Pain Assessment Terminology in the NCBO BioPortal: Evaluation and Recommendations
Werner Ceusters

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

This mixed-scope paper probes Bioportal accessible terminologies on coverage in a specific area of pain terminology (9 IASP terms) and at the same time intents to evaluate the BioPortal with regard to quality schemes and policies in place to improve ontologic rigor in the proliterated artefacts and their mappings.
→ That is correct

At the moment the paper is of interest to pain-domain people on one side and ontology engineers on the other, assuming an intersection of limited numbers.
→ Should it not be a *union*?

To separate the peers maybe the pain ontology should be separately discussed in [4].
→ It is very hard to add an extra discussion section to a paper that was published three years ago. Am I missing something?

But then what remains is the discussed BioPortal critisism but some people could argue that this part in the paper is not substantial enough to warrant more than a mapping feature correction request to the Bioportal email list.
→ Nobody else argued in this way thus far.

Although the authors note that some of the Bioprrtals automatic mapping approaches are faulty in some cases, this finding was not the authors archievement.
→ Really? Whose achievement then? There was indeed the one remark, without quantification, referred to in [16] about WHO-ART, but the work here reports on 14 other sources that have similar problems.

However I like to remark that critisizing a lack of clear best practice policies in the investigated RAs seems a little unfair when at the same time the analysis itself does not separate clearly between vocabularies of different semantic expressivity (they evaluate Terminologies according to Cimino and DL ontologies according to S. Schulz, but let the results appear in the same large tables).
→ I acknowledge of course the distinction between these RAs, but the point here is rather that it does not make any sense to have them all as 'ontologies' in the BioPortal. I added this point to the paper. Thanks for the suggestion.

The coverage of IASP pain terms was not good in the smaller artefacts and only large nedical terminologies like Medra and SnomedCT covered most terms (This however doe not state that the atoms for creating terms in postcoordination are not available in smaller untangled RAs).
→ Valid point. I added this to the result section.

In general this paper follows the establshed pattern of critisizing terminologies with a 'realist' view, which is sometimes perceived as unfair as the evaluated resources were never created with DL or the OBO Foundry best practices in mind.

→ Again, it is not so much the resources that we criticize for their structure, but the presence in and representation thereof in the BioPortal. We do criticize some of them for lack of content on the basis of clearly explained principles and arguments.

Also the Bioportal is deliberately open with regards to the formal expressivity of the contained RAs. My personal feeling would be that such critisism would be more just if done on OBO Foundry ontologies. The Bioportal is not the OBO foundry and its role is NOT mainly quality assurance of the RAs made accessible. However one should also appreciate that such papers are always a good tutorial resource in making ontology editors aware of fallacies.

→ Indeed, and there are unfortunately not enough papers that do so.

The result section is hard to follow, as the tables are large and their content (key findings of them) is/are presented and discussed in various sections of the result text. For each key result a clear pointer/reference to the place and the table should be provided. As this is not easy, I recommed to source out the key findings from the large boring tables into digestible diagramms appearing nearer to the text references. As is, the result presentation is very dry. So, a portion of what is key from the Tables should be visualized graphically i.e. using (Pie Charts or Venn?) diagrams.

→ The problem is then that certain conclusions are not obvious anymore. I decided to keep the tables – except minor modifications where appropriate – as they are, but will provide such diagrams as part of the presentation which will also be publicly available after the event.

The Discussion misses a reference to the OOPS ontologic error classification (http://oeg-lia3.dia.fi.upm.es/oops/catalogue.jsp). This scheme should not be ignored in such an error analysis effort, but rather be applied and expanded where error types are missing.

→ The reference provided is – despite its claim of being about 'ontological' errors – rather about misuse of the OWL representation language, which is a very restricted and computer-science oriented view on the problem and only a tiny part of the mistakes that tend to be committed. No action taken.

The 'research question' in the discussion section "Are resources in the Bioportal intrinsically flawed" sounds a bit too general to be able to judge on a simple yes/no answer; and indeed what follows are a few rather anecdotal erros instances drawn from the various artefacts.

→ Anecdotal?

The next research question "B. Is the BioPortal itself, or are some design or quality assurrance principles behind it, intrinsically flawed?" The distinctive placement of BioPortal and Portal content in the respective result and discussion parts seem a bit arbitrary to me, e.g. why are the concrete Ontologic errors are reported in the discurrion rather than in the Result section ?

→ Each result section has its section in the discussion. I don't see any issues here.

The authors note the interesting finding that "when both the LOOM and CUI-methods suggest a mapping, the error rate increases to over 46%, thus almost the equivalent of flipping a coin". They also should elaborate on potential reasons for that. As is, this statement belongs into the Result section.

→ I am very sorry but I have absolutely no idea why this is, as of course one would expect the opposite. The reasons for that should be explained by the designers of the BioPortal, and I am not one of them.

Minor Remarks:

Sarcasm such as "for the bioportal to become an instrument which is useful for other purposes than determining that its content is of poor quality, ..." should be removed.
→ I believe a bit of humor has a place in bringing sad news. No action taken.

In general the author does not go beyond the rather flat model of pain as used by the IASP. There could however be valuable ontological connections to be explored, i.e. the purpose of pain for the ontogeny of an organism, i.e. triggering damage avoidance and learning behaviour, conditioning, its pure signalling function, ... In this systemic interpretation of pain we leave the purely sensoric/materialistic/physiological interpretation level and enter a more systemic/functional/psychiological view. Here also Ghost pain should be discussed. Then of cause one could add emotional pain, i.e after the loss of an close relative, but all this should be part of the authors separate paper [4].
→ That is indeed discussed there and has no bearing with this paper.

in Fig 1 use bold for IASP term Superclasses. Add legend or state what dotted edge means.
→ Bold used. Dashed arrow removed.

Table 2-6 should have a table caption to support interpretation of the table structure, i.e. explaining and resolving Acronyms, what is the differrence between n/a and blank fields? What are 'foreign Classes',...
→ To save space, we added the missing explanations in the main text. 'n/a' was removed.

English: "Each class was on the basis of its preferred term (with disambiguation where required as for instance for analgesia) classified into one of these groupings." correct style.
→ I reformulated.

Bullet 3 under Results/A. Quality of BioPortal Resources Retrieved, states "11 resources provide textual definitions for at least some of the classes (AP2, AP4)--> Why AP2 ?
→ If a class is missing – what can be assessed in AP2 – it has of course no definition (AP4). Highlighting AP2 in addition to AP4 prevents wrong interpretations. Take CRISP as an example: AP4 alone tells you that only 2 CRISP classes have definitions. Through AP2 you can see that CRISP has only 2 of the 9 classes. Thus, all classes that CRISP has, are defined. It allows to make further conclusions such as the % of defined classes, which for CRISP would be 100%.

Header "B. Adequacy of the OBO BioPortal" should be "B. Adequacy of the NCBO BioPortal"
→ That has been corrected.

in the following text section on Tab.5 the method descriptions should be sourced out into the Methods section and the Table abbreviations and Acronyms (t, b, S?) belong into a Table caption, so that the result section gains in readability.
→ 'Some people' might argue differently. Main issue is the page limitation.

----------------------- REVIEW 2 --------------------
OVERALL EVALUATION: -1 (weak reject)
REVIEWER'S CONFIDENCE: 4 (high)

The paper addresses some important general issues in the ontology world, and specifically some issues in the Bioportal resource. Taking a defined set of concepts concerned with "pain" in the sense of clinical concepts, mainly abnormal neurological phenotypes assessed by clinical symptoms, the author discusses the problems in generating a classical subsumption hierarchy based on expert definitions of different pain concepts.

→ I don't do anything of this sort, unless you mean Fig 1 which is designed to make a point about the inadequacy of 'expert definitions'. I did not generate this hierarchy, but, of course implicitly, the 'experts' did.

One problem here is that he conflates syndromes and phenotypes and so of course this generates problems in generating a well-formed hierarchy.

→ I don't conflate anything. Are you serious? The critique I usually get is exactly the opposite, that I see too many distinctions.

It would have been better to consider all of the 29 IASP terms rather than just the selected ones he presents.

→ This reviewer should explain how more definitions for other terms would be able to change anything about the inadequacy of the 9 definitions covered here. It would make matters even worse, though of course provide more insight in why the definitions do not make sense taken altogether. That is covered in another paper.

He then looks for how these concepts are treated in ontologies contained within the Bioportal and assesses the quality of inter-ontology mappings, subsumption and structure of the ontologies containing the concepts. He concludes that the quality of several ontologies is very poor and that almost all are incomplete in that they do not contain all terms considered to be needed to describe different types of pain within the same ontology. He also concludes that Bioportal does not, in several instances, represent the ontology correctly and raises justified criticism about the quality of the ontologies themselves and maintenance of the Bioportal resource and its consequent utility.

→ That is correct.

In many ways the paper is trivial in that the methods used are simple and provide no metrics other than counts. This could have been much more general and sophisticated and consequently one can only regard this as a preliminary publication following a rather "back of the envelope" study.

→ Seriously? I wonder how this reviewer would fit the data file I made available (ref [8]) to check my claims in this paper on the back of an envelope.

It does raise some serious concerns, many well known. He underestimates the huge burden of errors of the type he describes, as a consequence of failure of ontology developers to engage experts and their use of computational methods, particularly lexical matching, to generate ontologies from other resources. In many ways these are chickens coming home to roost.

→ Having designed one of the first large scale formal ontology authoring system myself and having written numerous papers about this and related topics, I don't underestimate anything. I rather question the appropriateness of the NCBO BioPortal to do what it does: offering a platform for

junk without doing anything about it, since its criteria for success is being measured in how many systems are represented in it. See, for instance ref [17]: Brochhausen M, Burgun-Parenthoine A, Ceusters W, Hasman A, Leong TY, Musen M, Oliveira J, Peleg M, Rector A, Schulz S. Discussion of "Biomedical Ontologies: Toward Scientific Debate", Methods of Information in Medicine, 2011;50(3):217-36.

It is extremely difficult to detect and correct these errors in content computationally and only experts are in a position to judge them, although the type of egregious error he points out (which will be the tip of the iceberg), might be apparent to a generally well read biomedical scientist.The structural problems and mapping problems are again in many ways often down to poor content knowledge, or legitimate differences of opinion, and many of these would be picked up by expert audit and would then be easy to correct algorithmically. It is true that the kind of approach he suggests might assist in finding some classes of error and help in "semi-automated" ontology audit.

→ Amen

The issue of how to structure a "pain " ontology was not really resolved and to an extent left hanging in the paper.

→ That was not the topic of the paper. You can't reject a paper because you prefer the author to have written another paper on a distinct topic. Furthermore, work to that effect was written three years ago, and referred to: [4]

More important is the finding that Bioportal is misrepresenting some of the ontologies. This is very serious but the author does not attempt to look at this problem systematically or apply any detailed metrics.

→ I would be happy to do so when funded for this purpose.

This clearly needs to be done. What he does not assess is how these errors affect his conclusions about mappings and content, as from my reading he has only used Bioportal and not compared the representation of these ontologies with independent source files.

→ That is correct. See previous remark.

The question of whether Bioportal could or should act as Gatekeeper for its ontologies is a fraught one. The investment of effort in such an exercise would be immense and without expert domain knowledge i can see a huge number of problems arising due to differences of opinion about strategies for ontology structure. This is not considered in this discussion.

→ Such issues are discussed in other papers, for instance ref [17].

In general public databases carry with them a "caveat emptor' warning and unless the development of all the Bioportal ontologies were to be actively coordinated (impossible) I cannot see any way around this.

→ I agree, except for the 'impossible', and that is why the OBO Foundry is created. The problem would be solved if developers would join the Foundry. I made this suggestion in the conclusion and recommendations section.

One must also asks where the imposition of a "party line" on ontology construction is justified or indeed would be scientifically defensible - the author does allude to this problem in his discussion of the "realist" approach to ontology development. The example given is something of a straw man and no-one " outside the Foundry" would possibly consider this example to be good practice despite differences of opinion about foundational ontologies etc.

→ We referred the reader to [17] where positions pro- and con realism are provided.

Specific issues:

The IASP terms (the IASP taxonomy) actually form an unstructured defined lexicon of many different types of entity. The reference to the current online version should be given.

→ That was already the case: ref [3]

Table 1 in the paper is not a summary of these terms but a subset. This is not accurately described in the paper.

→ We nowhere wrote this table is a summary. Point made by the reviewer unclear.

2. Logical types are confused in some cases - for example pain therapy as a preferred term in table 2 is listed for anaesthesia. It is clear that there are two senses of the term, one a procedure and the other a reported sensory deficit - why do they appear like this together?

→ I don't discuss logical types, and have no idea what the reviewer has in mind here. Distinct senses have nothing to do with logic. There is also no confusion here. The point is precisely that some sources use preferred terms which have two senses – thus not have face value which is a violation of good terminological practice. Table II and the respective discussion explains that. Furthermore, in the paper it was also stated that terms with more than one sense were disambiguated on the basis of what the resources were intended for. So we did not count for an individual resource the term 'anaesthesia' to be problematic if that resource covered exclusively procedures, or exclusively symptoms. (see ref [8]). We did also not, as explained in the methodology section, include a mapping as a mistake when we could not disambiguate one of the two mapped terms, what led us to claim that our error counts might actually be upper bounds, and the situation even more dramatic.

3. In general legends would greatly improve understanding of the tables.

→ Agreed, but unfortunately, the 6-page limit is problematic. That is why legend information is interwoven in the text.

4. Specific illustrations should be given of the different types of error in table V. Were these expertly determined where the assessment suggests "WRONG"?

→ An exhaustive list of all instances for all error types was provided in the 'back of the envelope' data file (ref [8]). I will provide examples therefrom in the presentation. The procedure concerning how the types of error and other categories were assigned, was explained in section IIIb.

5. Table 6 is difficult to understand and needs more careful description.

→ Also the explanation is found in IIIb.

6. English needs some attention in places.

→ Has been taken care of.