

Response to reviewers for pHealth 2017
A Realism-Based View on Counts in OMOP's Common Data Model
Werner CEUSTERS and Jonathan BLAISURE

Dear Mr. Ceusters,

Thank you for submitting your contribution titled A Realism-Based View on Counts in OMOP's Common Data Model to the pHealth 2017 Conference to be held from 14-16 May 2017 in Eindhoven, The Netherlands.

We are delighted to report that your submission has been accepted for oral presentation. All submitted papers were reviewed by a minimum of two reviewers and members of the Scientific Program Committee (SPC). Below are the comments from the reviewers and SPC members. Please submit a revised version as soon as possible, but latest by March 17th, carefully taking into account all the reviewers' suggestions. Even if you have not been asked to make any revisions, please check your manuscript to ensure that it is of high quality and complies with the IOS Press Instructions to Authors published on the conference website as well as at <http://www.iospress.nl/service/authors/latex-and-word-tools-for-book-authors/>. Accepted papers will be published in the Series on Health Technology and Informatics at IOS Press which is indexed at Elsevier's EMCare, Elsevier's SciVerse Scopus, Google Scholar / Google Books, MEDLINE, PubMed, Thomson Reuters' Book Citation Index, Thomson Reuters' Conference Proceedings Citation Index. Please provide complete references, strictly following the Vancouver Style. Papers not complying with the IOS format and the reference list requirements will not be published.

You have been allocated a total time of 25 minutes for your presentation - 20 minutes for the presentation with 5 minutes for questions and discussion. These time limits will be strictly monitored. Please time your presentation prior to your arrival. Please check the day and time of your session on the congress website <http://phealth2017.eu//>.

Please note the following:

- At least one author named in your presentation must be registered by 21th March 2017 for your paper to be included in the program and proceedings.
- You must STRICTLY adhere to the aforementioned author guidelines and layout requirements, also published on the pHealth 2017 website, when resubmitting your revised paper. Ensure that your paper in all details complies with the specified IOS MS Word format. PDF Files are NOT acceptable as they cannot be finally edited by the responsible book editor.
- The ultimate deadline for re-submission through the submission website is 17th March 2017. Please submit your revised version via the conference website at <https://phealth2017.online-registry.net/> by logging in and clicking on 'Re-submission of contributions'. For meeting urgent requirements of reviewer regarding missing items or deeper explanations, the paper might be extended by 1-2 pages.

For any further general conference information you may require please send an email to info@phealth2017.eu.

We congratulate you to your success and we look forward to welcoming you in Eindhoven.

Kindest regards

on behalf of the Scientific Programme Committee
Bernd Blobel
Chair of the SPC

Comments to authors:

Reviewer 1:

General appraisal:

A well-written paper, addressing once again the hard problem of “use mention confusion” in clinical documents and supporting artefacts. Similar critiques have been issued more than one decade ago targeting similar problems in HL7.

- That is correct. Aren't we relentless? See: Smith B, Ceusters W. HL7 RIM: An Incoherent Standard, Stud Health Technol Inform. 2006;124:133-138. (Presented at MIE2006) ([draft](#))

Following the RBO track seems to be very hard, and the authors could have more discussed the rationale of this lack of “ontological purity” in the intuitive production of whatsoever content, which is probably deeply rooted in our cognitive setup and the pragmatics of human language.

- Thanks to the additional 2 pages allowed after this review, we have indeed included a discussion about this in the introduction, while moving the original introduction to a 2nd section now called ‘Background’.

RBO conformant thinking can therefore only be expected as an outcome of an intensive educational process authors of data models need to undergo.

- We agree. We added that as a suggestion in the conclusion.

The paper shows negative effects of a naïve, non-RBO approach. However, matters are not so easy given the continuous nature of biological entities, with hard questions of identity (... is this tumour still the same thing?), unity (is the primary tumour plus the nearby satellite one or two things?) and delineation (does this part of X still belong to Y).

- Thanks for these examples! They are indeed perfect to demonstrate that dealing with such situations is difficult without an RBO-approach, but can be easily dealt with within such view. We added them to the discussion where we elaborated on the distinction between diagnoses, disorders and disease courses.

Some arguments why this paper is of special relevance to the pHealth context, would be helpful.

- We added a sentence to this effect in the conclusion.

Minor comments

PERSON is not an ideal example for types, because it is disjunctive (Physical person OR Legal person, why not “Human being”?)

- We agree. We preferred to stay with the terminology OMOP uses but there is no reason to do so. We changed it, while of course still using the notion ‘person-table’ per OMOP's terminology.

Not clear why RBO is mostly on relations between types. Relations between types tend to be ontologically shallow (“A interacts_with B”) or rooted in relations between individuals (“A part of B”). The axiomatic foundation of BFO2 relies on relation between individuals. The RBO perspective should also permit axioms ranging over classes of individuals ...

- We are not sure on what basis this reviewer makes this comment. We gave several examples in the paper of particular-to-particular (PtoP) relationships and particular-to-types (PtoT)

relationships. Is it perhaps the first part of this sentence that made the reviewer frown: *'Whereas the RBO-perspective of the PoR derived from the CDM should primarily, if not exclusively, reference types and relationships between types, the RBO-perspective on the data sources translated in terms of the OMOP CDM should reference particulars and what types these particulars instantiate'*? The sentence in its totality makes clear that PtoP relations are included in the view. Anyhow, we added this sentence to clarify our point: *'This is because CDMs themselves, in contrast to the data repositories built according to these CDMs represent what is general in the domain in the same way as ontologies do'*.

... not all of which are extensions of types. E.g. PERSON, as a disjunctive class. Also classes that are currently empty might be target of the aboutness relation, e.g. "122 year old man".

- RBO does not work with disjunctive classes at all. '122 year old man' could be introduced as what is called a 'defined class' in application ontologies, but the practice is preferably to be avoided. This discussion is relevant in the larger scheme of things, but not for the purposes of this paper.

The problem identified that a provider can also be a patient could also be addressed by identifying (and counting) roles instead of humans.

- On the condition, as we suggested, that roles are clearly separated from persons, what is not the case in the current version of OMOP. Even then, counting persons is still a different process than counting roles.

"first order entity": mostly not known by the audience of this conference.

- We added, after the first time we used the term, the phrase: 'i.e. an entity that is not about something'

Minor language and formatting issues

remove dash from „data-repositories“, „Small-Caps“, „EHR-data“, „RBO-perspective“, „person-table“, „condition-era“. The mostly accepted grammar rule is that there is no dash between two nouns. In contrast, a dash is mostly set in the case of collocations of adjectives with preceding (qualifying) nouns

- done

"AMI" introduce acronym! Much of the audience are neither native English speakers nor clinicians.

- done

41001: wrong format for ICD-9. Insert period. The same with some other example ICD-9 codes.

- This is how it was literally in the source from which we copied the sentence. We changed it anyhow.

"f.i." : do you mean "e.g." ?

- Those are synonym indeed, English / latin. See: <https://www.merriam-webster.com/dictionary/f%20i>

Most papers in the RBO context have adhered to certain un-written formatting rules, such as the use of italic for classes/ types, as well as relations between classes, and bold face for individuals and relations between them. Therefore we do not need small caps.

→ We used the formatting we used in all our papers. The reason for the small caps is to highlight that in these instances we are really referencing the universal itself or the extension thereof, and not the corresponding generic term in a looser way.

→ Many thanks for this careful review.

Reviewer 2:

The paper provides insight into problems that might occur using the OMOP CDM based on scrutinizing OMOP Specifications. This contribution is extremely timely as OMOP is currently experiencing a lot of uptake. The technical work in this paper, which is based around carefully reading OMOP specifications and pointing out potential problems in representing counts, is correct.

→ Thanks, we agree.

Minor issues:

(Introduction)The reviewer proposes to specify that IHI is part of the University at Buffalo to ensure that readers do not mistake it with other institutes using the same acronym (Institute for Healthcare Improvements, I know different name, but you know readers...)

→ We did so

- *(Methods; paragraph 2, line 4): RBO's listed above. It is not clear which list you are referring to here. If the list isn't too long, listing them might be helpful to the reader.*

→ We did so, third paragraph of what is now the 'background' section.

- *(Results, paragraph 2, line 9) "unspecified, what is awkward" should be ""unspecified, which is awkward"*

→ corrected

- *The reviewer thinks it would be helpful for the reader to understand earlier in the paper (Introduction) that the review that will be presented is not an assessment of data problems based on ontological misrepresentation of counts based on data, but that what is done is scrutiny of the specifications and theoretical problems of the specification are brought to light. The reviewer holds that this work is just a useful and important as a data-driven assessment, but preparing the reader will help to prepare the authors' point.*

→ We made that clear in the last paragraph of the background section.