Reviewer's report

Title: OGMS: The Ontology for General Medical Science

Version: 1 **Date:** 2 July 2013

Reviewer number: 1

Reviewer's report:

OGMS's merit is the ontological analysis of the entities involved in pathological processes in the most general sense, and the tripartition between "disordered" material entities, which can be bearers of dispositions, which can be realized by processes.

This paper focuses on these three "core" universals, and describes them in a clear way with numerous examples. However, the paper could be technically improved in several aspects, and the current makeup of OGMS raises several issues, which could be further discussed.

- Global remarks on the organisation of the paper:

The manuscript focuses on the three "core" classes on the one hand, and on the applications of OGMS on the other hand. This gives the impression that OGMS is a minimalist ontology, just consisting of disorder, disease, and disease course. Other important classes like 'clinical finding', 'disorder', 'pathological bodily process', 'treatment' etc. are only addressed once, but not further introduced. The picture of OGMS given in this paper would be less biased if the scope of classes described were extended. Even if this means some overlap with already published material, it would be good to have it in this paper, too, at least in the form of a table. In compensation, the description of the ontologies that consume OGMS could be much more concise. In the current form of the manuscript this description includes five pages, which is disproportionate.

- Global remarks on terms and formalisms used in the paper

It seems that the words "term", "class", "type", "universal" (as well as "subclass", "subtype") are used as synonyms which is certainly not the intention of the authors. I recommend using the term "representational unit" (RU) and making clear what RUs denote. If this is different between the OBO (where the RUs are supposed to denote universals) and the OWL version (where RUs denote classes), this should be elucidated in the beginning; and the choice of terms should be consistent. This will improve the readability of the paper, as it reduces confusion especially for those who are not familiar with the domain.

I do not agree with the choice of "term" for RUs, because this obfuscates the distinction between terminologies and ontologies. Terms are entities of human language, but ontology is not about human language. Although ontologies should have human-readable labels, there naming should obey other rules than the

formation of terms in a domain sublanguage. For instance, there are good reasons for naming a RU 'pathological bodily process' (clarity, unambiguity), although this highly artificial phrase has possibly never be used by any medical author, and it is not part of any medical terminology. The missing distinction between "label" and "term" has already caused enough confusion and misinterpretations (e.g. that most GO "terms" are never found – literally – in scientific text). The OBO Foundry should play a more active role in standardizing the way to talk about representational artefacts, their components, and the things they denote.

There are currently two syntaxes used in OBO Foundry ontologies, viz. OBO and OWL. OWL expressions can be rendered by different syntaxes. In this paper it is not always clear which syntax is used. For instance, "x R SOME y" looks like OWL Manchester Syntax, which, however, requires an equivalence operator, such as subClassOf or equivalentTo between "x" and "R". In addition, "SOME" should be lower case. In contrast, "ido:infectious disorder is_a ogms:disorder" looks like OBO syntax, because "is_a" is not part of the OWL syntax (unless it is introduced as an object property, which would be very unorthodox). Finally, labels that consist of more than one token should be quoted (like in Protégé).

- Issues of OGMS proper

The acronym OGMS alludes to medical science, but it seems to be much more related to health care (the authors refer to "main types of entities involved in a clinical encounter"), which I would not necessarily subsume by "medical science". Here it naturally competes with SNOMED CT, which is, however only indirectly addressed. It would certainly be interesting to know which the obstacles of an integration between OGMS and SNOMED CT would lie.

The reference to ICD is less important, due to two reasons: It has become positively accepted by the WHO that SNOMED CT should be the vocabulary in which clinical information is primarily encoded. ICD – in the future available as different, purpose dependent ICD-11 linearizations would then be mainly used for purposes like health statistics and reimbursement. The meaning of ICD codes is therefore more and more conceived as proxies for queries, and not as classes in an OWL sense. A future common ontology between ICD and SNOMED CT is being prepared by both the IHTSDO and the WHO, which should facilitate the bridging between both terminologies. It therefore be appropriate in this article to put more emphasis on SNOMED CT than on ICD.

What about structural changes that come into being during the course of a disease, e.g. the swollen nasal mucosa that evolves in a manifestation of allergic rhinitis? Is "Allergic Rhinitis (Disorder)" the same as "Swollen Nasal Mucosa due to Allergen" the same?

The authors write that "for disordered anatomical parts this *WILL* mean using terms from the Foundational Model of Anatomy". I would be less prescriptive regarding anatomy ontologies, especially because the FMA lacks a clear anatomical commitment, particularly regarding ill-formed anatomical entities.

The author write that "OGMS can be used to factor ambiguous terms such as "HIV" into non-ambiguous terms whose meaning and reference is clear". On the other hand they write "To insert a single term into the BFO framework that accommodates all of these would violate the disjointness of continuants and occurrents, which is a presupposition of the BFO architecture"

Whereas I fully agree that the ontology must ensure that there are no common instances of disjoint universals, it does not appear clear how OGMS actually deals with ambiguous terms like "HIV". Clinicians won't care about whether this term denotes a disorder or a disease, and might consider this distinction as pointless. But they want to have one code for HIV. One possibility would be to allow for disjunctive classes such as "clinical condition", as suggested by Schulz et al. "Scalable representations of diseases in biomedical ontologies, which leaves the instances uncommitted.

In ordinary English, the meaning of the terms "disorder" and "disease" is diffuse, let alone in other languages in which "disorder" and "disease" translate to the same term. OGMS should consider using neutral terms, which make the ontological distinction clear, e.g. "clinical disposition", "clinical structure", "clinical process".

Furthermore, an ontology should be neutral regarding normality vs. abnormality. It is not only the continuum characteristics with the need to introduce thresholds, as rightly discussed by the authors. It is also the problem that the same kind of entity can be considered normal or abnormal, according to the circumstances. E.g., gingival bleeding is normal when children lose their milk teeth, whereas it is pathological in adults. What is normal and what isn't is at least partly epistemic. I don't see that OGMS would lose its expressivity and usefulness if it maintained complete agnosticism regarding canonicity. The fact that a doctor considers a body structure, process, or disposition as abnormal would be better encoded as an attribute of an information entity, such as a record entry.

By the way, there are many record entries referring to phenomena that are generally not regarded as pathological such as body height, skin colour etc., which are nevertheless worth reporting and may play a role in decision making, risk estimation etc.

In the definition of "extended organism" the authors also mention material entities that overlap the organism. This could be a venous catheter, an artificial limb, a cochlear implant, etc. (Examples would be helpful here). The statement "If an organism without any holes or cavities were to exist, the terms 'organism' and 'extended organism' would for that organism be synonomous", suggests that the only differentia between extended organism and organism is the presence of cavities in the former. This is not consistent with the before mentioned overlapping material entities

The authors restrict the bearers of clinically relevant dispositions (diseases) to the whole organism. I could accept this as an axiomatic assumption. However,

the rationale given in the paper ("part of a disorder may be located in an anatomical entity without being a disorder of that entity") is not convincing. I would not object considering a lung tumour metastasis in the liver a liver disorder. Why shouldn't we identify and name pathological dispositions of organism parts (e.g. the disposition of nasal mucosa to engage in an allergic reaction when exposed to an allergen) ?

"Terms" are used to form singular referring expressions (the authors give the example 'John's subglottic stenosis'). What about plural referents such as populations or pluralities of molecules or cells?

Major compulsory revisions:

1. Insert a table presenting the most important categories of OGMS (besides disease, disorder, disease course)

2. Reduce the length of the subsections in which you describe each ontology. A third of a page should be sufficient on average, per ontology.

3. Make rational use of the words "term", "class", "type", "universal" (as well as "subclass", "subtype"). If you think that the word "term" really is appropriate for naming representational units in UMLS, give a good reason.

4. Make a clear distinction between the syntaxes you use. Analyze whether OWL syntax would not be sufficient. If not, give convincing reasons why you use two syntaxes in parallel. Use OWL Manchester syntax properly.

Minor essential revisions

5. Revise the definition of "extended organisms"

6. Explain better why plural reference is not accepted and show how you represent collections without plural reference – or, drop the remark on plural reference

7. The acronym "PVL" should be spelt out.

Minor discretionary revisons

8. All other topics addressed under " Issues of OGMS proper"

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.