Biomedical Ontologies: Toward Sound Debate

Werner Ceusters

Ontology Research Group, Center of Excellence in Bioinformatics & Life Sciences, and Department of Psychiatry, University at Buffalo, Buffalo NY, USA

Corresponding author:

Werner Ceusters, MD

New York State Center of Excellence in Bioinformatics & Life Sciences

701 Ellicott Street

Buffalo NY, 14203

USA

ceusters@buffalo.edu

www.org.buffalo.edu/RTU/

Summary

BACKGROUND: In '*Biomedical Ontologies: toward scientific debate*' [1] Maojo et al. discuss various aspects of 'computational biomedical ontologies' developed using 'classical philosophical assumptions'. They conclude that (1) ontologies following philosophical principles cannot be tested empirically, (2) many issues remain open, and (3) further scientific debate is needed.

OBJECTIVE: The goal of the work presented here is to provide directions towards (1) the priorities for such discussion, (2) the way in which it should be entertained, and (3) what authors and reviewers of ontologies or papers thereof should pay attention to to avoid that ontologies become the 'neural *not* works' of the future.

METHODS: The paper was studied from a logical discourse perspective, assessing the validity of the underlying data (references and citations), the soundness of the arguments.

RESULTS: No argument could be found supporting their first claim, while the two other conclusions are correct, be it trivially true.

CONCLUSION: Further debate is required, but not exclusively under the narrow view of 'science' entertained by Maojo et al. Preconditions are (1) development of a vocabulary usable by all parties to express exactly what each party means, and that it is clearly understood by all other parties irrespective of whether they agree with statements made in terms of that vocabulary, and (2) agreement about (2a) an 'ontology of ontologies' that clearly distinguishes the various sorts of artifacts currently denoted by this term and (2b) the distinct quality criteria instances of these various sorts of artifacts can (not necessarily *should*) adhere to. This will make users and reviewers better equipped to identify and evaluate the evolution of high quality work, how controversial, preliminary or non-mainstream it might be.

Keywords: Ontology, Ontological Realism, Science, Philosophy of Science

1 Introduction

For several years I have been teaching at the University of Buffalo the one-credit course 'Solving crimes through Referent Tracking' under the form of a Discovery Seminar. These optional seminars offer freshmen and sophomores a small class experience providing them with the opportunity to engage around a thought-provoking, unfamiliar and challenging topic with the goal to improve skills in critical observation and thinking, and in oral and written expression. Semester after semester, the first student to whom I would ask 'Can you please introduce me?', fell in the trap, and started to introduce him- or herself, rather than introducing me. When I asked to tell me something about Referent Tracking¹, I would either get no answer, or a shy-hesitant 'tracking criminals?'. Pretending to wonder why they selected the course, not knowing anything about me or the topic, I would get the answer I anticipated: because they like the popular television series 'Law and Order' and 'CSI: Crime Scene Investigation'. I fulfilled for sure some of their expectations, not by analyzing CSI plots, but by making an old philosophical puzzle the central theme of the seminar: so it is claimed that in 1935 the medical doctor Carl Austin Weiss shot Louisiana Senator Huey Long inflicting a wound which caused Long to die thirty hours later. Weiss, on the other hand, received at least forty bullets from Long's bodyguards and died immediately. Thus assuming that Weiss killed Long, the question is: when? If one answers 'when *he shot him*', than one has to explain how somebody who is killed can live for another 36 hours. If one answers 'when Long died', then one has to explain how somebody who is already dead for 36 hours can still kill somebody. I invite the interested reader to solve the puzzle before reading further.

What has this story to do with a commentary on a paper arguing for scientific debate on biomedical ontologies, one may ask. Sadly, way more than I would like it to be the case, and this for many reasons, three of which involve issues that form the basis of my analysis about what should be debated and what is required to make such a debate possible and useful.

The first reason for bringing it up, of course, is the content of the discovery seminar in which students come to understand, painstakingly slowly, that the solution can only be found if they are able to distinguish the *ontological* aspects from the *terminological* ones. I published in this

¹ Referent Tracking is a paradigm for data annotation using Ontological Realism as basis. The reader should therefore understand that I am 'on the side' of those ontologists who by Maojo et al. are qualified as being 'philosophical'.

journal about this crucial distinction already in 1993, using a number of clinically more appealing variations of the killing paradox [2]. It was, by the way, the first time that I used the word 'ontology', not in the sense of a representational artifact but in the original philosophical meaning. This important difference, i.e. the distinction between *terminology* and *ontology*, is too often neglected or not well understood, neither by ontology developers or authors of papers about either topic. I mean here not only the distinction between terminologies and ontologies as representational artifacts, but also the distinction between the disciplines of terminology and ontology and ontology both of which involve activities of developing representational artifacts of a different sort.

The second reason is the effect of *popularity*. The first year, my seminar's title was 'Solving biomedical problems using Referent Tracking' and I had only one. The change in topic - solving crimes - caused the seminar to be fully booked short after opening. Popularity is clearly also an issue in 'ontology' in the sense of representational artifacts. Indeed, while Pubmed references only 53 papers in which the word 'ontology' is used between 1912 (year of the first appearance) and 1993, Maojo et. al. report to have found 4,557 documents while since then, after merely two months, another 155 have been added. In non-biomedical domains, the topics of ontology (in the computer science sense) and the very closely related semantic web are real hypes. As will be demonstrated later, this has, unfortunately, a negative impact on the quality of the scientific efforts.

The third reason, also related to quality, or lack thereof, involves the striking similarities between form, presentation and content of my students' first essays and representations about various aspects of the killing puzzle on the one hand and what I continue to read in the majority of papers describing ontology development or analysis that I am invited to review, or, worse, that are already published in scientific journals, on the other hand. In the latter case, the reviewers and editors are more to blame than the authors and one could wonder whether a scientific debate about the current peer-review process in this domain would not be urgently required as well.

2 Background

In 'Biomedical Ontologies: toward scientific debate' [1] Maojo 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) [3], the Relation Ontology [4], and the biomedical ontologies that are designed following the principles of Ontological Realism [5] with the goal to become accepted in the Open Biomedical Ontologies Foundry [6].

Maojo et al. start by questioning the scientific value of the top-level categories identified in BFO, in particular the continuant/occurrent and independent/dependent distinctions. They do so on the basis of (1) an analysis of disputes in the literature, (2) the identification of phenomena such as emergence which they claim to be unexplainable by resorting to such distinctions, and (3) the difficulties they experienced to apply these distinctions and related principles in their own ontology development efforts. They then discuss the relationship between ontologies and scientific theories, their application as knowledge representation in artificial intelligence, the problems principle-based ontologies seem to face in rapidly changing fields such as biological classification, and the perceived inability of such ontologies to deal with probabilities, uncertainty and various forms of reasoning. They finally propose a traditional concept-based ontology of shapes, thereby explaining terms such as '*classes*' and '*inheritance*', but unfortunately not what '*shapes*' are, nor how this effort relates to the analysis reported on in the rest of the paper.

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.

3 Objectives

The goal of the work presented is to provide some directions as to (1) the priorities of what needs to be discussed, (2) the way in which such a discussion should be entertained, and (3) what authors and evaluators of ontologies on the one hand and reviewers of papers about the former on the other hand should pay attention to in order to avoid that ontologies are facing the same destiny as the sort of representational artifacts that are commonly known as 'neural networks', but were quickly nick-named, not totally justified, 'neural <u>not</u>works'.

4 Methods

Maojo et al.'s paper was studied from a logical and discourse perspective, thus assessing the validity and soundness of the arguments and the correctness of the premises, in this case the data

the work is based on, including, to some extent, the citations and references provided. Good's *Classification of Fallacious Arguments and Interpretations* was used to guide the analysis [7]. A limitation of this effort to date is that, given the short time frame (6 weeks) allowed by the editor to prepare this commentary, it was not possible to verify the appropriate use of all 153 references.

The proposed morphospatial ontology was analyzed using the methodology proposed in Ontological Realism [5]. The goal of Ontological Realism is to foster consistency in the ways scientific results are described for purposes of more effective data-integration, thus counteracting the many tendencies leading to *ad hoc* and non-interoperable coding of data, and thus to the formation of data silos. Ontologies, when adequately designed, are ideal instruments to achieve this goal, but unfortunately, their very success has led to the creation of ever new ontologies, and thus has resurrected the very silo problems which ontologies were designed to counteract. This can only be solved by *minimize the number of ontologies* that are being constructed and at the same time *maximize their mutual consistency* what requires ontology developers to accept certain common *constraints on how they build their ontologies* in such a way that we *do not endanger the flexibility that is needed to keep pace with scientific advance and empirical research.* The realist methodology is thus 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* [5].

5 Results

Many of the arguments used by Maojo et al. can be classified into one or more of Good's fallacy classes [7]. The validity and soundness of some arguments could not be assessed because of the use of ambiguous language.

No sound argument could be found supporting their first claim according to which OBO Foundry style ontologies - whether or not in contrast to traditional concept-based ontologies - would not be empirically testable.

As a disclaimer, empirical testing of Maojo et al.'s paper was not possible due to time constraints. It is however completely feasible to subject the paper to a quantitative analysis by inviting a panel of evaluators with competency in the domain to classify all fallacies according to Good's classification, with or without resorting to agreement facilitating setups such as Delphi rounds, and applying appropriate statistical error estimation algorithms to the final result.

There is however no need to do so, since Maojo et al. arrive at two correct conclusions: that many issues remain open, and that further scientific debate is required. But, as classical logic tells us, it can from the correctness of a conclusion not be inferred that the premises are true also. In addition, these conclusions are also *trivially* true.

Finally, the proposed morphospatial ontology was found to be flawed from the perspective of Ontological Realism as its foundational element '*shape*' is not defined. The usefulness of the ontology under a traditional concept-based view is also questionable since Maojo et al. *themselves* misinterpret their own ontology when they explain in appendix 2 an example of its applicability.

6 Discussion²

6.1 What are we talking about?

Essential, so I believe, in a paper about ontologies and a request for scientific debate, is that the authors clarify what ontologies are and science is.

Maojo et. al. do raise the first question, but don't answer it. They use terms such as *'computational ontologies'*, *'biomedical ontologies'*, *'upper ontologies'*, *'philosophical ontologies'*, and so forth. They make some assertions about what computational ontologies *do*, how they are *built*, what they are *used* for, but not what they *are*. Perhaps they are right in not doing so. One of the first principles I teach my students is to be very careful with questions of the form *'what is/are X'*. For in order to answer such questions, one must know what the terms that would replace *'X' denote*. Nobody, except perhaps children or foreign language learners, will ask the question *'what is a bank?'* if it is not already made clear what the precise context is. In other words, before one is able to answer the question *'what is an ontology?'*, one must provide first an answer to the question *'what does the word "ontology" mean?'*. Only when this question has exactly one answer, i.e. if the word *'ontology'* would have only one meaning or, stated differently, if there is only one type of things that is denoted by the word, then the first question makes sense and a single answer can be provided. If there are different sorts of entities denoted by the term, then the first question stated in that specific way is non-sensical.

 $^{^2}$ In order to make this commentary more entertaining and educational, I deliberately introduced in this discussion a few elements from Aristotelian rhetorics in such a way that some arguments are examples of fallacies categorized by Good, however not to an extent that it would undermine my position. Arguing the opposite would be a fallacy of overgeneration.

Clearly, the word '*ontology*' does not denote just one sort of entity. Maojo et al. correctly refer to the fundamentally distinct meanings in philosophy and knowledge representation. But what is however cumbersome is that 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 what is expressed in OWL or some other formal language? In case of the latter, any collection of mathematical formulae would constitute an ontology. What is essential for ontologies following the principles of Ontological Realism is that their representational units denote entities that to our best scientific understanding exist in reality and that the structure of such an ontology mimics the structure of that reality (again to our best scientific understanding). This principle does not exclude, as is often misunderstood [8-9], representational units denoting 'happy thoughts' - to use a term from Feynman's quote found in [1] on which I will elaborate later - but requires 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 formulae 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 formulae [10]. And one surely remembers the hypothesis about the existence of the planet Vulcan on the basis of mathematical formulae that were found not earlier than more than half a century later to be inadequate [11]. 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 - in absence of a clear definition being given I can only make an educated guess - 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 Philosophy of Science confirm - and Maojo et al. do include indeed a discussion on the debate - that method has many caveats too, a conclusion which Maojo et al. however do not express.

6.2 Do we talk clearly ?

Clearly, introducing an unambiguous vocabulary alone is a necessary, but by far not a sufficient condition for high quality papers or high quality documentation of ontologies. Equally important, I believe, is consistent use of a vocabulary once introduced, and this, once more, specifically in the context of ontologies.

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 pure 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, as a third possibility, are Maojo et al. just introducing a terminology for independent continuants which they wish to categorize, as 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, Fig.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 to date.

6.3 Do we take too much for granted?

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? For one, a skilled ontologist would resort to appropriate relationships defined following the principles explained in the Relation Ontology [4] to assert that a corresponding essential relation in reality holds. And second, he would not use phrases of the sort 'viewing them as X'. Does viewing cells as 'cells' and viruses as 'viruses' capture the interrelationships essential for biological function and life? That would only be the case if 'cell' and 'virus' would be representational units in an ontology in which also these essential interrelationships are expressed. Similarly, such essential interrelationships are expressed for dependent and independent entities in high quality realism-based ontologies. Maojo et al. - so we can infer from their proposed morphospatial ontology - being strong believers in the value of inheritance for sure understand that when an ontology represents cells and viruses as independent continuants, these extra assertions come through inheritance to their disposal as well. They might not be useful for the purposes for which they use such an ontology, but they are for sure useful for researchers that have other goals in mind.

Also the statement that 'the distinction between continuants and occurrents 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 to be blatantly trivial.

Another one is '*Einstein's theory of relativity changed the concepts of light and time in a way that no pre-relativistic ontologies could have anticipated*'. Of course ontologies based on Ontological Realism have to change with the advance of science! Is that not also the case for other sorts of ontologies? But what ontologists that follow Ontological Realism have in their favor is that (1) they clearly understand that Einstein's theory didn't change anything about the nature of light and time - they continued to be the way they always have been - and (2) they can use a machinery that allows such changes in understanding, as well as changes in reality, to be tracked formally and coherently [12-14].

6.4 Do we argue correctly?

I always inform my students about the appropriate use of references and citations. The most important rules for referencing are (1) to have read the materials completely and not just the abstract, (2) to be sure to understand precisely what is argued for (or against), and (3) to verify whether claims made are not invalidated in later papers. In addition to this, citations should not be taken out of their context, and one should not judge a priori and without verification statements made by highly respected scholars of greater value than of unknown scientists, for doing so 'scientists may uncritically follow paths of investigation that are popularised in prestigious publications, neglecting novel ideas and truly independent investigative paths' [15]. As a rule of thumb, in domains as complex as knowledge management, terminology and ontology, one should count one hour per page to do a reasonable job, and perhaps the double if one is in the beginning of his career, or unfamiliar with the topic. This means that Maojo et. al. must have gone through a titanic job in light of the 153 references provided, many of which being books and of recent publication date.

References and citations are often used to build a case, but unfortunately, there are multiple ways to build bad arguments and student essays resulting from class assignments, thus work that is not

mentored or supervised, contain typically both formal, for instance logical inconsistencies, and informal fallacies such as hasty generalizations, appeals to authority, etc., some of which, such as the latter, are already committed when violating the reference and citation principles [7].

Maojo et al. do appeal a lot to authority, citing, for instance, from Noble price winners. Quite often, however, I feel lucky to be ignorant about awardees or impressive achievements and to have a failing memory for names: it avoids me taking for granted everything they state. So I confess, when reading Maojo et al., I had initially no clue who Feynman was. But the 'eminent scientist' qualification did make me frown when reading the quote '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'3. What does Feynman mean here by 'know' and 'known'? That what in terms of Ontological Realism would be phrased as being '*objectively the case*' or by others as '*the truth*' or 'facts'? If so, what made Feynman so sure about the infallibility of his knowing? Remember the examples given earlier about the interpretation of formulae and the pitfalls of the scientific method. Would 'believe we know' and 'believe to be known' not be more appropriate? For the old 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 hanging in the Matrix or be living, way less happy but evenly ignorant, in Dark City. Any first year student of Philosophy of Science or Epistemology would spot the problem with Feynman's quote. Why didn't he? If the quote 'Philosophy of science is about as useful to scientists as ornithology is to birds' (http://en.wikipedia.org/wiki/Philosophy_of_science) is rightfully attributed to him, I can see why.

Granted, with the last three sentences, one might argue that I am following an argumentation strategy which constitutes a fallacy of another sort: *argumentum ad hominem*. That would indeed be the case if I were abusing Feynman's first quote to discredit the rest of his work. But that of

³ Maojo et al. reference this quote erroneously as 'Feynman R. The **Feynam** Lectures on Physics, Vol. 2. Addison-Wesley; 1977.' [bold/underline emphasis added], a mistake I found also on class notes prepared by Ronald Kriz (http://www.sv.vt.edu/classes/ESM4714/Gen_Prin/vizthink.html) on the topic of visual thinking, an issue addressed by Maojo et al. as well.

course is not my intention. After all, he also wrote: '*The next great era of awakening of human intellect may well produce a method of understanding the qualitative content of equations. Today we cannot. Today we cannot see that the water flow equations contains such things as the barber pole structure of turbulence that one sees between rotating cylinders. Today we cannot see whether Schrodinger's equation contains frogs, musical composers, or morality--or whether it does not.' [16]⁴, thus confirming the point I made earlier.*

I do mention it nevertheless because this tactic of undermining the credibility of a scholar is applied in Maojo et al. with the goal to discredit Aristotle - note that Ontological Realism is based on only certain aspects of Aristotle's thinking, the main point being the formulation of definitions - using a quote from Bertrand Russell: '*Aristotle maintained that women have fewer teeth than men; although he was twice married, it never occurred to him to verify this statement by examining his wives' mouths*'. Unsuccessful, however, since it turns out that Aristotle was probably right: several papers do confirm that hypodontia is more prevalent in women than men, although due to sample sizes statistical significance cannot always be demonstrated [17-19].

6.5 Go popularity and quality hand in hand?

Maojo et al. did an extensive literature study covering ontologies of various kinds as a result of which they claim the '*strong agreement about the advantages of computational ontologies*' on the basis of all the positive experiences reported. Nevertheless, I believe that these positive experiences must be taken with a pinch of salt. In [20], empirical support is provided that the reliability of findings published in the scientific literature indeed decreases with the popularity of the research field, as was suggested earlier on a pure statistical basis in [21]. Specifically relevant here is 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 [22].

I am thus also very leery about the value of ontology quality measurement frameworks that use uptake and popularity as a positive quality criterion [23] and I prefer quality review over democratic ranking [24]. The drawback of popularity based measures is clearly visible on the NCBO BioPortal [25], where the Medical Dictionary for Drug Regulatory Affairs (MedDRA) and the National Cancer Institute (NCI) Thesaurus occupy 2nd and 3rd place in the ranking. The NCI Theaurus was in 2005 found to violate several ontological and terminological principles [26], claimed to have been cleaned up in 2009 [27] but then found in 2010 to violate dramatically

⁴ I found the quote on Google Books, but did not read the entire volume.

the semantics of OWL in which it has been converted [28]. The structure of MedDRA was in 2005 reported to violate standard terminological principles [29-30] and to correspond merely to what already in 1997 was termed a 'first generation system' [31]. MedDRA's structure remained to date unchanged.

The presence of these two systems - as well as of so many others that I am surprised to find in the collection - clearly raises serious questions about the 'principles' behind the 'NCBO-recommended formats and methodologies for ontology development, maintenance, and use' (http://www.bioontology.org/mission), principles, by the way, which are not those of the OBO Foundry [6].

It is therefore argued in [20], 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'.

6.6 Should usability be favored over quality?

Ontological Realism is a methodology for ontologies that want to be maximally re-usable, at the risk, as often pointed out by Alan Rector in the GALEN project, 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 in the development of high quality ontologies, embracing extreme forms of (model-theoretic) semantics that cut computational constructs loose from reality. So many computational scientists and semantic web engineers have indeed stated that their job is to design algorithms and systems that are guaranteed to reason consistently with the input provided to them but that they have no control about whether the input is any good or bears any relation with what is the case in 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, discussed earlier, is an example of the contrary [28].

7 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.

We do agree with Maojo et al. that further discussion is needed, although not exclusively under the very narrow view of 'science' that they seem - in absence of a definition thereof in [1] - 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 [32]. If he is right, then we know who to blame.

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 [33].

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. As an example, if a concept-based system as defined in [33] is developed for a specific application, then representational adequacy may perfectly be limited 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, both by Maojo et al. and me, 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 nonmainstream it might be.

References

- Maojo V, Crespo J, Garcia-Remesal M, Iglesia Ddl, Pérez-Rey D, Kulikowski C: Biomedical Ontologies: Toward Scientific Debate. *Methods of Information in Medicine* 2011.
- 2. Buekens F, Ceusters W, Moor GD: The explanatory role of events in causal and temporal reasoning in medicine. *Methods of Information in Medicine* 1993, **32**:274-278.
- Bittner T, Smith, B. : Normalizing Medical Ontologies Using Basic Formal Ontology. Tagungsband der 49 Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie 2004.
- 4. Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, Mungall C, Neuhaus F, Rector AL, Rosse C: **Relations in biomedical ontologies.** *Genome Biology* 2005, **6**:R46.
- Smith B, Ceusters W: Ontological Realism as a Methodology for Coordinated Evolution of Scientific Ontologies. *Applied Ontology* 2010, 5:139-188.
- Smith B, Ashburner M, Ceusters W, Goldberg L, Mungall C, Shah N, Bard J, Eilbeck K, the OBI Working Group, Leontis N, et al: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology* 2007, 25:1251-1255.
- Good IJ: A Classification of Fallacious Arguments and Interpretations. *Technometrics* 1962, 4:125-132.
- 8. Lord P, Stevens R: Adding a little reality to building ontologies for biology. *PLoS One* 2010, 5:e12258.
- 9. Dumontier M, Hoehndorf R: Realism for scientific ontologies. *Proceedings of the 6th International Conference on Formal Ontology in Information Systems (FOIS)* 2010:387-399.
- Goldstein S: Bohmian Mechanics}. In *The Stanford Encyclopedia of Philosophy*. Spring 2009 edition. Edited by Zalta EN; 2009

- 11. Baum R, Sheehan W: In Search of Planet Vulcan, The Ghost in Newton's Clockwork Machine. New York: Plenum Press; 1997.
- 12. Ceusters W, Smith B: A Realism-Based Approach to the Evolution of Biomedical Ontologies. In Biomedical and Health Informatics: Proceedings of the 2006 AMIA Annual Symposium. Washington DC: American Medical Informatics Association; 2006: 121-125
- Ceusters W, Capolupo M, Moor GD, Devlies J, Smith B: An Evolutionary Approach to Realism-Based Adverse Event Representations. *Methods of Information in Medicine* 2011, 50:62-73.
- 14. Ceusters W: Applying Evolutionary Terminology Auditing to the Gene Ontology. Journal of Biomedical Informatics; Special Issue of the Journal of Biomedical Informatics on Auditing of Terminologies 2009, **42:**518-529.
- Young N, Ioannidis J, Al-Ubaydli O: Why Current Publication Practices May Distort Science. In Book Why Current Publication Practices May Distort Science (Editor ed.^eds.), vol. 5. pp. e201. City; 2008:e201.
- 16. Feynman RP, Leighton RB, Sands ML: *The Feynman Lectures on Physics: The electromagnetic field*. Addison-Wesley Pub. Co.; 1965.
- 17. Celikoglu M, Kazanci F, Miloglu O, Oztek O, Kamak H, Ceylan I: Frequency and characteristics of tooth agenesis among an orthodontic patient population. *Med Oral Patol Oral Cir Bucal* 2010, 15:e797-801.
- Vahid-Dastjerdi E, Borzabadi-Farahani A, Mahdian M, Amini N: Non-syndromic hypodontia in an Iranian orthodontic population. *Journal of Oral Sciences* 2010, 52:455-461.
- 19. Fekonja A: Hypodontia in orthodontically treated children. European Journal of Orthodontics 2005, 27:457-460.
- 20. Pfeiffer T, Hoffmann R: Large-Scale Assessment of the Effect of Popularity on the Reliability of Research. *Plos ONE* 2009, 4:e5996.
- 21. Ioannidis J: Why most published research findings are false. In *Book Why most published research findings are false* (Editor ed.^eds.), vol. 2. pp. e124. City; 2005:e124.
- Goodman SN, Royall R: Evidence and scientific research. American Journal of Public Health 1988, 78:1568-1574.

- 23. Aranguren ME, Wroe C, Goble C, Stevens R: In situ migration of handcrafted ontologies to reason-able forms. *Data & Knowledge Engineering* 2008, 66:147-162.
- 24. Smith B: The Evaluation of Ontologies: Editorial Review vs. Democratic Ranking. *Proceedings of InterOntology 2008 - Tokyo, Japan 2008*:127-138.
- 25. Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, Jonquet C, Rubin DL, Storey MA, Chute CG, Musen MA: **BioPortal: ontologies and integrated data resources at the click of a mouse.** *Nucleic Acids Research* 2009 **1:**37.
- 26. Ceusters W, Smith B, Goldberg L: A terminological and ontological analysis of the NCI Thesaurus. *Methods of Information in Medicine* 2005, 44:498-507.
- 27. Coronado Sd, Wright LW, Fragoso G, Haber MW, Hahn-Dantona EA, Hartel FW, Quan SL, Safran T, Thomas N, Whiteman L: The NCI Thesaurus quality assurance life cycle. *Journal of Biomedical Informatics* 2009, 42:530-539.
- 28. Schulz S, Schober S, Tudose I, Stenzhorn H: The Pitfalls of Thesaurus Ontologization the Case of the NCI Thesaurus. In AMIA Annual Symposium Proceedings. Washington D.C.: AMIA; 2010: 727-731
- 29. Bousquet C, Lagier G, Lillo-Le Louët A, Le Beller C, Venot A, Jaulent M: Appraisal of the MedDRA conceptual structure for describing and grouping adverse drug reactions. *Drug Safety* 2005, 28:19-34.
- 30. Henegar C, Bousquet C, Louët AL-L, Degoulet P, Jaulent M-C: Building an ontology of adverse drug reactions for automated signal generation in pharmacovigilance. *Computers in Biology and Medicine* 2006, 36:748-767.
- 31. Rogers J, Rector A: Terminological Systems: Bridging the Generation Gap. In Proceedings of the Annual Fall Symposium of the American Medical Informatics Association. Philadelphia PA: Hanley & Belfus Inc; 1997: 610-614
- 32. Russell B: *The Impact of Science on Society*. London: Unwin Paperbacks; 1976 (reprint 1952).
- 33. Klein GO, Smith B: Concept Systems and Ontologies: Recommendations for Basic Terminology. Japanese translation in Journal of the Japanese Society for Artifical Intelligence, 25 (3), 2010, 317-325 2010, 25:433-441.