

Response to Reviews

on

Ms. No.: JBI-08-124

Title: Applying Evolutionary Terminology Auditing to the Gene Ontology

Corresponding Author: Dr. Werner Ceusters

Special Issue Editors' Comments on original submission:

Recommendation: major revision necessary

Overall, the reviewers found this paper innovative and interesting. However, they collectively raised a variety of issues that must be addressed.

Major issues include the concern of Reviewer 1 regarding the "last version equals gold standard" principle.

→1 this has been addressed. See below under 'reviewer 1'

Questions about the specification of Formula (1), as well as its application and interpretation, are also raised.

→2 this has been addressed. See below under 'reviewer 1'

Reviewer 3 was very critical of the length of Section 3, and from our own reading of the manuscript, we concur with this assessment. The section contains a great deal of background information that should be tightened up.

→3 this has been addressed. We removed those parts that are not directly relevant.

Reviewer 3 also noted that the author misinterpreted the content of the GO monthly reports. Implications of this must be determined, and any inaccuracies resulting from this must be corrected.

→4 this has been addressed. See below under 'reviewer 3'

Reviewer 3 also had trouble reading some of the figures. Figures 1 and 2, in particular, could benefit from some reformatting.

→5 We have added colours as requested by the reviewer and removed also a number of data series from figure 2.

Tables 1 through 3 each seem to have an extraneous row with content: (1) (2) (3)

→6 These rows contain column labels which are extensively used in the text for easy reference, thus avoiding that the reader has to count the columns. To eliminate the possibility that the rows would be misinterpreted as extra data, rather than labels, I removed the separating line with the row above in Table 1 and 2. I removed this row from Table 3 since no references are used in the text.

Reviewer 2 limits remarks to the organization of the Introduction. In particular, there is a request for some material at the outset on the need for ontology auditing.

→7 Has been done.

Furthermore, it would be helpful to include a brief statement of the specific problem being addressed by this paper earlier on in the introduction.

→8 Has been done. See last paragraph of the main introduction

As a related issue, the editors feel that it is important for the author to identify exactly the theoretical differences between reference [19] and the current paper. (There certainly is a difference from an application standpoint: GO as opposed to SNOMED.)

→9 Has been done. See the beginnings of the conclusion.

Additional editorial remarks:

** There are too many numbered lists in the midst of the text that may cause confusion for the reader. For example, see the opening paragraph of Section 3.1.*

→10 We removed these lists, except for 1 in which case it would make the paper less readable

** Certain words found in the paper are not English: for example, "precisification" and "obsolescence." If they are technical terms from the literature, then provide references.*

→11. Words have been removed and corresponding sentences rephrased

** It should be "an RU" rather than "a RU" throughout.*

→12. This has been fixed.

** Use proper abbreviations: "Nr" should be "No." in Table 8. It is "et al." in the first paragraph of Section 5.1.*

** Use consistent capitalization.*

→13. Has been fixed.

** The formatting of the references needs to be cleaned up. See [27] and [44], for example.*

→14. Has been fixed.

Reviewers' comments:

Reviewer #1:

The topics of this paper are important and despite we are currently at the first steps, the results are encouraging and research in this direction should be published. A number of comments are given now, some of them could lead to localized changes in the paper:

1) In 1.2 (and therefore 3.2) the assertion that "the structure of a terminology should mimic the pre-existing structure of reality" should be commented. I have serious doubts about the general validity of this assertion. In a domain like disease, I really wonder if we can speak of a pre-existing structure. Are the human diseases organized by system, by causative agent or by type of disease (like tumors)?

→15 I added to the sentence: "..., rather than being determined – and usually limited – by, for example, what the representation language is able to express, by mixing ontology with epistemology, or by incidental features related to the context in which the terminology is built, thus confusing the 'model of meaning' with the 'model of use'." including references

2) Following 1), I wonder if the conclusions are too much GO-centric? Other arguments for other metrics could be developed in the future.

→16 After re-reading the conclusions, I don't understand this reviewer's first remark concerning the 'too GO-centricness'. The conclusions are phrased generic, so I believe. I added a phrase concerning other similar metrics that might be developed.

3) The principle of "last version equals gold standard" is a serious limitation and should be more extensively discussed. From this principle, we always produce better versions. This is an optimistic assertion! What happens if we introduce new errors. The intend is always to improve an existing version, but the result is not guaranteed. Only the result is based on reality, but your metric measures the intend. Please comment.

→17 . We added the following sentences: "This is motivated, as described further in detail, by the assumption that new versions of a terminology are better than previous ones, despite the possibility that with each version new errors are introduced. But if terminology curators take their work seriously, such errors are likely to be corrected in later versions, for instance on the basis of remarks from the community when the version is used in practice. It seems obvious that using other terminologies as gold standard has at least the same risk, if not a bigger one. After all, if one is sure about the correctness of another terminology covering the same domain, why then bother to develop a new one?"

4) The formula (1) becomes less and less sensitive when the size of ontology increases. In formula (1), RU should not be part of the formula, the dependance on RU is expressed by e (as far as I understand).

→18. The reviewer is right about this notational mistake. We corrected the formula.

5) There is no explicite distinctions between horizontal growing of an ontology (new children on existing nodes) and vertical growing (new descendants on existing leaves). Horizontal growing is typically an "absence" but vertical growing is just continuation of work with more details. I suggest at least to make the distinction by some comment (not to modify the metric).

→19 . The discussion contained the sentences: "But the current method does not take into account at what level in the classification hierarchy the mistake is made. This applies not only to omissions but also to unjustified additions and property-RU changes. Therefore, another strategy, to be tested in the future, is to base the magnitude of the error also on the difference in the hierarchical position of a RU in an older version as compared to the newer one." But in order to make this clear earlier in the paper, we added the following sentence: 'Note that we do not assign a higher or lower error magnitude to unjustified absences that occur at the level of leaf nodes in a terminology as compared to absences at higher levels in the hierarchy.'

6) Tables 2 and followings have 9 entries (lines), but all the calculation with the formula 1 are made with 8 entries. This may slightly change the values!

→20 . That is not correct. Terminology 1 uses indeed 9 entries, as witnessed in Table 2 by the sum of m and n (8 +1). Reality, and then other terminologies use indeed only 8 entries because justified absences are not accounted for: there is indeed, as indicated in the original version of our paper, an infinite amount of such justified absences.

7) The magnitude of errors in Table 1 is arbitrary. Please comment on that. This is certainly a valid experimental choice.

→21 . We do not really believe the magnitudes are so arbitrary since they are based on Granular Partition Theory (see reference). We gave our motivation in the methodology section. But indeed, as already stated in the first submitted version, alternatives must be studied.

Last but not least: your metric does not measure all criteria for a better quality.

→22 . That is correct and so intended: it measures only differences in the structure of reality as compared to the structure of the terminology

For example the fit between a RU and the reality in an ideal ontology should be documented by a definition (formal or encyclopaedical). This is not represented in the metric.

→23 . We disagree here. First such a definition does not change the nature of a fit, or in other words, whether there is a fit does not depend on any such definition (although, as pointed out by Cimino in his desiderata to which we refer in the discussion section, it might be easier for a terminology auditor to assess what terms stand for. But that is an epistemological issue); Second, I am not sure what would be the value of a definition given for an erroneous RU, and lastly, definitions might themselves be problematic, see for instance the definition for water in the ICPC: “Water is a type of Nursing Phenomenon of Physical Environment with the specific characteristics: Clear liquid compound of hydrogen and oxygen that is essential for most plant and animal life influencing life and development of human beings”.

There are other aspects which are not taken into account. They could be discussed in the paper.

→24 . The discussion section starts with a description of other approaches. It is possible to design other metrics indeed, a point that we acknowledged in the conclusion in response to this remark.

Thank you for this important contribution. I hope that my comments may help.

→25 . Yes they did. Many thanks !!!

Reviewer #2:

This manuscript fills a much needed void in ontology management-- that in the area of ontology evaluation.

Overall, the manuscript is very well written and clear.

One aspect that may be improved is more introductory background. The current manuscript jumps right into description of the Gene Ontology; however, it might be helpful to have a few introductory paragraphs about the need for ontology auditing.

→26 this has been done by adding an introductory paragraph pointing out an analogy with financial auditing. Because the lengthiness of our paper was already a point of concern for one reviewer (and confirmed by the editor), and because we are dealing here with a paper for a special issue on terminology auditing, we expect the editorial to this special issue to cover the need for terminology auditing more in detail, and thus preferred to be brief here.

Reviewer #3:

The authors present a novel methodology for assessing the quality of ontologies, and apply this methodology to the Gene Ontology (GO).

Their methodology works by assuming that incremental changes to an ontology bring it ever closer to perfection, such that the most recent version is used as a 'gold standard'. Each type of ontology modification is assigned a magnitude score and using these scores, older versions of the ontology are compared with the gold standard. The GO monthly reports were manually mined to make the comparisons.

Assessing the quality of biological ontologies is notoriously difficult, mainly because there is no way of knowing how close they are to completely representing their domain because the full scope of the domain is unknown, and the authors make a fair attempt to address this difficult problem.

Major compulsory revisions:

My first major criticism of the manuscript is its length, the materials and methods section in particular, which alone comes in at 12 pages. Making this section shorter would make this paper much more accessible, so I suggest the authors shorten the first half of the materials and methods section considerably (particularly the background information not directly related to their methodology). Perhaps some of the reference information regarding terminology could be summarized as a table?

→27. We removed the material not directly related to the methodology from that section. Bits and pieces had then to be added to the discussion to make certain claims clear.

My second major criticism is that the authors have misinterpreted some of the information found in the GO monthly reports. Results are broken down by the different 'databases' that made the changes, but in fact the acronyms they believe to be databases are actually the initials of individual GO editors who don't belong to any specific database (e.g. MAH = Midori Harris) and sometimes encode other information (e.g. OS = GO 'original set'). In fairness, this isn't entirely the authors fault as the key in the monthly reports does indicate that the column in question is the contributing database, but the paper and figures do need to be revised with this in mind. The authors can contact the GO helpdesk to find out what the different abbreviations stand for.

→28. We were indeed led by the term used in the reports. As requested we contacted the GO helpdesk to clarify the issue and the reviewer is right. We modified the paper and used the term 'source' instead of database.

Here the answer we obtained on the basis of our requests:

“The entries refer to the person or database group that added the term. Members of the GO Editorial Office have a whole number range to themselves as they are likely to be adding large numbers of terms. Database groups with members who occasionally add terms are assigned an ID range for their database, rather than per-person. If you look in the file go/numbers/go_numbers, there's a list of what the various abbreviations stand for. Cheers, Amelia.

--

Amelia Ireland

GO Editorial Office

<http://www.ebi.ac.uk> || <http://www.berkeleybop.org>

BBOP Plant Project: <http://bbopgarden.blogspot.com>”

Discretionary revision:

I found figures 1, 2 and 3 (particularly fig 2) tangled and difficult to interpret properly - is there any way these could be made clearer, e.g. different colours for different lines?

→29. Fig 1 and 2 have been reformatted using colours.

Special Issue Editors' Comments on first revision:

Dear Dr. Ceusters:

We have received the reviews for your revised manuscript. Two of the reviewers are now satisfied. The third reviewer has added a small number of follow-up requirements. The special issue editors feel that the only mandatory change for final acceptance would be to appropriately include the references mentioned by the reviewer. We are looking forward to your final submission.

Reviewer #2: This resubmitted manuscript is an improvement from the original submission. There still seems to be some lacking background information about what terminology auditing is, why it matters, and how it is important for something like the Gene Ontology. As a result, it may be difficult for the reader to completely grasp the context and the importance of the work presented. The addition of the financial statement analogy is helpful, but still does not provide needed background on terminology auditing or evaluation.

→ 1 We believe this will be covered in the intro of the special issue.

Understanding the concern about space, there is still disproportionately less needed background compared to the rest of the manuscript.

→ 2 We disagree. There is a serious lack of understanding amongst our peers about what the realist agenda of terminology and ontology authoring and evaluation is all about. The background provided makes the issues, so we hope, clear.

While ETA as proposed in the present manuscript is a great advance in terminology auditing, it might be helpful for direct discussion about how ontologies are currently evaluated. Consider how a reader might evaluate ETA versus other evaluation methods -- e.g., (Brewster C, Alani H, Dasmahapatra S, Wilks Y. (2004) Data Driven Ontology Evaluation. International Conference on Language Resources and Evaluation) or those described in (Brank J, Grobelnik M, Mladenic D. (2005) A Survey of ontological evaluation techniques. SiKDD).

→ 3 Has been done: we added an entire new paragraph with the title 'Concept-based ontology auditing'