parties are arguing for – or against – in a debate. Debating parties must be able to understand each other's language and agree to introduce additional terms where existing terms that are perfectly clear to one side, lead to too much confusion on the other side. Here, I confess, Ontological Realism has still a long way to go since it does take a long time before novices, however skilled and competent, in the field become totally familiar with the methodology. The difficulties for scholars in other fields than philosophy to understand the basic distinctions become already apparent with a question such as 'How can cells or viruses be entirely independent entities, even within a controlled laboratory environment?'. Maojo et al. make it thus clear that they do not understand what ontological dependence means, which has nothing to do with the inability of organisms (an example of independent continuants) to survive in absence of certain other independent continuants: the dependence of human beings on oxygen is not ontological dependence.

But despite such misunderstandings, certain reflections made by Maojo et al. are quite astonishing. For instance, so they continue, 'Viewing them as independent entities may serve as a practical simplification for philosophical, cognitive or even computational purposes, but does not capture the interrelationships essential for biological function and life'. No, of course not, I would say; what serious ontologist would make such a claim? A skilled ontologist would resort to appropriate relationships defined following the principles explained in the Relation Ontology [28] to assert that a corresponding essential relation in reality holds.

Also the statement that 'the distinction between continuants and occurrences does not account for the contrast between reversible and irreversible processes in biology, chemistry, computation, or quantum mechanics', is an odd one to make in a 'scientific' discussion. This is like saying: the distinction between males and females does not account for the difference between nuns and housewives. I believe that for assertions to be qualified as 'scientific' they should not just be true – Maojo et al. make undeniably a lot of true statements, at least under a common sense meaning of 'true' – but they should also not be blatantly trivial.

Argue Soundly

Maojo et al. use a debatable rhetoric by appealing, for instance, a lot to authority, citing from Nobel prize winners and eminent scientists such as Feynman: 'whatever we are allowed to imagine in science must be consistent with everything else we know; that the electric fields and the waves we talk about are not just some happy thoughts which we are free to make as we wish, but ideas which must be consistent with all the laws of physics we know. We can't allow ourselves to seriously imagine things which are obviously in contradiction to the known laws of nature'.

But eminent scientist or not, what does Feynman mean here by 'know' and 'known'? Would 'believe we know' and 'believe to be known' not be more appropriate? For the ancient eminent scientists, it was a 'known law of nature' that the sun orbited around the earth. If Copernicus would have followed Feynman's advice uncritically, we might still believe so. And for all we know, we might, full of happy thoughts, be hang-
Don’t Equate Popularity with Quality

An extensive literature study covering ontologies of various kinds leads Maojo et al. to claim the ‘strong agreement about the advantages of computational ontologies’ on the basis of positive experiences reported. Unfortunately, in [39], thus confirming suggestions made in [40], empirical support is provided that the reliability of findings published in the scientific literature decreases with the popularity of the research field, and that in hot research fields one can expect to find some positive finding for almost any claim, while this is not the case in research fields with little competition [41]. It is therefore argued in [39], and I agree wholeheartedly, that ‘for increasing the reliability of research it is essential to assess the negative effects of popularity and develop approaches to diminish these effects’. That is for sure applicable to ontologies.

Separate Usability from Quality

Ontological Realism is a methodology for ontologies that want to be maximally reusable, at the risk of being less usable. It favors representational correctness, to our best scientific understanding, over short-cuts. It is thus not paradoxically at all, as stated by Maojo et al., that in their experience ‘using the kind of philosophical assumptions currently specified for computational ontologies from the OBO Foundry has considerably complicated some of our work’, specifically not in light of the problems of understanding the methodology as I documented earlier in this paper. We do not however follow Goguen who according to Maojo et al. ‘takes an even more critical position, considering that philosophical ontology is a step backwards in computer science, embracing extreme forms of realism and reductionism’. Reductionism is exactly what Ontological Realism avoids, and avoiding reductionism is exactly what makes Ontological Realism more complex than what computer scientists believe to be necessary for their purposes. I would rather argue the opposite: computer science is a step backwards for the development of high-quality ontologies, embracing extreme forms of (model-theoretic) semantics that cut computational constructs loose from reality. This does not seem to be understood well enough by many enthusiast computational ontologists who seem to believe that it is sufficient to have something stated in OWL, for that something to be a high-quality representation. The NCI Thesaurus is an example of the contrary as recently has been demonstrated once more [42].

Conclusion

Although Maojo et al. claim to have analyzed various aspects of current computational biomedical ontologies, ‘philosophical’ ontologies in particular, the bulk of their work was a rather one-sided analysis of the literature thereby cherry-picking references and citations that favor an anti-philosophical position. Based upon an analysis of their discourse, their argumentation is not convincing, and often flawed.

Ironically, they are saved by Aristotle: logic tells us that it is possible for a valid argument to have false premises and a true conclusion. That is what Maojo et al. achieved in part: further discussion is indeed needed, although not exclusively under the very narrow view of ‘science’ that they seem to entertain, and not with mere pragmatism in mind. If Bertrand Russell was wrong in his analysis about the observational skills of Aristotle, he might also be wrong in his view that the victory of pragmatism is greatest in those countries where science is most advanced and that the question is if, in the end, this is not to the detriment of science and of the scientific spirit as scientists become slaves of research projects geared towards technical mastery [43]. That would be bad news for the utilitarian scientists.

Further debate cannot be fruitful, however, if not first, or at least in parallel, additional steps are taken. One such step is the development of a vocabulary that can be used by all parties to express exactly what each party means, and in such a way that it is clearly understood by all other parties irrespective of whether they agree with statements made in terms of that vocabulary. Attempts in this direction are already made upon initiative of Gunnar Klein, former Chairman of CEN/TC251 [44].

Next, there need to be agreement about an ontology of ontologies that clearly distinguishes the various sorts of artifacts that currently are denoted by this term as well as about the distinct quality criteria instances of each of these various sorts of artifacts can, but not necessarily should, adhere to. Concept-based systems developed for a specific application may limit representational adequacy to what is relevant for that application. But it would be confusing and even wrong to assign such a small-scale initiative the same ontological status as a reference ontology that tries to describe an entire domain, independent of any purpose for which it might be used.

With all this in place, it will become possible to address all issues raised more appropriately. And it will make reviewers of ontology papers better equipped to identify and evaluate the evolution of high quality work, how controversial, preliminary or

---

4 I found the quote on Google Books, but did not read the entire volume.
non-mainstream it might be, and eliminate inferior work.

4. Comment by A. Hasman

Introduction

Maojo et al. [1] raised a number of questions about the role of computational ontologies in scientific research. In this contribution I present my opinion about some of these questions.

Many papers nowadays deal in one way or another with ontologies. The term itself has already been used in philosophy for about four centuries but in informatics it appeared only some 20 years ago. According to Gruber [18], who introduced the term in our field, an ontology is a specification of a conceptualization of (some part of) reality. It is a description of the concepts and relationships that can exist for an agent or a community of agents. Gruber has designed ontologies for the purpose of enabling knowledge sharing and reuse.

A computational ontology contains classes, links between classes and properties of these classes, described in a uniform and standardized way. Ontologies support the semantic interoperability of heterogeneous information systems. And because the concepts described in computational ontologies are standardized, computers are able to reason with them.

Ontologies are especially important for knowledge management purposes, as already indicated by Gruber. Guideline systems for example can only efficiently communicate with health information systems when they use the same terminology for the classes, relations and properties. Expert systems can work much more efficiently and effectively if they make use of ontologies and inheritance. Maojo et al. [1] however want to discuss the role of computational ontologies in scientific research instead of in knowledge management applications. This role has not been discussed extensively until now.

Discussion of Some Roles of Ontologies in Scientific Research

Maojo et al. review various aspects of biomedical ontologies. One issue concerns the philosophical foundation of computational applications of ontologies. In response to the critique of some classical ontologists that early versions of computational ontologies show deficiencies in how they relate to classical formalisms for representing reality, the OBO Foundry adopted an approach based on classical Aristotelian assumptions. However, according to Maojo et al. some of the philosophical distinctions now present in e.g. BFO might not be suitable for scientific purposes. BFO distinguishes for example independent and dependent entities. According to the BFO a virus and a cell are independent entities, a fact which is difficult to accept. Also they demonstrate that the distinction between continuants and occurrences does not account for the contrast between reversible and irreversible processes in biology, chemistry, etc.

Apparently there are differences between a classical philosophical approach and a computer science approach towards defining an ontology. In my opinion we should be pragmatic and create computational ontologies with which we can describe our knowledge of a domain incrementally starting with the distinctions that are necessary for the current purposes. Having too many distinctions in a computational ontology may make the ontology less efficient and usable and may easily lead to errors. Ontologies should be developed in an iterative way: their use will determine which distinctions need to be modified, deleted or created. In other words, learning by doing.

Ontologies can be considered as descriptions of scientific theories. They could be used to explain certain phenomena on the basis of the incorporated knowledge. From such an ontology also new knowledge may be inferred. This knowledge was then already present in an implicit way. I agree with the authors that on the basis of an ontology alone no really new theories can formulated. A scientist, however, can deduce hypotheses using an ontology in for example text mining. An hypothesis is not generated on the basis of the ontology alone, its terms or concepts are used to find new relations in the reported results of empirical studies. Swanson [45, 46] et al. discovered new knowledge in this way (e.g. the treatment of Raynaud’s disease by fish oil). They used a word-based approach, linking entities by intermediate words that appeared frequently in the contexts of both entities. Van Haagen et al. [47] used a concept-based approach for inferring protein-protein interactions: different terms denoting the same concept (i.e. synonyms) were mapped to a single concept identifier. In both approaches ontologies are used as a tool to mine knowledge but that knowledge was not implicitly present in the ontologies itself.

Conclusion

The suggestion of Maojo et al. to discuss the advantages and challenges presented by biomedical ontologies from a scientific perspective is a good one. As stated earlier I think that ontologies can be best developed and used for knowledge management. Knowledge management in itself is a very important area and this area can be used to learn about the possibilities and limitations of developed ontologies. Also ontologies can be used as a tool to discover new knowledge from text bases or databases. I do not believe that ontologies alone will enable us to arrive at new theories.

5. Comment by T. Y. Leong

Toward an Adaptive Ontology in Biomedicine

Real-world problems in health care and biomedicine are often complex, uncertain, and changing. Many limitations of the current ontologies used in health and biomedical information systems and solutions are due their limited or missing capacity in dealing with change. A biomedical ontology that effectively supports information management, analytics, and decision making should be grounded on an operational conceptualization framework that would facilitate reasoning and discovery in a non-
stationary environment. This commentary aims at exploring some feasible directions and sketching some potential solutions. Based on the observations by Maojo and colleagues [1] in this issue, we examine some alternate approaches to developing ontology systems, and postulate a set of desiderata for an adaptive ontology for biomedicine.

Introduction

In the article “Biomedical Ontologies: Toward Scientific Debate” [1] in this issue, Maojo and colleagues highlight significant limitations of biomedical ontologies in particular, and computational ontologies in general. Some of the main challenges lie in the limited or missing capacity of the existing ontologies in supporting representing and reasoning with change. Based on the observations by Maojo and colleagues, we identify some requirements for computational ontologies to effectively deal with change, especially in supporting knowledge discovery and decision making, and postulate a set of desiderata for an adaptive ontology for biomedicine.

Ontologies are structured organizations of concepts related to, in our case, one or more biomedical domains. As pointed out by Maojo and colleagues, the conceptualization of an ontology is usually pre-determined and manually designed, with or without philosophical or semantic interpretation, and targeted at recording or describing natural phenomena in the domain. The descriptive nature of existing ontologies assumes that the underlying phenomena are stationary—whether it is a concept of a class, individual, incident, or process, the description is defined in terms of a stationary domain environment. Changes are either not incorporated, or only included as part of the description in the ontological definition. For example, the change by “breaking down” is part of the definition of the metabolic process in Gene Ontology (GO)4:

"L-lysine catabolic process to glutarate, by acetylation"

Meaning: The chemical reactions and pathways resulting in the breakdown of L-lysine into other compounds, including glutarate, by acetylation.”

Real world problems, however, are non-stationary. The underlying world model evolves, and new perceptions are often different from expectations. In other words, the world model as the problem solver understands it is always changing. There are many common examples of changes that impact reasoning and discovery in biomedicine. As pointed out by Maojo and colleagues, for example, a pre-defined or pre-established ontology has difficulty in supporting reasoning with emerging patterns of biological processes. At a higher level, for example, a new infectious disease may be emerging in a rural population, but epistemological studies did not pick it up automatically as the number of incidences is small. How can a surveillance program with a fixed ontology be made to examine the relevant patterns in the population data in a statistically significant manner? On the other hand, a hospital may extend or adapt the electronic medical record systems to a hospital in another community, where the demographics and the disease types and the treatment strategies may be different. How to ensure that the existing knowledge, including ontologies and processes, can be reused without having to spend an enormous amount of money and effort? Even when such transfers are internal to a single hospital or hospital system, the patient groups, diseases, and test and treatment types may still be different or change over time. How can a health information system with a fixed ontology transform or evolve “gracefully”?

Challenges of a Changing World

To effectively support representing and reasoning in a changing world or domain, and to discover new scientific knowledge, facilitate paradigm shift, or theory formation, the underlying ontology must be adaptive to the changing environment. The main research questions to be addressed include:

● What are the characteristics of an adaptive ontology?
● What types of operations or tasks should be supported?
● When and how would the ontology adapt to the changing world?

Maojo and colleagues point out that most scientific discoveries and advancements are not made by ontologists. The statement is, however, partially accurate as ontology only makes up part of the requirements for scientific discovery and theory formation; the other part—reasoning and learning new knowledge that would lead to new insights—is not addressed in the existing approaches to ontological design and use.

Existing ontologies are mainly designed to support direct information retrieval, e.g., in health information systems, or deductive reasoning for explanation and prediction, e.g., in specialized diagnostic systems. For these functions, design and organization of the ontology focus on a comprehensive indexing system that allows efficient and accurate incorporation and retrieval of the relevant concepts. Even with these targeted functions, most existing ontologies are restricted by the deterministic, static, and manual designs. While tremendous effort has been invested and some progress made in the past few decades, many biomedical systems built on such ontologies are fragile, hard to scale up, hard to maintain, hard to reuse, and often do not support interoperability.

Some improvements may be made to the descriptive and Aristotelian approach to ontology designs to address the challenges related to change. We believe, however, that a paradigm shift in the ontology design and development process itself is needed to achieve the main objectives of an adaptive ontology. In particular, an ontology must be able to change with new subjective or objective information. The process of incorporating new information and the ability to reconfigure representation structure should be seamlessly integrated. In other words, a biomedical information system that can cost-effectively support integration, organization, and analysis of biomedical information to arrive at actionable recommendations must allow learning to change (in representation), and changing (in representation) to learn. Traditionally, representing change and reasoning with a changed representation are considered sep-

---

A Cognitive Approach to Representing Change

According to Maojo and colleagues, a major disadvantage of adopting Aristotelian semantics in ontological design is the rigidity and complexity in using the resulting conceptual constructs in reasoning tasks. Confusion and hence inaccuracy are introduced when many assumptions are made and extensive simplification is needed in the applications. As an alternate to adopting a philosophical semantics, a cognitive science perspective in ontology design may shed some new lights and add some new capabilities. The dual process theory of mind has been widely regarded as a viable cognitive and computational model of representation and reasoning [48]. The theory separates reasoning and learning activities into two separate processes, often working in parallel – one of them fast, unconscious, automatic, contextualized and effortless, and the other slow, conscious, controlled, general, and effortful. Any effective mental representation or conceptualization system should therefore be able to support these two processes.

In her seminal work The Origin of Concepts [49], Carey advocates that the conceptual systems in humans for understanding a particular domain also operate in two phases. The first phase includes concepts or representations that are more elementary, basic, and concrete, akin to the structural and analogical relations to the domain, and close to the current descriptive and Aristotelian ontologies. The second phase is more abstract and socially constructed, in other words, formed by context-dependent sensing and assimilation of information with the original concepts. Contradiction or inconsistency detection is the engine that underlies conceptual change.

While such a theory may need to be further validated, it provides a basis for developing an adaptive ontology that can be grounded in an operational framework and supports reasoning and leaning in a systematic, sound manner. Most of the debates and disagreements in ontological designs stem from the need for a solid semantic foundation, and the lack of consensus about the classification, perspective, and naming of the concepts involved. A cognitive approach to ontological design would avoid the ad-hoc definitions that arise in a purely computational approach, while still accommodating the philosophical underpinnings that one may wish to attach to the semantics.

A Probabilistic Approach to Reasoning with Change

Maojo and colleagues point out another major limitation of current biomedical ontologies is the inability to manage uncertainty, both in representation of and reasoning with the knowledge captured. In recent years, probabilistic graphical networks, e.g., Bayesian networks, dynamic Bayesian networks, and their decision-theoretic variants, have been applied in many biomedical applications where the domain parameters – the concepts and the conditional dependence among them, may change with new observations, either in terms of subjective knowledge or statistical data [50]. Probabilistic graphical networks, based on statistical and philosophical (logical) definitions, support a wide range of explanation, predictive, and decision-analytic inferences in face of uncertainty, incomplete data, and noise. Bayesian networks, in particular, has been featured extensively in recent cognitive and computational modeling work [51], exhibiting many desirable properties and functions that help exploration and explanation of various cognitive processes [52]. While Bayesian networks have not traditionally been directly related to ontological frameworks, the domain concepts can be considered as propositional variables in the networks. Relational or first-order interpretations have recently been incorporating into the general class of probabilistic graphical networks [53].

Probabilistic graphical networks are a unique class of representation and reasoning frameworks that can support both learning to change and changing to learn.

On the one hand, learning algorithms can be deployed to update both the model structure, in terms of the links between the nodes in the network, and the probability distribution, in terms of the conditional probabilities among the nodes. On the other hand, in view of incomplete, unexpected, or unexpected information, updating of the model through learning can be restricted to specific components or partial instantiations of the networks, with respect to certain contexts or local structures. Other dynamical systems, including neural networks, exhibit some similar properties, but the ontological or conceptual definitions, and hence the philosophical or cognitive underpinnings, are less clear.

Toward an Adaptive Ontology

While Maojo and colleagues discuss the main limitations of existing biomedical ontologies and highlight some important characteristics and targeted functions, it is unclear what an “ideal” or even just appropriate design approach for the next generation of biomedical ontologies should be. With the vast number of information sources, including text, signals, images, categorical and numerical data, and the wide range of functions and operations that the ontologies aim at supporting, such an “ideal” or prototypical approach may never emerge. Instead, a number of different approaches to building the next generation biomedical ontologies may work together, depending on the tasks at hand. An example is the morphospatial ontology that combines qualitative and quantitative information as proposed in Maojo and colleagues’ article. Another example may be a combination of the cognitive and probabilistic approaches described earlier.

Based on the observations by Maojo and colleagues, we postulate a set of design considerations for the next generation biomedical ontologies, some of which may be more relevant for supporting different classes of functions and operations in different information management, analytic, and decision support systems.

Specifically, an adaptive ontology in biomedicine should:

1. Support conceptualization of the domain, including classes, individuals, properties, incidences, events, processes
2. Support both qualitative and quantitative representations and descriptions
3. Support uncertain representation of the facts and the inference processes
4. Support reasoning and learning in multiple perspectives along different dimensions, e.g., is-a, part-of, inverse-of, etc.
5. Support reasoning and learning at multilevel of abstractions – first with basic, elemental and concrete definitions, then with emerging concepts upon some “triggering effects.”
6. Support reasoning and learning in different contexts and over time.
7. Support definition of the triggering effects that would lead to conceptual change – e.g., by manual definition, or by outlier detection, or by “surprise” detection.
8. Support formation of abstract concepts, with possibly different names or identifiers in different contexts.
9. Support mapping of naming conflicts among different ontologies and defining similarity measures
10. Support reconfiguration of conceptual structures, where some concepts may become obsolete, or their relations may be revised.

The current state of the art in ontology development, knowledge representation, and inference and learning frameworks in Artificial Intelligence systems can potentially address items 1 to 6; advancements in these general computational technologies will eventually be incorporated into biomedical applications. Most of the issues involved in items 7 to 10, however, have not been extensively explored.

Maojo and colleagues have articulated some important limitations related to the core area of biomedical ontologies, and initiated an important debate that would advance scientific pursuits in this area, with potential far and wide reaching impact in the field of biomedical informatics. Based on their observations, we have identified some specific challenges and explored some possible directions of investigation. The changing world poses a set of rich motivations, in inspiration, and testbed for new ideas for a set of representation, inference, and learning techniques that can effective deal with changes. We are hopeful to see more future research efforts in this direction.

6. Comment by M. Musen

Biomedical Ontologies: Dogmatism Considered Harmful

The use of standard ontologies is exploding in both the life sciences and in clinical medicine. The stakes are high, because ontologies are providing key infrastructure for electronic health records, for the exchange of clinical data, for the labeling and indexing of biological experimental results, for the analysis of high-throughput data, and for a host of other key applications [54]. Historically, most biomedical ontologies have emerged directly from the communities that are most in need of data interchange and analysis, not in any kind of top-down manner from standards organizations or professional societies [55]. Not surprisingly, many of these “grass roots” ontologies have been created without careful consideration of existing knowledge-representation standards or best practices, and they have exposed serious tensions between biologists who may be seeking pragmatic solutions and computer scientists arguing for coherent, decidable logical inference [56].

The Open Biomedical Ontologies (OBO) Library7 represents a collection of ontologies originally assembled by the developers of the Gene Ontology [57] to offer an online archive of the ontologies that they and their colleagues have developed. The OBO Library now includes several ontologies that aspire to be admitted to the OBO Foundry [25], a collection of reference ontologies that adhere to certain design criteria articulated by the Foundry’s curators. To date, only five ontologies, in addition to the Gene Ontology, have graduated from members of the OBO Library to the OBO Foundry.

In this issue of Methods of Information in Medicine, Maojo and his colleagues [1] argue for a scientific debate regarding the principal requirement for admission to the OBO Foundry: an ontological commitment to Aristotelian realism. The authors also suggest that our community openly discuss a related OBO Foundry requirement: adherence to the Basic Formal Ontology (BFO) – the upper-level ontology developed and promoted by philosopher Barry Smith and colleagues [59]. Smith is a principal driver of the OBO Foundry.

Although Maojo and his colleagues seem to suggest that all (or most) biomedical ontologies are moving in the direction of the OBO Foundry, it is important to recognize that inclusion in the OBO Foundry is not a universal objective in the biomedical community. The vast majority of ontologies available through the BioPortal archive1 of the United States National Center for Biomedical Ontology [59], for example, are not included in the OBO Library and have not been developed by authors who see incorporation within the OBO Foundry as a necessary or even desirable goal. The OBO Foundry may receive enormous attention in the literature, but the Foundry has not been uniformly accepted by the biomedical community by any means.

Most of the OBO Foundry criteria have obvious face validity. No serious ontology developer will argue with the need for good documentation, standard syntax, and careful versioning. The controversy surrounding the OBO Foundry stems from, among other things, the required commitment to philosophical realism and to embracement of the BFO.

For some time, there has been almost no open discussion in the scientific community of the appropriateness of requiring a specific philosophical stance for an ontology to be considered for inclusion in the OBO Foundry and to be blessed as “good.” Few biomedical investigators have any knowledge of the metaphysical nuances that keep philosophers up at night; few have any professed view on universals in reality. Few biomedical investigators even have an inkling that Aristotelian realism

---

might not be a mainstream philosophical position. As most life scientists know very little of metaphysics and are unaware of the competing philosophical perspectives on reality, there has seemed to be no need for debate — until now.

The past year has seen the initiation of serious scientific discussion of the role that philosophy should play in the development of computational ontologies for scientific purposes. Although it is surprising how long it has taken for the conversation to commence, it is important to the scientific community that these philosophical issues finally are on the table for examination, discussion, and empirical evaluation.

Gary Merrill, a philosopher by training, now at North Carolina State University and for many years head of the Semantic Technologies Group at GlaxoSmithKline, made the first salvo in the summer of 2010 with a paper entitled, "Ontological realism: Methodology or misdirection" [60]. Merrill argued that Aristotelian realism could not be viewed as a "best practice" within the philosophy community and that realism had nothing specific to offer the computational ontology community as an engineering method. Like Maojo and his colleagues [1], Merrill argued that adherence to Aristotelian principles is actually inappropriate for the modeling of science. Smith and Ceusters [29] replied to Merrill with an exhaustive analysis of their position. Their primary defense of realism as an ontology-development methodology, however, was to point to the large number of developers who have embraced their particular philosophical stance when developing contributions to the OBO Library. Merrill offered a stinging rebuttal [61], arguing that the realist perspective articulated by Smith and Ceusters is fundamentally "incoherent". As Merrill, Smith, and Ceusters exchanged barbs in their journal articles, hundreds of postings on the subject of realism in biomedical ontologies appeared on the OBO-discuss mailing list8.

Meanwhile, Phillip Lord and Robert Stevens [31] presented several scenarios where the realist perspective breaks down when attempting to capture scientific phenomena. They argued that science often deals with inherently psychological phenomena, such as color, that do not correspond to unique universals in reality and hence may be construed in alternative ways. They pointed to numbers and equations — essential components of scientific discourse — that are purely conceptual entities with no referent in an objective reality "out there".

Adding to the dialog was Dumontier and Hoehndorf [32], who criticized the realist approach in general and the BFO in particular as being inappropriate for the representation of scientific entities. Like Lord and Stevens [31], these authors emphasized the centrality of conceptual entities in science. Dumontier and Hoehndorf also pointed to the importance of being able to discuss entities that scientists impute to exist on the basis of data (e.g., certain subatomic particles), but whose existence in reality has never been confirmed.

Maojo and his colleagues should be commended for adding to this discussion, and for bringing it to the biomedicalinformatics literature. The scientific debate has been ongoing for some time, but the paper by Maojo and his co-authors makes it clear that these issues need to be articulated more broadly and discussed with far more openness.

Despite all the claims that have been made about the role of alternative methodologies for ontology engineering, it is striking that there is scant empirical evidence to help developers identify what works and what does not. The advocates for the OBO Foundry claim that the realist perspective enables ontology authors to create more principled, more interoperative ontologies [29]. Other ontology builders, such as Maojo and his colleagues, complain that, "using the kind of philosophical assumptions currently specified for computational ontologies from the OBO Foundry has considerably complicated some of our work" [1]. The arguments about philosophy and ontology engineering remain entirely philosophical.

Whereas many elements of the realist position and its relationship to modern science will need to remain the subject of philosophical debate, there is much that we actually can measure and evaluate. We can assess in controlled experiments the complexity of using the BFO to model conceptual entities. We can record the time it takes ontology developers who use the BFO to structure ontologies of standardized phenomena, and we can contrast their efforts with those of developers who use alternative upper ontologies — or who use no upper ontology at all. We can compare the error rate and search time required for users to locate entities in ontologies that have been constructed using alternative approaches, and we can try to measure the incremental benefits that may be derived from the different methods. Maojo and colleagues do a good job of showing that philosophical ontology is not the same as computational ontology; members of the biomedical informatics community can follow their lead and begin to evaluate the engineering of computational ontologies with the same kinds of metrics that they might apply to the engineering of software.

Given the centrality of biomedical ontologies to so much of our work, it is time for our community to move beyond philosophical debates and to begin to collect real data. There is too much at stake not to subject ontology engineering to the same kinds of experimental analyses that we apply routinely to everything else that we do in science. It is remarkable that, while biomedical science remains a highly empirical discipline, many of us are all too quick to abandon empiricism altogether in deciding how to build the biomedical ontologies on which the future of biomedical science depends.

Acknowledgments
This work has been supported in part by the National Center for Biomedical Ontology, supported by grant U54 HG004028 from the United States National Institutes of Health.

7. Comment by J. L. Oliveira
In their paper, Maojo and colleagues [1] present an interesting review of ontologies, starting from their philosophical foundations up to current applications and challenges in biomedicine. The importance of

8 https://lists.sourceforge.net/lists/listinfo/obo-discuss
philosophical grounding is unquestionable and many questions raised by the authors may be addressed within these roots. However, as an engineer, and sometimes ontology user, my opinion about the theme is certainly biased by a perspective that is more technical than philosophical. I also share the idea that the philosophical versus computational debate might not exist if a different name was chosen to coin this concept in computer science.

During the last decades, computer and information scientists have been using ontological artifacts, such as data models and object-oriented methods, to develop programs and databases. However, with the unprecedented increase of biomedical data and information successively available in scientific papers, more powerful methods were necessary, in order to organize and manage all the emergent entities and miscellaneous relationships. It was time for ontologies to come out from their traditional philosophy landscape and start to be used in daily applications.

Following the success of the Gene Ontology (GO), ontologies have been increasingly used in the biomedical domain due to their capacity to organize knowledge into formal and machine-readable methods. In fact, they are beginning to proliferate in step with accumulated biomedical data. While ontologies are good instruments for data annotation, their proliferation may lead to the same kind of problems we are already seeing with data, and that was the main driving force for the increasing importance of ontologies: the risk of being overwhelmed with so many dissimilar, complex, and sometimes incomplete solutions. The NCBO BioPortal, for instance, had already 255 ontologies and 4,809,250 terms by the end of February.

To deal with the diversity of ontologies, we can use integration techniques. This may be done through cumulative gathering of complementary but autonomous contributions, such as UMLS, or through top-down approaches, such as the OBO Foundry, which ensures consistency between integrated ontologies but may fail to reuse legacy ontologies. The OBO Foundry principles are mostly concerned with the establishment of best practices related to engineering and pragmatism. But even the integration of ontologies can lead to somewhat antagonistic requirements – an engineer may be more interested in the ontology as a metadata technology, while an ontologist will be more concerned with how the ontology accurately mirrors the entities and relationships of a specific knowledge domain. Therefore, the ontologist is someone that needs to be sensitive to ontological semantics so that he can formalize knowledge representation, according to the complexities of the real world. From a practical perspective, despite the importance of formalisms, they often create a burden when modeling data. The authors stress their experience with the use of the Basic Formal Ontology (BFO), which typically produced unexpectedly long and complicated expressions to refer to the data themselves.

Besides the increase of biomedical ontologies, another problem hindering their wide adoption is related to the heterogeneity and isolation of several groups: some (ontologists and computer scientists) spend their career developing ontologies and tools, while other researchers (biologists and physicians) know how to use ontologies. The diversity of availabilities of philosophical formalism, but much more from some degree of uncertainty and the unknown associated with any scientific discovery process. However, this is a major problem since the poor quality of an ontology may result in inaccuracies in the applications it supports. The diversity of available ontologies also creates a problem when an ontology has to be chosen for a given application. Besides the accuracy of statements in ontologies, other factors such as popularity [62] and governance [63] also need to be taken into account. Another challenging goal for ontologies is their integration in current scientific literature. Semantic publishing, i.e. embedding ontologies in manuscripts, will help to support efficiently text mining and information extraction [64], and will intensify the common practice of strategic reading [65].

While ontology completeness is a major goal, seeking the ‘perfect’ solution may lead to unnecessarily complex models which, despite being computationally tractable, will be difficult to handle by humans. From a computational perspective, this dilemma is similar to the one software engineers face when planning and developing applications: there is no recipe for the perfect solution, just a set of good rules and practices that lead the way to the best result. Solution diversity is desirable even if it often leads to increased complexity and confusion. Lessons can be learnt from several fields, namely computers and network communications – the most accepted solutions are typically the simplest and not necessarily the most complete and powerful. From a technical point of view, pragmatism typically succeeds, while good and bad ontology solutions will continue to appear.

8. Comment by M. Peleg

Maojo and co-authors [1] discuss computational biomedical ontologies from different perspectives, including their philosophical foundations, and in particular issues of emergence, the ability (or lack of ability) of ontologies to generate new discoveries and theories, and to evolve. Pointing to a new direction of research in biomedical ontologies, the authors raise open questions related to visual reasoning and spatial ontologies.

Emerging novel biological knowledge and its integration with current scientific views is not always easily represented in ontologies, even when the new knowledge is deduced by scientists and is not expected to be discovered purely by reasoning with ontologies. Maojo and co-authors note that different ontologies that cover the same domain but view it from different perspectives (e.g., of molecular biologists, physiologists, and clinicians) may address different biological levels using a different level of granularity and different context. Indeed, the different communities of researchers and practitioners have different points of view that change over time, as the scientific knowledge about a domain grows. For example, the cell cycle stages,
such as interphase and its phases, G1 (first gap), S (synthesis), and G2 (second gap), and the mitosis stage with its phases, prophase, prometaphase, metaphase, anaphase, and telophase, were defined based on cytological observations made by examining dividing cells through a microscope. For example, at the onset of prophase, chromatin condenses together into a chromosome and at metaphase the centromeres of the chromosomes converge along the metaphase plate. However, ontologies need to consistently represent the dynamic knowledge about high-level biological processes in the context of their molecular-level sub-processes [66]. For example, they should be able to relate the interphase G1 phase to processes of protein synthesis, organelles production, and cell growth. Furthermore, ontologies can also be used to provide even more detailed molecular definitions. Such definitions could relate the cell cycle checkpoint between phase G1 and phase S, in which the cell duplicates its DNA, to the specific molecular-level processes and their participating molecules (e.g., cell cycle kinases CDK4/6-cyclin D and CDK2-cyclin E) and molecular complexes (for details see http://www.biocarta.com/pathfiles/h_g1pathway.asp). Peleg et al. [67] provide another example where a bio-ontology was used to specify the process of protein translation in a high-level representation that is further nested into a detailed specification of tRNA mutations that cause abnormal protein translation resulting in clinical disease phenotypes.

The authors reiterate Noble’s criticism of the Gene Ontology (GO), which excludes physiology and most aspects of evolutionary biology and therefore cannot adequately address modeling phenomena across different biological levels. It is important to note, however, that bio-ontologies have different purposes; GO [68] is meant to provide a common vocabulary used to annotate gene products according to the biological processes at which they participate, the function that they exhibit, and their cellular localization. In this way, GO can facilitate data aggregation (e.g., retrieval of all mouse genes that participate in lipid metabolism) [69] and data integration [70], where data from different biological data bases is retrieved and joined based on common GO annotations. Other bio-ontologies have been used to formulate queries for accessing and integrating data from multiple biological databases [71] or for modeling and simulating biological processes in the context of their participating molecules and their resulting disease phenotypes [66, 67, 72].

Another interesting question raised by the authors is whether computational ontologies could do more than provide a means of representing descriptive knowledge of a field in a computationally useful way, or in other words could they go beyond “what you put in is what you get out”. The answer to this question, in my view, is affirmative. Existing ontologies already provide not just descriptions but can be the basis for planning care pathways [73], providing explanations for decision-support recommendations [74], and enabling simulation-based prediction of system behavior over time [72]. Furthermore, by representing libraries that contain definitions of different types of exceptions and exception-handling mechanisms [75], even unanticipated exceptions that occur during execution of clinical-guideline based decision-support systems could be handled by generic exception-handlers. Similarly, using description-logics-based ontologies, default behavior could be defined and executed for controlling the access to electronic medical records when specific organizational policies have not been defined [76].

Moving on to the second topic addressed by the paper by Maojo and co-authors [1], the focus on visual reasoning and spatial ontologies is an important emerging need for biomedical ontologies. Geometric shapes are important means for biologists for conceptualizing and comprehending biomolecular structures as well as for understanding how structure relates to function and to process. Biologists are used to creating diagrams to convey how biomolecular complexes act in biological processes, e.g. the replication complex that replicates DNA, the protein translation mechanisms, etc. Biologists find it more intuitive to comprehend such diagrams in contrast with more abstract representations of biological process models. To make computational biological process models more intuitive for biologists, tools such as Cell Illustrator [77] allow biological scientists (users) to intuitively model and simulate complex dynamic interactions and processes in biopathways comprising of hundreds of entities within and among cells, using intuitive icons that correspond to classes in the biological pathway ontologies named Cell System Ontology.

Maojo and co-authors note that when considering shapes, fundamental concepts that arise are connectedness and adjacency. These distinctions are also important when considering cellular organelles; in the Bio-Workflow ontology [66], cellular components are classified into membranes (e.g., nuclear membrane), spaces (e.g., cytoplasm, lysosome lumen), membrane bound compartments (e.g., membrane-bound organelles, such as nucleus, chloroplast), non-membrane-bound cellular compartments (such as cytosome and nucleolus), and cellular structures (e.g., centriole).

There are many additional topics that are important in the context of biomedical ontologies and have not been covered by this paper. One of those topics is the use of computational methods to discover knowledge in data [78–80] and in biomedical publications [81, 82] in order to populate ontologies. Such methods make the task of ontology construction more scalable.

9. Comment by A. Rector

Overview

This paper [1] opens with the statement: “The role of computational ontologies in scientific research, as opposed to knowledge management applications, has not been extensively discussed. We aim to stimulate further discussion on the advantages and challenges presented by biomedical ontologies from a scientific perspective.” We take this to mean that they wish to explore whether ontologies provide new means of formulating scientific theories and/or whether they provide new insights or new means of analysis, analogous to other new developments in mathematics – e.g. Bayesian networks, chaos theory, etc.
The authors sketch a taxonomy of “shapes and forms” near the end of the paper as an exemplar. However, they never make clear what added value they expect to get from the use of the “ontology”. What could they not do with existing approaches? What problems are they trying to solve?

Critique of existing ontologies

Despite the stated objective, much of the paper is taken up with critiques of existing work on ontologies, much but not all, directed at Smith’s Basic Formal Ontology (BFO) [83] and the related OBO Foundry [84] family of ontologies. Smith’s approach, which has had wide effect on the Gene Ontology and related OBO Foundry ontologies, has been extensively critiqued by Lord and Stevens [31] and by Merrill [60], the critique here is marred by two omissions:

1. The ontologies that are being critiqued were largely created for the purpose of mark-up of databases – i.e. for “knowledge management applications”, so the relevance of the critique to their use “in scientific research” is questionable.

2. The authors never make clear what they mean by a “computational ontology”. Given the many different sorts of artifact-termed “ontologies”, many referenced in this paper, the lack of a clear statement of what the authors mean by “ontology” leads to issues being conflated rather than differentiated. What is the scope of “ontology”? Where are its boundaries? What should be expected of one? What not?

For example, in their discussion of “emergence” (Section 3), the authors criticise the Gene Ontology (GO) because it does not have the “higher level insight needed to establish direct causal relations between genes and the phenotype”… and “cannot adequate address modelling phenomena across different biological levels”. They seem almost to criticise GO because it does not predict emergence. Whichever is intended, the natural response is: “Why would anyone expect it to?” GO was created as a catalogue of genes at the level of our then understanding. To do so, most genes were characterised at a micro level. The authors go on to state: “Without introducing quantitative approaches and the capacity to computationally interact with such empirical models and simulation tools at various biological levels, current computational ontologies cannot fully address the functionalities involved in biological processes and their ecologies.” True, but an answer to the wrong question. The right question is whether current (or potential) biological ontologies are useful as part of the ensemble of tools needed to address these functionalities; i.e. whether and how including ontologies enhances those tools’ effectiveness. Are their scientific theories that we can formulate or test using ontologies that we could not otherwise? Or could do so only with greater difficulty?

Ontologies, Knowledge Bases, Inference Engines, Indexes and Catalogs

This brings us up squarely against the question of what an ontology is or should be. This question has become more difficult in the past few years. With the authors I wonder “…what might have happened differently if Gruber had proposed a different name than (computational) ‘ontologies’ like, for instance, ‘conceptual constructs’ or ‘semantic structures’.

…that inference was computationally tractable. This allowed “classifiers” to organise complex hierarchies of definitions that would have been difficult or impossible by manual means. Indeed they might better have been termed “definition logics”.

In compensation, it is possible to define their semantics precisely using standard model theoretic methods from logic, and the subset of first order logic was chosen so that inference was computationally tractable. This allowed “classifiers” to organise complex hierarchies of definitions that would have been difficult or impossible by manual means. Indeed they might better have been termed “definition logics”.

All of these representations and notions had in common that they were symbolic and not quantitative. Some included numerical comparisons, but few used calculations except as “attached procedures.”

Meanwhile, the name “ontology” and the obvious analogous with philosophical ontology attracted the interest of various philosophers, most notably Smith who applied their skills and doctrines to clarify ambiguities but also to introduce dogmas from their own philosophical background.
Given this lengthy digression, we can sketch a coarse grained scheme that distinguishes various sorts of knowledge artifacts that have sometimes been called ontologies. Using such distinctions, perhaps refined, it might be possible to make Maojo et al.’s questions sufficiently precise to be tested and their critiques sufficiently focused to avoid confusion.

1. **Definitional and necessary (aka “universal”) knowledge.** Roughly speaking what Kant (since the authors refer to him) would have called “analytic” knowledge; what logicians using model-theoretic semantics would refer to as being true in all possible models (sometimes phrased: “all possible worlds”). Typically represented in description logics / OWL.

2. **Generic contingent knowledge.** The background knowledge that is assumed in addition to the definitional and necessary knowledge, e.g. that “most gram positive bacteria are sensitive to penicillin”, that “that renal failure is a complication of diabetes”, “that appendicitis is a cause of right lower quadrant pain”, etc. This is knowledge that is not universally true, but that we do not expect to have to communicate explicitly when exchanging data and which we find more convenient to express declaratively rather than procedurally. Typically represented in frames, semantic nets, or RDF networks.

3. **Conventional data and facts.** Additional background information based on legislation, standards, etc. e.g. which drugs are licensed for use in which countries for treatment of which diseases; the list of SI units, etc. Typically represented in databases, but also possibly in frames or RDF.

4. **Annotations, meta-knowledge, and higher order knowledge about the kinds themselves** – e.g. data on authorship, use of symbols within the artifact, and higher order statements such as that organisms always have children and parents of the same kind. Representation varies by system.

5. **Patterns and pathways:** e.g. Plans, patterns, clinical protocols, etc. Currently a topic of investigation, but lacking any consensus representation in biomedicine.

6. **Quantitative knowledge of complex systems,** e.g. the functioning of networks of cell components involving complex feedback loops. Typically represented as networks of partial differential equations, complex Bayesian networks, complex simulations, etc.

7. **Procedural knowledge:** Rules and procedures for using the above to come to conclusions, extrinsic to any built mechanisms, e.g. query and rule languages, other computations, etc. sometimes extending to “attached procedures”.

Orthogonal to this categorization, representations of the domain must be distinguished from specifications of data and message structures for holding information about the domain – e.g. typically UML, entity relationship diagrams, XML Schema, etc. Specifications of data structures may be informed by knowledge, but they are not themselves knowledge. Maojo et al. do not fall into the common trap of confusing knowledge representation and data structures, but it is so widespread as to be worth mentioning here for completeness.

**Rephrasing the Question**

Given this background, we can come back to rephrase the question: Which if any of the above should be referred to as ontologies? Which if any can contribute to new “scientific research”? What added value does each bring?

**Definitional and necessary knowledge** (1) comes closest to the traditional philosophical use of the word “ontology.” Conveniently, it also is a reasonably close match to what can be represented naturally in description logics. The philosopher’s notion of “universal” and the logician’s notion of “all possible models” are closely related. Unfortunately, the logician is also concerned with what is computationally tractable, and so excludes many notions that the philosopher might wish to include, e.g. higher order representations, cyclical representations, etc. However, we must be careful not to identify “ontology” with what can, or cannot, be expressed in a specific language such as OWL.

We can then ask if separating off the **definitional and necessary knowledge** is useful in representing the other forms of knowledge. One answer is pragmatic. If partitioning off this form of knowledge makes certain computations of the cognitive task easier, then it is effective. If those effects are useful in performing scientific research, then it is useful scientifically.

There is at least one published study of such a discovery with respect to the relation of the structure and function of proteins analysed using the classification methods [90] and numerous others in progressb In these cases, discovery results from precise definitions of both structure and function linked to computational methods for analyzing large knowledge bases against the definitions.

For **generic contingent knowledge and conventional facts** (2, 3), few would argue the usefulness of a declarative representation. However, how best to combine definitional knowledge (1) with contingent knowledge (2) remains, surprisingly, an open question.

**Annotations and higher-order knowledge** (4) are often the primary purpose of ontologies such as the GO, but probably an adjunct to doing “scientific research.”

Patterns and pathways (5) constitute a special problem of great importance. Pathways, clinical protocols, and other patterns tend to contain a rich array of cycles, alternatives, feedback loops, sequences, more complex temporal phenomena, etc. These clearly go beyond definitions, generic knowledge and facts (1–3). Using the word “ontology” for them seems to stretch it to its breaking point. Rather, we would prefer to consider whether and how ontologies and generic representations of contingent knowledge can contribute to the representation of patterns and pathways and facilitating new means of manipulating them. Despite various efforts, e.g. [91, 92], this also remains an open question.

**Complex systems and procedural knowledge** (6–7) are again clearly beyond the scope of ontology, per se. The best mechanisms for integrating ontologies and such representations remain to be explored, with many proposals but no consensus.

---

b Personal communication, Robert Stevens, 2011.
Conclusion

In moving forward a scientific debate on the use biomedical ontologies in “scientific research”, we would first focus on distinguishing the various intellectual and computational artifacts sometimes labelled “ontologies” or “knowledge representations”, and then on the contributions that they can make to “scientific research” singly and in combination, and on how these contributions interact with the various tools – description logics/OWL, frames, RDF, other logics – used to implement them. Without precision about what we are debating, the debate will not be productive.

Having made such distinctions, we can return to the issues raised in the paper, for example, whether, and if so how, ontologies and knowledge representations in any of the senses 1–4 are useful in addressing questions of organisation and emergence or how they might interact with other mechanisms such as 6–7 (or others) to do so. What are the requirements? Of specific artifacts, do they have the necessary constructs and relations? For example, can they deal with collective effects? Do they have relations that cross scales? etc. With respect to the shapes ontology, we can ask whether it is simply a set of definitions (1) or something more (2–4) and how it is proposed to be implemented. What are the new issues that it raises? Does it alter fundamentally the categorization above? What are the alternatives? Are there functions that are performed better with it than without it? In achieving their goals, new methods will undoubtedly be needed. Can they indicate what is novel more clearly?

If we understand the authors’ intent in the introduction correctly, the evidence for success or failure of any of the methods proposed lies in the empirical demonstration of their ability to add value – that they proposed lies in the empirical demonstration of success or failure of any of the methods performed correctly, the evidence for what is novel more clearly?

Expressiveness vs. Performance

What ontologies aren’t

Ontologies Are Not Knowledge Representations

In philosophy, ontology (the study of being) is opposed to epistemology (the study of knowledge), and we could argue that ontologies are not at all about representing knowledge. Recognizing the broader meaning of “knowledge” in computer science I would be less strict emphasizing that ontologies are about a very special kind of knowledge, viz. about only what is considered, believed, and defined to be universally true about a segment of reality in a given scientific community at a given time. As Alan Rector puts it [96], there are “very few interesting items of knowledge that are truly ontological in this strict sense”: Or according to Woods, “an ontology is no more than a conceptual coat rack” [97] on which more complex and scientifically more “interesting” information and knowledge is held. We would therefore not do justice to computational ontologies if we expected that they should contain probabilistic associations, implement Bayesian models or permit the deduction of emergent properties of a compound from the properties of its components.

Ontologies Are Not Thesauri

Thesauri, such as the MeSH, the UMLS Metathesaurus, or the NCI thesaurus aggregate lexical entities according to informal semantic relations such as (quasi-) synonymy, hyponymy and hypernymy. As such they address important use cases, particularly in information retrieval. For example, the MeSH thesaurus is a highly optimized tool that has become indispensable for literature retrieval in MEDLINE. However, thesauri have no formal semantics and no consolidated ontological commitment. Ontology-motivated critique to thesauri is therefore meaningless, unless thesauri are dubbed as ontologies or distributed in ontology languages, such as the NCI thesaurus [42].

10. Comment by S. Schulz

Introduction

Is a Swiss Army knife a knife? No, it isn’t. It is a complex tool which has a knife as one of its parts. Who considers a Swiss Army knife a typical knife will be disappointed by a normal kitchen knife as it lacks scissors, screwdrivers, bottle openers etc. Maojo et al. [1], throughout their paper, seem to expect computational ontologies to be Swiss Army knives for knowledge representation and, not surprisingly, they are dissatisfied with the features and limitations of existing biomedical ontologies. Nevertheless, their ample notion of ontologies is backed by various references which, above all, demonstrate how the term “ontology” is used in multiple and partly contradicting senses, partly motivated by Gruber’s famous but unfortunate definition of ontologies as “explicit specifications of conceptualizations” [93]. Here advocate, for sake of clarity, a less flashy but more distinctive, modest, and comprehensible definition of formal (or computational) ontologies: Whereas Ontology – as a discipline – is the “study of what there is” [94], formal ontologies are “theories that attempt to give precise mathematical formulations of the properties and relations of certain entities” [95]. With this definition in mind I will pick out just a few aspects which have direct impact on the judicious use of computational ontologies and the role they have to play in the field of biomedical semantics and knowledge representation.

What Ontologies Aren’t
cause their meaning cannot be translated into axioms in current ontology languages. First order logics would be very appropriate to encode ontology axioms, but it is undecidable. Therefore, computable subsets have been proposed by the description logics [88] community, such as the recent W3C recommendation on the OWL 2 Web Ontology Language [98]. If ontology-enhanced applications depend on computationally effective reasoning such as satisfiability checking and classification, a price must be paid in terms of ontology expressiveness. For instance, SNOMED CT with about 300,000 classes uses the inexpressive EL++ language, which lacks important constructors such as value restrictions and negations. If the whole range of OWL-DL constructors is used, performance problems may already complicate the use of ontologies with just a few thousand classes, and for effective reasoning, simplification approaches [99] need to be put in practice.

**Ontologies and Terminologies**

Surprisingly, Maojo et al. did not address at all the important connection between computational ontologies and domain terminologies. Assertions about what is universally true for all instances of a natural kind, as introduced above, is not the only rationale for building computational ontologies. Equally important is their use for precisely defining the meaning of domain terms, as an important step toward terminology standardization [100]. For instance, “viral hepatitis”, as used by medical practitioners and researchers, delineates the class of all inflammatory diseases of the liver which are caused by some virus. Conversely, every individual hepatitis which is caused by some virus is classified under viral hepatitis. The meaning of terms which are used to label and classify objects in reality often depends on arbitrary (“flat”) boundaries, e.g. a threshold diameter to distinguish two different tumour stages, a temperature threshold which delineates the meaning of the term “fever”, or logical combinations of characteristics for authority-dependent concepts like “rheumatoid arthritis”.

Classification is a major rationale for automated reasoning, enabling interoperability between syntactically different encodings of the same thing. Legacy terminological systems such as SNOMED International did not offer solutions to state the equivalence between concurring expressions of different degrees of aggregation, such as in Table 1 [101].

Only the ontology-based successors SNOMED RT and CT [100] provided Aristotelian definitions like:

**Table 1** Semantically equivalent expressions for Acute Appendicitis in SNOMED International [101]

<table>
<thead>
<tr>
<th>SNOMED ID</th>
<th>Expression</th>
<th>SNOMED ID</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-A231</td>
<td>Acute Appendicitis, NOS</td>
<td>G-C006</td>
<td>Acute inflammation, NOS</td>
</tr>
<tr>
<td>T-59200</td>
<td></td>
<td>In</td>
<td>Appendix, NOS</td>
</tr>
</tbody>
</table>

For personal or educational use only. No other uses without permission. All rights reserved.
even if, on the level of the ontology specification, the class Function is excluded from the domain of the relation hasParticipant, semantically wrong expressions can still be assembled in ontology editors like Protégé.

**Conclusion**

Computational ontologies are not Swiss Army knives for knowledge representation. But in knowledge-based systems they should be seen as indispensable components like the blade in a Swiss Army knife. As such they need to be well-sharpened, faultless and handled with caution by experienced users. Improvised fabrication makes them useless and improper use can cause considerable damage. User-friendly guidance and tooling is therefore of utmost importance for the breakthrough of computational ontologies as important foundations for intelligent systems in biomedicine.

**References**

Addresses of the Authors

Mathias Brochhausen
Saarland University
Institute of Formal Ontology and Medical Information Science
POB 151150
66041 Saarbrücken
Germany
E-mail: mathias.brochhausen@ifomis.uni-saarland.de

Anita Burgun
INSERM U936
Departement of Medical Informatics
Rennes 1 University
Rennes Cedex 35033
France
E-mail: anita.burgun-parenthoine@univ-rennes1.fr

Werner Ceusters
University at Buffalo
Ontology Research Group
Center of Excellence in Bioinformatics & Life Sciences
Department of Psychiatry
701 Ellicott Street
Buffalo NY 14203
USA
E-mail: ceusters@buffalo.edu

Arie Hasman
AMC-University of Amsterdam
Department of Medical Informatics
Meibergdreef 15
1105 AZ Amsterdam
The Netherlands
E-mail: a.hasman@amc.uva.nl

Tze Yun Leong
National University of Singapore
Medical Computing Laboratory
Department of Computing Science, School of Computing
Computing 1, 13 Computing Drive
Singapore 117417
Republic of Singapore
E-mail: leongty@comp.nus.edu.sg

Mark Musen
Stanford University
The National Center for Biomedical Ontology
Stanford Center for Biomedical Informatics Research
Medical School Office Building
251 Campus Drive
Stanford, California 94305–5479
USA
E-mail: musen@stanford.edu

José Luis Oliveira
University of Aveiro
DETI/IEETA
3810-193 Aveiro
Portugal
E-mail: jlo@ua.pt

Mor Peleg
University of Haifa
Department of Information Systems
Faculty of Social Sciences
Haifa, 31905
Israel
E-mail: morpeleg@is.haifa.ac.il

Alan Rector
School of Computer Science
University of Manchester
Oxford Road
Manchester M13 9Pa
United Kingdom
E-mail: rector@cs.manchester.ac.uk

Stefan Schulz
Medical University of Graz
Institute of Medical Informatics, Statistics and Documentation
Auenbruggerplatz 2
8036 Graz
Austria
E-mail: stefan.schulz@medunigraz.at