Response to reviewers of ICBO2015 for the paper

An ontological analysis of diagnostic assertions in electronic healthcare records
Werner Ceusters and William Hogan

---------- REVIEW 1 ----------

OVERALL EVALUATION: 0 (borderline paper)
REVIEWER'S CONFIDENCE: 4 (high)

This paper is a report of the two authors efforts to model two sentences in a hypothetical EHR. I must admit that I am a little baffled as to what the exact point of this paper is.

--> 1. The point was explained in the 1st sentence of the last paragraph of the introduction: "The purpose of the work reported on here was to assess to what extent each of the authors of this paper—two experts in Referent Tracking—would be able to develop independently from one another a collection of RTTs that describe the same portion of reality in a semantically-interoperable way."
→No action taken.

Since we do not know the ground truth (what the patient actually had), I think it is obvious that there is no "optimal" way of representing these sentences ontologically.

--> 2. As explained in the same sentence, the purpose was not to model sentences – we did not work with sentences at all but with diagnostic entries made in the problem list of an EHR as shown in Table 1 – but to represent the portion of reality described by means of such entries.
→ No action taken for this point

--> 3. The 'ground truth' is much broader than just what the patient had (this being part of the non-assertional part of the portion of reality (POR) under scrutiny): it includes what the clinicians stated about what the patient had (these statements being part of the assertional part of the POR under scrutiny). If what the patient precisely had cannot be inferred from what was stated, it would be wrong to construct a collection of RTTs that states that the patient has such or such a specific type of disorder. To represent the non-assertional part of a POR that a collection of assertions is about, one has to resort to these assertions ***and*** to what has already been established to be the case through other means. The optimal collection of RTTs would be the one which satisfies the following criteria: (1) it consists of RTTs which describe the non-assertional part of the POR only to the extent to which there is enough evidence for what those RTTs themselves assert to be true (e.g. there is sufficient evidence that the patients are human beings, there is not sufficient evidence that the diagnoses are correct) and (2) it consists of other RTTs which describe the assertional part in relation to the RTTs referenced under (1).
→ We added this explanation to the discussion section of the paper.

The fact that two expert practitioners largely agree on a representation is not particularly surprising. But it is not particularly enlightening to read about the differences (e.g., "X represented P1 as a human with a patient role. Y represented P1 as a material entity without assigning a patient role,..."). Since any

representation is a model of reality, I am sure that the two authors would wind up agreeing that the patient was in fact a human being if they wound up using a model that needed to specify it.

--> 4. The differences might not be enlightening to this reviewer, but they are nevertheless important. In Ceusters W, Smith B. A Realism-Based Approach to the Evolution of Biomedical Ontologies. Proceedings of AMIA 2006, Washington DC, 2006;:121-125, the authors argued that '*an optimal ontology should constitute a representation of all and only those portions of reality that are relevant for its purpose*'. So it might very well be that the two authors have used two distinct ontologies to describe the same portion of reality, but that the two ontologies do so from two distinct perspectives which are despite both being veridical not derivable from each other because axioms allowing to do so would be missing for the simple reason that such axioms would fall outside the purpose of the specific ontologies. That would lead the representations by each of the authors not to be semantically interoperable, the assessment thereof being one of the purposes of the work described (see response --> 1 above).
→ We added this explanation to the conclusion of the paper.

Minor

reception of the other's result. => receipt is the better word

--> 5. corrected

Nevertheless, the persistent, glaring flaws of DO => I agree that the DO is a work in progress, but the authors are implicitly stating that the DO is flawed because it is in direct contradiction to OGMS. I am not involved with either ontology, but I think it is fairer to say that each has its own particular view on reality, and each ontology has its own use cases.

--> 6. We added 'glaring flaws' of DO 'in terms of OGMS'.

---------- REVIEW 2 ----------

OVERALL EVALUATION: -1 (weak reject)
REVIEWER'S CONFIDENCE: 4 (high)

In their paper, the authors provide a case study on the application of reference tracking to a single episode of medical documentation with two diagnoses of one patient. The two authors developed independently from each other a set of Reference Tracking Tuples and compared them informally. The identified particulars were mapped to existing ontologies.

7 out of 9 cited references were self citations.

--> 7. This reviewer might have been less confident in his or her assessment had he or she been minimally familiar with the minimal rules for submitting a review. Here is not too bad a start: http://www.wikihow.com/Write-an-Article-Review
→ No action taken.

---------- REVIEW 3 ----------

OVERALL EVALUATION: 0 (borderline paper)
REVIEWER'S CONFIDENCE: 3 (medium)

The paper describes the application of the reference tracking (RT) methodology by the two authors to the same clinical description and the qualitative analysis of similarities and differences, such analysis being based on the ontologies used and the meaning of each entity. The main conclusion is that RT is helpful to obtain explicit interpretations of the clinical content so they can be systematically and formally analyzed. Besides, the analysis can help to detect flaws in existing ontologies since, for instance, differences in the results of different people could be due to such flaws in the ontologies.

--> 8. That is correct

The paper provides a good example of the application of RT and its support for ontological analysis. Practically speaking, the results of the application of RT are somehow like semantic annotations which can be then represented as triples and even compared.

--> 9. RTTs are actually time-indexed triples, thus quintuplets as exemplified on page 1 of the paper. Although we stated that the work presented here was only the first step in reaching the goal expressed in response --> 1 above, we failed to note that the analysis conducted for this paper did not include the time-indexing aspects and that therefore all results are expressed as triples.
→ We clarified this in the methods section of the paper.

A limitation of the study is the participation of only 2 subjects and the analysis of one report. The participation of a larger number of experts would have made this study more interesting.

--> 10. We acknowledged this limitation in the paper and of course would welcome very many more ontologists to resort to referent tracking.

It seems that all the entities are given the same importance in the analysis and discussion of the results.

--> 11. That is correct. We cannot think of a good reason why some entities would be more important than others for the goal we wanted to achieve.

In the conclusion, the authors mention that the authors agreed on key entities. The objective of the paper was to see to what extent the authors are able to create "interoperable RTTs" but I would have expected the formulation of some research hypotheses in the paper that could lead to such conclusions through the analysis. The two collections of RTTs could be approached as the content of two EHR systems. I would have expected some answer to what would have been then possible for them to exchange according to such interpretations.

--> 12. Point taken. Although we wrote that we report on the first step here, we did not specify what the next steps would be. We added this to the conclusion.

---------- REVIEW 4 ----------

OVERALL EVALUATION: 1 (weak accept)
REVIEWER'S CONFIDENCE: 5 (expert)

Carving up a domain of reality is a matter of personal preference. At least this seems to be the moral that emerges from this paper (and this despite the fact that, I am guessing, the authors seem to be aiming at the opposite conclusion).

--> 13. We are indeed aiming at the opposite conclusion and will make this clear in the paper. Interestingly, reviewer #1 by contrast was not surprised at all that we came to similar results. The divergence of opinion between these two reviewers adds additional support for our claim that our results are indeed reportable in the scientific literature and noteworthy.

Quite interestingly, we do not necessarily disagree with this moral. Choosing between two competing "conceptualizations" of the same domain of reality is also a matter of personal preference, which should not necessarily mean that the "conceptualization" one has not chosen is wrong. One of the frequently used criteria for deciding between two ontologies of the same domain is Ockam's razor (whereby the ontology that postulates the least entities should be preferred), though that is more of a *convenience* principle rather than anything else.

--> 14. The notion of competing conceptualizations is valid when the underlying reality has not been completely discovered yet. But here we made sure – with one oversight: whether the two diagnoses were correct or not – that the underlying reality for the example was fixed. What we studied and finally agreed upon, were thus not two distinct conceptualizations, but the very same conceptualization expressed in different, though semantically equivalent, ways. We also note that 'personal preference' is a barrier to interoperability.
→ We clarified this in the paper.

At any rate, the current paper attempts precisely that: give two biomedical researchers the task of representing a chunk of biomedical reality, and you'll end up with two rather different conceptual schemes. Comparing the two schemes might reveal some similarities, though, as far as we can tell, the degree of difference between the two is pretty significant--despite the fact, and this should be emphasized, both researchers seem to have had major exposure to very much the same array of contemporary controlled biomedical vocabularies/ontologies. This is, in our view, quite surprising: we would have expected much more of a convergence than simply having less than half particulars in common.

--> 15. Although we acknowledge that each of us selected to a certain extent distinct entities as basis for our representations, our hypothesis is that the two representations are nevertheless compatible. Since this paper is only the first step in this endeavor, that is indeed not obvious. Again, we note that this reviewer reached the opposite conclusion of reviewer #1, making our result non-obvious and not predictable by two experts in the field.
→ We clarified this in the paper, also in response to similar remarks by other reviewers.

Be that as it may, this remains an interesting onto-sociological case study, and deserves to be reported. Our issues with this paper do not necessarily address its core, but more some of its secondary aspects. Here they are (I will not bother during the rest of the discussion to distinguish between an RTT and a

"concretization of an RTT," a table and the "concretization of a table" etc., distinctions that seem to be consuming so much of the authors' attention; nothing in the discussion below hinges on such distinctions (which should *not* mean that they are illegitimate)):

--> 16. We agree that we are pedantically precise in our representations and strongly believe that every ontologist should be pedantically precise, but we can accept that reviewers are mere precise

1. We could not find the definition of an ICE in the paper, and that makes the purportedly philosophical discussion touching upon issues in abstraction, abstract/concrete objects etc. hard to follow. If anything, the philosophical musings in section #4 stay testimony to the fact that, for someone wishing to master the intricacies of the act of referring, there is no alternative to a serious background in the philosophy of language, abstraction, and mind.

The problem with ICBO is the 5-page restriction which does not make sense for an electronic publication. As a result, there is not enough space to provide the background needed to understand everything. Therefore we provide references to the background, at the risk of being slapped for self-citation (a risk that was indeed realized but we digress). We could add the current definition for ICE from the IAO, but that definition is itself currently under review, an issue dealt with in another paper accepted for ICBO 2015. → No action taken.

2. Nor is the notion of an RT tuple (RTT) very clear. Up to a point one might suppose that an RTT is a row in any of the Tables 2-5, though the authors also state that "each row in the tables represents [*]part of[*] an RTT... " (section 3, 1st par.). It is also stated that RTTs are concretization of ICEs, but this does not help given that ICEs are not defined anywhere.

--> 17. An example of an RTT was given on page 1. The transition to the tables involved chopping of the time indexing. We clarified this also in response to another reviewer.
→ No additional action taken.

3. Hinging on the above are some questions regarding the reasons for identifying certain "Ind." values: for example, why is it that the 14th row in Table 3 has the same value of the "Ind." field as row #4 in Table 2? What is the basis for this determination? This needs further explanation.

--> 18. This is because in the representations of each author exactly the same entities were referenced. Of course, when the individual results were disclosed, the identifiers used were different. But during comparison, with the goal to make the tables more compact, we introduced one global unique reference scheme. This was explained in the Methods section.
→ No action taken.

4. We would be interested in learning what possible use could emerge from making a distinction between an ICD code, and a *concretization* thereof, at least in the context of this paper. If none, we would advise dropping such distinctions, which might also relax the overly pretentious phrasing.

--> 19. It is relevant – as was explained in the paper – to the analysis of what from an ontological perspective on ICE counts as a diagnosis in an EHR, the current situation being that the notions of ICE and concretizations thereof are currently not well enough developed to do so. Surely a proper ontological analysis of "diagnosis" is useful?