The Evaluation of Ontologies: Editorial Review vs. Democratic Ranking

Barry Smith
Department of Philosophy, Center of Excellence in Bioinformatics and Life Sciences, and National Center for Biomedical Ontology
University at Buffalo

from *Proceedings of InterOntology 2008* (Tokyo, Japan, 26-27 February 2008), 127-138.

ABSTRACT. Increasingly, the high throughput technologies used by biomedical researchers are bringing about a situation in which large bodies of data are being described using controlled structured vocabularies—also known as *ontologies*—in order to support the integration and analysis of this data. Annotation of data by means of ontologies is already contributing in significant ways to the cumulation of scientific knowledge and, prospectively, to the applicability of cross-domain algorithmic reasoning in support of scientific advance. This very success, however, has led to a proliferation of ontologies of varying scope and quality. We define one strategy for achieving quality assurance of ontologies—a plan of action already adopted by a large community of collaborating ontologists—which consists in subjecting ontologies to a process of *peer review* analogous to that which is applied to scientific journal articles.

1 From OBO to the OBO Foundry

Our topic here is the use of ontologies to support scientific research, especially in the domains of biology and biomedicine. In 2001, Open Biomedical Ontologies (OBO) was created by Michael Ashburner and Suzanna Lewis as an umbrella body for the developers of such ontologies in the domain of the life sciences, applying the key principles underlying the success of the Gene Ontology (GO) [GOC 2006], namely, that ontologies be (a) *open*, (b) *orthogonal*, (c) *instantiated in a well-specified syntax*, and such as (d) *to share a common space of identifiers* [Ashburner et al. 2003]. The OBO library now comprises over 60 ontologies, and its role as an ontology information resource is supported by the National Center for Biomedical Ontology (NCBO) through its BioPortal [Rubin et al. 2006].

In 2005, the developers of a subset of OBO ontologies initiated the OBO Foundry, a collaborative experiment based on the voluntary acceptance by its participants of an evolving set of principles extending those of the original OBO (Table 1), and designed to maximize the degree to which ontologies can support the needs of working scientists. The OBO Foundry represents a strategy for coordinated reform of existing OBO ontologies, and for the creation of new ontologies, on the basis of shared principles governing ontology development [Smith et al. 2007]. What follows is a personal view of this strategy, though I believe that it comes close to the positions accepted by all of those involved.

Table 1: OBO Foundry Principles (Version as of Dec. 2007)

1. The ontology must be *open* and available to be used by all without any constraint other than (a) its origin must be acknowledged and (b) it is not to be altered and subsequently redistributed under the original name or with the same identifiers.
2. The ontology is in, or can be expressed in, a *common shared syntax*. This may be either the OBO syntax, extensions of this syntax, or OWL.
3. The ontologies possess a *unique identifier space* within the OBO Foundry.
4. The ontology provider has procedures for identifying distinct successive *versions*.
5. The ontology has a clearly specified and clearly *delineated content*.
6. The ontologies include textual *definitions* for all terms.
7. The ontology uses relations which are unambiguously defined following the pattern of definitions laid down in the *OBO Relation Ontology*.
8. The ontology is *well documented*.
9. The ontology has a plurality of independent *users*.
10. The ontology will be developed *collaboratively* with other OBO Foundry members.

2 Models of Good Practice

The OBO Foundry provides models of good practice in ontology development along a number of axes, conformity to which is serving as a framework for the evaluation of ontologies submitted to the Foundry. Foundry ontologies share a common top level ontology—effectively, that of Basic Formal Ontology—which divides entities into the basic categories of *continuants* (which endure through time) and *occurrents* (which unfold themselves in time) [Grenon et al. 2004]. In addition, a distinction is made between *independent* and *dependent* continuants, which is to say between the things (e.g., molecules, cells, organisms) and the qualities and functions which inhere in (depend on) these things. This threefold division is illustrated, for example, in the tripartite division of the GO into Cellular Component, Molecular Function, and Biological Process ontologies. The Foundry also provides guidelines in the formulation of definitions, where the model of good practice is the Foundational Model of Anatomy (FMA), a representation of types of anatomical entities built around two backbone hierarchies of *is_a* and *part_of* relations [Rosse and Mejino 2007].

Each Foundry ontology forms a graph-theoretic structure, with terms connected by edges representing relations like *is_a* or *part_of*, in assertions such as, "serotonin *is_a* biogenic amine" or "cytokinesis *part_of* cell proliferation." Because relations in OBO ontologies were initially used in inconsistent ways, the OBO Relation Ontology (RO) was developed to provide guidelines to ontology builders in the consistent formulation of relational assertions [Smith et al. 2005].

List of OBO Foundry Ontologies (as of December 2007)

1. Mature ontologies undergoing incremental reform
- Cell Ontology (CL)
- Gene Ontology (GO)
- Foundational Model of Anatomy (FMA)
- Zebrafish Anatomical Ontology (ZAO)

2. Mature ontologies still in need of thorough review
- Chemical Entities of Biological Interest (ChEBI)
- Plant Ontology (PO)
- Sequence Ontology (SO)

3. Early versions exist
- Common Anatomy Reference Ontology (CARO)
- Environment Ontology (EnvO)
- Infectious Disease Ontology (IDO)
- Ontology for Biomedical Investigations (OBI)
- Ontology for Clinical Investigations (OCI)
- Phenotypic Quality Ontology (PaTO)
- Protein Ontology (PRO)
- Relation Ontology (RO)
- RNA Ontology (RnaO)

3 The Process of Peer Review

The OBO Foundry is designed to be a community resource of public domain ontologies in the health and biological sciences. Those involved in the Foundry initiative have agreed, in advance, to adhere to a jointly developed, evolving set of principles that exemplify good practice in ontology development and use.

Each OBO ontology is subject to a process of peer review, the reviews being carried out by two types of editors: *Coordinating Editors*, whose primary responsibility is harmonizing interactions between ontology development projects, and *Associate Editors*, who are the editors of the ontologies already accepted for inclusion within the Foundry. In addition, ad hoc reviewers with special expertise are included in the reviewing process as occasion demands.

As with traditional journal submissions, this will be an iterative process, with recommendations for revisions that are published openly on the web, and being addressed in successive versions of the ontology until a stage is reached where the ontology is deemed suitable for inclusion in the Foundry. Thereafter, the ontology will be subject—like all other evidence-based ontologies—to successive updates in order to keep pace with the advance of scientific knowledge. One criterion for review is that the authors of the ontology have a process in place to support necessary changes.

One principle adopted by the Foundry is that of orthogonality, which asserts that for each domain of biomedical research there should be convergence upon a single ontology within the community of those involved with the Foundry initiative. This is, for example, to provide maximally useful assistance to new users of ontologies, who need to know where to look in finding an ontology relating to a specific subject-

matter which has already been validated to work well with other established ontologies. The principle serves also to ensure a division of labor, whereby experts in each domain will be responsible for creating a common, high-quality controlled vocabulary, for that domain, for use by all those who wish to take advantage of the resources of Foundry ontologies. This does not mean that Foundry members hold that all ontology development should take place within the Foundry. Indeed, those involved with the initiative are fully aware that scientific advance rests on the existence of competing strategies for development, as well as on a constant to-and-fro of criticism and exchange between the representatives of these different strategies. At the same time, however, given its commitment to the scientific method, the Foundry is committed to taking on board all relevant results of scientific investigation, independently of the strategy which was used to achieve them. Hence, all those advances in both ontology content and methodology proving fruitful to scientific advance will, in course of time, be adopted by the Foundry, and even the Foundry principles are modifiable in light of new learning gained from experience of what works in practice.

4 Pros and Cons of the Foundry Method of Evaluation

The editorial process of the OBO Foundry applies only to ontologies developed in support of scientific research or in support of activities closely related thereto, as for example in the domain of clinical medicine. This methodology may be less salient, say, where one is dealing with the administrative ontologies developed by industrial or government agencies and where different conditions apply.

Peer reviews are widely used by editors of scientific journals and funding agencies to ensure scientific quality. This method of evaluation is applicable in any given scientific domain provided that (i) the scientific community agrees about the criteria for good science in their domain, (ii) the peer reviewers adhere to these criteria, and (iii) the peer reviewers are trusted by the community.

The hypothesis is that, because the ontologies at issue within the Foundry are intended to serve as contributions to, and to support advances in, scientific knowledge, they can be subjected to the same editorial processes as scientific journal articles.

In favor of the Foundry methodology for evaluation, it can be said that this methodology already has proven successful, even though only in the testing stage. Currently, there are some 20 ontologies within the Foundry, covering a vast range of biomedical topics. Several more are under review. The evaluation process involves a large scientific community, which contributes to the widespread use of the ontology and encourages cooperation between members of a community with similar interests. The peer review process also matches well with the systems of rewards and of hiring and promotion still operative in the scientific realm (and, it must be said, with the motivations of traditional publishers). Finally it jibes well with a certain philosophical view embraced by many in the natural sciences, according to which science consists in the search for truth and that there is a truth to be found (however difficult the search might be). This view is consistent with the acceptance that many aspects of science—including many aspects of terminology use—are a matter of convention or

fiat. But it departs from postmodernist or relativist views according to which the whole of science is a matter of convention, or that science reflects merely the whims of fashion.

Against the Foundry methodology, it can be argued that:

(1) the review process relies too narrowly on human experts and is thus both expensive and open to questions concerning scalability.
(2) by making the evaluation process open, a key benefit of double blind reviews is lost. The behavior of people in publicly accessible forums is influenced by their social status, personal relations, and political considerations.
(3) ontologies are not like journal articles, and it is difficult to judge methodological soundness simply by inspection.
(4) the evaluation process does not yield a quantifiable result.
(5) because the evaluation process does not happen at regular intervals and does not lead to quantifiable results, this method does not provide certain sorts of information, for example, as to how much an ontology improved in the last three months, which may be useful for the management of ontologies.

In response to (1) and (2), the Foundry editors are taking steps both to increase the degree to which computational metrics can be applied, as well as to diminish the possibility for peer pressure and similar factors to influence the outcomes of Foundry review. In response to (3) and (4), it should be noted that the peer review process employed by journal articles faces the similar task of judging methodological soundness simply by inspection, and that such peer review does not yield quantifiable results. We believe that ontologies, like other contributions to the theoretical resources of science, will always stand in need of evaluation by human experts, though we recognize that their work will increasingly be supported by software tools. In response to (5), we are developing sophisticated metrics designed to provide answers to questions concerning improvement of ontologies over time [Ceusters and Smith 2006].

5 Democratic Ranking as Alternative to Scientific Peer Review

In recent work, Musen, Guha, and Noy have advocated an alternative approach to ontology evaluation based on a democratically inspired ranking system analogous to that used by Amazon.com, eBay and similar bodies [Musen 2007], [Noy et al. 2005]. In defense of this approach, Musen and his colleagues point out that users of ontologies may well have ontology needs quite different from those taken account of by the Foundry editors. The utility of ontologies depends on the tasks for the performance of which they were constructed to a much higher degree than journal articles, and these tasks may be highly context-specific.

Against this, we note, first, that in relation to the five points listed above the democratic ranking methodology, too, faces analogous problems: it, too, relies on humans; it, too, is open; rests in large degree on inspection; and does not yield a quantifiable result (or at least: it does not yield any single quantifiable result); and finally it, too, because of its open-ended nature, would cause problems for the management of ontologies.

It is relevant, too, that the OBO Foundry does not seek to stop others from using and developing non-Foundry ontologies in addressing different needs. Indeed, as already stated, the members of the Foundry only see benefit from the development of alternative strategies by other groups. At the same time, the orientation of the Foundry around evidence-based ontologies developed in the spirit of the scientific method implies a certain non-locality and context-independence of the ontologies with which it deals. There is no private or local science, and if controlled vocabularies are developed by scientists to address the specific local purposes of specific groups without consideration of the needs of the broader scientific community, then the results of their work will, at best, make a diminished contribution to the cumulation of scientific knowledge.

In essence, the democratic approach of Musen and his colleagues addresses the same goals as those addressed by the Foundry editorial process. Both seek a particular kind of quality assurance when it comes to ontology selection. Musen, Guha, and Noy think that quality assurance can be provided reliably through user rankings. The allies of the Foundry initiative prefer to rely on a rule-driven editorial process.

Both sides need to address the fact that the expertise and biases of reviewers may vary widely with respect to different ontologies or to different portions of an ontology. In defense of democratic rankings, it can be said that ranking by large numbers of users will tend to counteract such biases (though it must also be said that we have, as yet, no way of knowing to what degree the ranking facilities will, in fact, be taken advantage of by users). In defense of the Foundry, it can be pointed out that problems of bias and differential expertise affect all editorial reviews. However, our systems of peer review seem, nonetheless, to have played a significant positive role in the advance of science and in the provision of guidance to successive generations of scientists as to what scientific claims should be taken with what degree of seriousness.

The process of democratic ranking as conceived by Musen and colleagues would incorporate, not merely numerical ratings of ontology quality, but also qualitative evaluations, some of which might look very much like editorial reviews, and some of which might indeed incorporate applications of OBO Foundry principles. Musen, Guha, and Noy, here, defend an intriguing vision of a new type of peer review, consisting of what they call ‘community-based annotation’ of ontologies. The vision rests on the assumption that the application of ontologies by actual users will lead to insights not achievable by inspection alone, and that users of the ranking service are likely to find such insights especially valuable to their work. Indeed, Musen and his Bioportal colleagues assert that users ‘should want the opinions of more than 2-3 hand-selected reviewers’, and ‘Peer review needs to scale to the entire user community’ [Musen 2007].

Here, again, there is considerable convergence between the two schools of thought. OBO Foundry reviewers will themselves be users of ontologies and will bring precisely the sorts of focused insights that other users will find useful. An important part of the rationale for the Foundry’s open process of reviewing is that users and developers of ontologies can learn important lessons from the reviewing process itself. If this process did not take account of such focused insights, then the community which the Foundry serves would quite rightly demand countervailing adjustments in the way reviews were being carried out.

There is an important difference between the two approaches, however, which turns on the fact that the relevant commentaries generated by the Foundry review are created precisely by the *peers* of the ontology authors themselves—by persons with established and recognized expertise and with a demonstrated willingness to invest due diligence in ontology development, use, and evaluation. We believe that the insights generated by this reviewing process will, therefore, have an enhanced likelihood of being of value and utility to outsiders. In the case of the democratic rankings as conceived by Musen and colleagues, in contrast, no restrictions at all would be placed on who can contribute to the bodies of textual commentary on ontologies which will arise through the ranking process. Musen *et al.* talk of creating a 'web of trust' which would be designed to enable a user to filter out certain comments or ratings in such a way as to avoid what for him or her will count as 'noise' within the system. It is possible that such a web of trust can provide genuine utility in bringing to light what is truly useful. But again, we have, as yet, no way of knowing to what degree this facility could be used to isolate insights of value to the working scientist. Indeed, we have no way of knowing to what degree the democratic ranking and reviewing facilities as a whole will be taken advantage of by those experts whose commentaries might be most likely to yield insights of this sort.

The mutual ranking of buyers and sellers carried out by eBay is of obvious value to all involved, and the same goes for Amazon.com-style ranking in domains such as that of home electrical items like refrigerators, where consumer reviews and rankings not only play an important role in the decisions of purchasers, but are taken into account also by the manufacturers of products. But these manufacturers, of course, employ *in addition* the services of expert reviewers when designing and marketing their particular products. In some domains, therefore, it seems that there is room for both kinds of approaches. It is however difficult for us, at this stage, to conceive of a role for democratic ranking—and of a strategy which embodies no demarcation between experts and non-experts—in the domain of scientific research. In the spirit of scientific inquiry, however, we strongly welcome all efforts to prove us wrong empirically.

Acknowledgements

This work was funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant 1 U 54 HG004028. With thanks to Robert Arp and Mitsu Okada for helpful comments, and also to the participants at the workshop on Methods and Metrics for Ontology Evaluation, sponsored by the National Institute of Standards and Technology and the National Institutes of Health in Gaithersburg, Maryland, October 25-26, 2007.

References

[Ashburner et al. 2003] M. Ashburner, C. Mungall and S. Lewis, Ontologies for Biologists: A Community Model for the Annotation of Genomic Data, *Cold Spring Harbor Symposia on Quantitative Biology* 68 (2003), pp. 227-236.

[Ceusters and Smith 2006] W. Ceusters and B. Smith, A Realism-Based Approach to the Evolution of Biomedical Ontologies, *Proceedings of the Annual AMIA Symposium* 94 (2006), pp. 121-125.

[GOC 2006] The Gene Ontology Consortium, The Gene Ontology (GO) Project in 2006, *Nucleic Acids Resources* 34 (2006), pp. 322-326.

[Grenon et al. 2004] P. Grenon, B. Smith, and L. Goldberg, "Biodynamic Ontology: Applying BFO in the Biomedical Domain. In D. Pisanelli, ed. *Ontologies in Medicine*. IOS Press, 2004, pp. 20-38.

[Musen 2007] M. Musen, If We Build It, Will They Come? Available at: http://ontolog.cim3.net/cgi-bin/wiki.pl?ConferenceCall_2007_12_06.

[Noy et al. 2005] N. Noy, R. Guha, and M. Musen, User Ratings of Ontologies: Who Will Rate the Raters? *Proceedings of the AAAI 2005 Spring Symposium on Knowledge Collection*. Available at: http://smi.stanford.edu/smi-web/research/details.jsp?PubId=1112.

[Rosse and Mejino 2007] C. Rosse and J. Mejino, The Foundational Model of Anatomy Ontology. In A. Burger, D. Davidson, and R. Baldock, eds. *Anatomy Ontologies for Bioinformatics*. Springer, 2007, pp. 59-117.

[Rubin et al. 2006] D. Rubin et al., The National Center for Biomedical Ontology: Advancing Biomedicine through Structured Organization of Scientific Knowledge, *OMICS* 10 (2006), pp. 185-198.

[Smith et al. 2005] B. Smith et al., Relations in Biomedical Ontologies, *Genome Biology* 6 (2005), R46.

[Smith et al. 2007] B. Smith et al., The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration, *Nature Biotechnology* 25 (2007), pp. 1251-1255.

>

Mark Musen, below, makes a number of valuable points, which are made all the more interesting in virtue of the fact that the NCBO's BioPortal, an ontology repository for which Mark is responsible, http://www.bioontology.org/bioportal.html, is carrying out an experimental test of the benefits of democratic ranking-based approach to ontology assessment.

Specifically, the BioPortal will test a thesis to the effect that democratic ranking based on user comments can 1. provide a valuable service which will scale as the population of ontologies grows and 2. allow true openness (no gatekeeping at all) of a repository (thus perhaps even allowing admission to the BioPortal of http://www.schemaweb.info/schema/SchemaDetails.aspx?id=163, which is, as I understand it, a bio-ontology-like

artifact pertaining to organisms with more than two legs).

However, his main argument against the alternative (expert peer review-based) approach currently being tested by the OBO Foundry, has been addressed already in earlier postings to this list: the committee of peer reviewers used by the Foundry will in every case involve expert users from the specific user communities.

Mark thinks that ontologies are much more like refrigerators than they are like journal articles. I think that most of them are in fact still much more like collections of refrigerator magnets. The question is: how can we motivate ontology creators (and potential ontology evaluators) to do a better job?

This question is also not addressed by Holger Lewen, in another interesting and useful post that is also appended below. Indeed, Holger expresses a touching optimism to the effect that large bodies of intelligent user comments will form around the ontologies submitted to a potential OOR; that software will allow potential new users to navigate through these comments to help them find the answers to just the questions they need; and that intelligent evaluators will keep on submitting new comments as ever new collections of refrigerator magnet products (sorry: ontologies) come onto the market.

Why should they invest this time and effort? Skilled users of ontologies are, I can imagine, busy people. They also form a rather small community, with limited resources to spend e.g. on training teams of document inspectors as proposed (also below) by Matthew.

The OBO aims to test one potential answer to this

motivation question, at least for ontologies developed to aid science. This answer has the advantage of resting on a methodology -- the methodology of peer review -- that has enjoyed some three hundred years of success that is roughly co-terminous with the advance of modern science. .

Crudely put: experts are motivated to review ontologies in their relevant domains of expertise because they get career-related credit for serving as reviewers.

Ontology developers are motived to create better ontologies because they get career-related credit for having their ontologies included (published) in what, if the peer-review process is successful will count as analogous to a peer-reviewed scientific journal. (We are working on the many tough problems which must be tackled to make this possible -- including all the problems mentioned by Mark and Holger, below.) The publishing process we have in mind will have the incidental advantage that it will allow the multiple developers typically involved in serious ontology endeavors to get appropriate credit, which they can use to justify spending the time and effort involved.

Both reviewers and developers will be further motivated to participate in this process because they can thereby directly influence the standard set of ontology resources which will be available to them, thereby also motivating the creation of: related ontology-based software, useful bodies of data annotated in terms of these resources, etc.

Note that I am not recommending this as an approach to be adopted by the OOR. It rests on too many features peculiar to the domain of science. However, if Patrick Hayes is right - that people like him can just as well publish

their ontologies on the web - then this suggests
the need for a real raison d'꽪e for the OOR,
and I am suggesting non-trivial and evolving
gatekeeping constraints in the cause of
incrementally raising the quality of ontologies as one such raison d'꽪e.

BS

BS

At 01:08 AM 3/21/2008, Mark Musen wrote:
>On Mar 20, 2008, at 8:56 PM, John F. Sowa wrote:
> > There are two independent issues here: reviewing and publishing.
> > Everybody would agree that reviewing is important, but ideally,
> > the readers/users should have the option of making their own
> > choices based on the reviews. When publication was expensive,
> > the publishers became gatekeepers because it was economically
> > impractical to publish everything.
>
>The analogy between peer review of journal articles and peer review of
>ontologies has been applied too glibly, I believe.
>
>The best reviewers of a journal article are scientists who can
>evaluate the methods described in the paper, judge whether the data
>presented are plausibly consistent with the methods, and assess
>whether the authors' interpretations of the data are reasonable. This
>process is all done rather well by scientists who are experts in the
>field and who can understand the work that is described in the paper.
>Although the system does break down, sometimes in notorious ways, it
>generally works rather well.
>
>Ontologies are not journal articles. Although there are many surface-
>level distinctions that can be assessed purely by inspection (OBO-
>Foundry criteria regarding representation language, namespaces,
>textual definitions, and so on), the key question one wants answered
>before using an ontology concerns whether the ontology makes the

right
>distinctions about the domain being modeled. This question cannot be
>answered by inspection of the ontology; it can be answered only by
>application of the ontology to some set of real-world problems and
>discovering where things break down. The people best suited for
>making the kinds of assessment that are needed are not necessarily the
>best experts in the field, but the mid-level practitioners who
>actually do the work. Any effective system of peer review has got to
>capture the opinions of ontology users, and not just those of renowned
>subject-matter experts or of curators.
>
>I think ontologies are much more like refrigerators than they are like
>journal articles. I view ontologies as artifacts. Not surprisingly,
>I am much more interested in the opinions of people who actually use
>refrigerators than I am of experts in thermodynamics, product
>manufacturing, or mechanical engineering. The latter are people who
>can inspect a particular refrigerator very carefully for surface-level
>flaws, but who may have no first-hand knowledge of what happens when
>you actually plug it in.
>
>Mark
> Delivered-To: phismith@mailspool09.dyn.acsu.buffalo.edu
Received: (qmail 20069 invoked from network); 21 Mar 2008 12:10:24 -0000
Received: from unknown (HELO mailscan1.acsu.buffalo.edu) (128.205.6.133)
  by mail3 with SMTP; 21 Mar 2008 12:10:24 -0000
Received: (qmail 23886 invoked by uid 26149); 21 Mar 2008 12:10:23 -0000
Delivered-To: phismith@buffalo.edu
Received: (qmail 23679 invoked from network); 21 Mar 2008 12:10:22 -0000
Received: from mccarthy.cim3.com (64.62.192.10)
  by front3.acsu.buffalo.edu with SMTP; 21 Mar 2008 12:10:22 -0000
Received: from mccarthy.cim3.com (localhost.localdomain [127.0.0.1])
      by mccarthy.cim3.com (Postfix) with ESMTP id B5E07108B62;
      Fri, 21 Mar 2008 05:09:20 -0700 (PDT)
X-Original-To: ontology-summit@ontolog.cim3.net

Delivered-To: ontology-summit@mccarthy.cim3.com
Received: from moutng.kundenserver.de (moutng.kundenserver.de [212.227.126.187])
by mccarthy.cim3.com (Postfix) with ESMTP id 6901A108B62
for <ontology-summit@ontolog.cim3.net>;
Fri, 21 Mar 2008 05:09:16 -0700 (PDT)
Received: from [192.168.2.37] (p5B3BE262.dip.t-dialin.net [91.59.226.98])
by mrelayeu.kundenserver.de (node=mrelayeu4) with ESMTP (Nemesis)
id 0ML21M-1Jcg3c3gtd-0003pk; Fri, 21 Mar 2008 13:09:14 +0100
Message-Id: <426E7162-BB50-46B5-9138-6A6FDA832D20@aifb.uni-karlsruhe.de>
From: Holger Lewen <hle@aifb.uni-karlsruhe.de>
To: Ontology Summit 2008 <ontology-summit@ontolog.cim3.net>
In-Reply-To: <808637A57BC3454FA660801A3995FA8F06A2D3FC@lonsc-s-031.europe.shell.com>
Mime-Version: 1.0 (Apple Message framework v919.2)
Date: Fri, 21 Mar 2008 13:09:12 +0100
References: <808637A57BC3454FA660801A3995FA8F06A2D3FC@lonsc-s-031.europe.shell.com>
X-Mailer: Apple Mail (2.919.2)
X-Provags-ID: V01U2FsdGVkX1+R3ClbCATiOgkDQJFy18mmqWctdIOMUbJ72r9
QE6P9VrSZh2z7Ulgng1GZbcetJRrDotabFTYuo9bIcMu02snoG
dPmpRMf7I56QBJktSJaMGlJEBXSfzrQ
Subject: Re: [ontology-summit] [Quality] What means
X-BeenThere: ontology-summit@ontolog.cim3.net
X-Mailman-Version: 2.1.8
Precedence: list
Reply-To: Ontology Summit 2008 <ontology-summit@ontolog.cim3.net>
List-Id: Ontology Summit 2008 <ontology-summit.ontolog.cim3.net>
List-Unsubscribe: <http://ontolog.cim3.net/mailman/listinfo/ontology-summit>,
<mailto:ontology-summit-

request@ontolog.cim3.net?subject=unsubscribe>
List-Archive: <http://ontolog.cim3.net/forum/ontology-summit>
List-Post: <mailto:ontology-summit@ontolog.cim3.net>
List-Help: <mailto:ontology-summit-request@ontolog.cim3.net?subject=help>
List-Subscribe: <http://ontolog.cim3.net/mailman/listinfo/ontology-summit>,
<mailto:ontology-summit-request@ontolog.cim3.net?subject=subscribe>
Content-Type: text/plain; charset="iso-8859-1"
Sender: ontology-summit-bounces@ontolog.cim3.net
Errors-To: ontology-summit-bounces@ontolog.cim3.net
X-UB-Relay: (mccarthy.cim3.com)
X-PM-Spam-Prob: : 7%

Dear Colleagues,

after having followed the quality discussion for quite some time now, I am glad to see that the majority of people seem to agree that peer review of ontologies can provide value and submission to an "open" system should not be too limited.

Assuming one would decide to have an Open Rating System as the basis
for peer review, as was already proposed in the literature, most of the points raised in the discussion could be accommodated.

Since everyone can write reviews about the ontologies, some of the reviewers can (and should) be what Barry would consider gatekeepers in
the restricted scenario. Namely experts that offer their opinion on certain aspects of an ontology in the system. The way the Open Rating System works, users can then decide which reviewer to trust and get their ontologies ranked accordingly.

Not only does this approach scale (everybody can review), it is also very personalizable. It is up to the user to decide whether she values the opinion of a "mere ontology user" more than the opinion of an "ontology expert". As was already pointed out, the ontology user can

provide feedback about actually working with the ontology, while the expert might just look at the ontology from a theoretical point of view and determine the usefulness based on that without even considering runtime implications.

One critique often raised when proposing this kind of solution is: Who will provide the input, who will review the ontologies and who is even able to review ontologies. While I certainly agree that reviewing ontologies is harder than reviewing consumer products, there seem to be a group of people that are knowledgeable enough for Barry to consider them part of his gatekeeping committee. If the only contribution of the rating system were to have their process of assessing submitted ontologies public, i.e. each expert writing a review based on his context as philosopher, computer scientist or scientist, I claim there is a benefit.

As several of you have already mentioned, one problem with restricted reviewing systems is that they are very vulnerable to personal preferences, prejudices and reviewer's egos. Also controversial ideas sometimes are not allowed because n people decide they are not worth publishing. I would gladly appreciate a peer review system that at least makes the reviews of papers with all submitted papers accessible. Then I could make my own decision of whether a review might have been biased or otherwise subjective, and whether I want to read a controversial paper.

I do not want to bore you with all the details, so in short my claims are:
-Open Ratings provide more transparency
-Open Ratings allow user personalization of ranking order based on trust in reviewers
-The reviews can and should come also from the people that are now thought of as potential gatekeepers
-This allows for a much wider exploration of the usefulness of an ontology in different scenarios, because people can provide reviews based on each specific setting
-The gatekeeping approach cannot scale beyond a certain number of new
ontologies per reviewing cycle

Regards,

Holger Lewen
Associate Researcher
Institut AIFB, Universit虞Karlsruhe
phone: +49-(0)721-608-6817
email: lewen@aifb.uni-karlsruhe.de
www: http://www.aifb.uni-karlsruhe.de/WBS

Am 21.03.2008 um 11:40 schrieb
<matthew.west@shell.com> <matthew.west@shell.com
>:
> Dear Pat, John, and Barry,
>
> I think the problem that many have with academic review is that it
> is open to abuse and personal prejudice.
>
> An approach that is aimed at being more structured is Document
> Inspection.
> This at least tries to be objective, and is designed to make being
> subjective
> harder.
>
> The approach is to measure a document against its purpose and target
> audience.
> It uses a team of trained inspectors (training is simple and
> straightforward)
> - Divide document so that (for total inspection) each part of the
> document is
> reviewed by 3 inspectors (diminishing returns after 3 in terms of
> identifying
> new issues). Author is one of the inspectors.
> - Identify issues:
> - Statements that are untrue, or unclear and/or ambiguous to

> target audience
> - Super-major - show stopper
> - Major - subverts the purpose of the document
> - Minor - incorrect but no major impact
> - Editorial - grammar and spelling, badly laid out diagrams
> Review the issues, determine whether document is fit for purpose (no
> Super Major,
> low count of majors).
>
> This gives a rationale for rejection, and provides the basis for
> improvement
> so that inclusion becomes possible. The issue list is publicly
> available so that
> people can see where the deliverable is, and whether the issues
> raised are a
> concern for them.
>
> This is of course the essence of reaching consensus in a
> standardization process,
> but if you are getting into any level of approval, you ARE doing
> standardization,
> however you choose to dress it up.
>
> Regards
>
> Matthew West
> Reference Data Architecture and Standards Manager
> Shell International Petroleum Company Limited
> Registered in England and Wales
> Registered number: 621148
> Registered office: Shell Centre, London SE1 7NA, United Kingdom
>
> Tel: +44 20 7934 4490 Mobile: +44 7796 336538
> Email: matthew.west@shell.com
> http://www.shell.com
> http://www.matthew-west.org.uk/
>
>

>
>> -----Original Message-----
>> From: ontology-summit-bounces@ontolog.cim3.net
>> [mailto:ontology-summit-bounces@ontolog.cim3.net]On Behalf Of Patrick
>> Cassidy
>> Sent: 21 March 2008 04:08
>> To: 'Ontology Summit 2008'
>> Subject: Re: [ontology-summit] [Quality] What means
>>
>>
>> John,
>> Among the 'reviewers' is there any reason not to have an
>> expert committee
>> that can create a binary distinction of, e.g.
>> "well-structured" and "not
>> well-structured"? The imprimatur can be an alternative to absolute
>> exclusion, and still serve the legitimate concerns that Barry
>> has about
>> poorly constructed ontologies.
>>
>> Pat
>>
>> Patrick Cassidy
>> MICRA, Inc.
>> 908-561-3416
>> cell: 908-565-4053
>> cassidy@micra.com
>>
>>
>>> -----Original Message-----
>>> From: ontology-summit-bounces@ontolog.cim3.net
>> [mailto:ontology-summit-
>>> bounces@ontolog.cim3.net] On Behalf Of John F. Sowa
>>> Sent: Thursday, March 20, 2008 11:56 PM
>>> To: Ontology Summit 2008
>>> Subject: Re: [ontology-summit] [Quality] What means
>>>
>>> Pat, Barry, Deborah, and Ed,

>>>
>>> Barry asked an important question that gets to the heart of
>>> the issues we have been discussing:
>>>
>>> BS> What are scientific journals for? Why do they employ a peer
>>>> review process?
>>>
>>> There are two independent issues here: reviewing and publishing.
>>> Everybody would agree that reviewing is important, but ideally,
>>> the readers/users should have the option of making their own
>>> choices based on the reviews. When publication was expensive,
>>> the publishers became gatekeepers because it was economically
>>> impractical to publish everything.
>>>
>>> But with the WWW, new options are available. Publication is
>>> almost free, and we have the luxury of decoupling the reviewing
>>> process from the gatekeeping process. Metadata enables that
>>> decoupling:
>>>
>>> 1. All submissions to the OOR can be made available as soon
>>> as they are submitted.
>>>
>>> 2. The metadata associated with each submission can indicate
>>> what tests were made, what the reviewers said, and what
>>> results the users, if any, obtained.
>>>
>>> 3. Users can choose to see ontologies sorted by any criteria
>>> they want: in the order of best reviews, most thorough
>>> testing, greatest usage, greatest relevance to a particular
>>> domain, or any weighted combination.
>>>
>>> PH> This is where I part company with Barry, and indeed where I
>>>> believe that the very idea of controlling the contents of an OOR
>>>> (noting that the first O means 'open') needs to be examined very,
>>>> very carefully. Of course we would not argue that majority voting
>>>> should be used to choose scientific theories; but ontologies,
>>>> even those used by scientists, are not themselves scientific
>>>> theories.

>>>
>>> Ontologies overlap philosophy, engineering, science, and
>> mathematics.
>>> The closest model we have is the metadata registry, but new policies
>>> can and should be explored.
>>>
>>> BS>> While refrigerator manufacturers may allow democratic ranking
>>>>> to influence e.g. size and color, they would use other strategies
>>>>> e.g. in matters of thermodynamics.
>>>
>>> PH> Perhaps so: but we are here discussing matters of ontology, and
>>>> in the current state of the art, this may have more in common
>>>> with consumer product choice than with thermodynamics.
>>>
>>> That is the point I was trying to emphasize. The application
>>> developers have deeper understanding of their specific needs and
>>> problems than any general gatekeeper or committee of gatekeepers.
>>>
>>> DM> CSI, the specification writing organization for building
>>>> architecture, says quality is "a mirror of the requirements."
>>>
>>> That's a good point, which implies that different set of
>>> requirements might lead to a different ranking of the same
>>> ontologies. No gatekeeper can anticipate the requirements
>>> of all possible users.
>>>
>>> DM> Do you think the gatekeepers can help define the OOR
>> requirements
>>>> and set up the dynamic tests?
>>>
>>> I'd prefer to keep the reviewers and replace the gatekeepers with
>>> caretakers who have a broader role along the lines you suggested.
>>>
>>> EB> I'm thinking about bureaucrats. I think that many ontologies
>>>> (and more broadly, concept systems including thesauri,

>> taxonomies,
>>>> etc.) have been and will be developed for use within the mission
>>>> areas of government agencies. There can be a vetting process to
>>>> "approve" a concept system/ontology for use within a community
>>>> of interest.
>>>
>>> That suggests a further refinement of the roles of reviewers and
>>> gatekeepers/caretakers. At the source, there are individuals and/or
>>> organizations, who develop ontologies and make them available.
>>> Among the users, there may be organizations, coalitions, or
>>> bureaucracies that evaluate the ontologies and determine which
>>> of them are best suited to their groups of users.
>>>
>>> That is another reason for replacing the gatekeepers in the OOR
>>> with caretakers. Any gatekeeping that might be useful would be
>>> better done by user groups at a level close to the applications
>>> than by any gatekeeper that is close to the ontology providers.
>>>
>>> John

Bug:

Mark is making a very import point regarding the problematic nature of the analogy between journal peer reviewers and OOR gatekeepers. To some extent this is an apples to oranges comparison, despite the apparent similarity in the gatekeeper function.

Having said that, I do believe there is another function journal reviewers provide which is very relevant to what I'd hope to see provided by an OOR gatekeeper - at least when it comes to ontologies used to help add reliable, highly-scalable automation to bioinformatic data processing. To go back to the realm of advertising, this other gatekeeper role relates to the Google Scholar "slogan" - "Stand on the shoulders of giants". The idea is the peer review process in the literature helps provide a certain level of distillation and reliable vetting to ensure every researcher need only read, absorb, and collate vetted manuscripts, as opposed to all the manuscripts submitted to the relevant journals. As imperfect as this process, I still tend to transfer my confidence in the reviewers having competently vetted the articles to the articles themselves, so that I can accept - at a given point in time - the validity of the assertions in a given article and build on them by performing additional work on their "shoulders".

Perhaps - as others have suggested - it is too early in the evolution of ontology development practice for us to expect we can produce a vetting process capable of functioning in this way, but until we do, some of the expectations informaticists - and funding agencies - have for the use of ontologies are probably not achievable. Until there is a reliable vetting procedure, we cannot expect to re-use and extend existing ontologies effectively or with confidence for the purpose of bringing like data together in novel ways from across the biomedical data diaspora. Without vetting, we cannot expect to provide other developers with clear advise on what are the reliable ontological shoulders to build on. If the OOR has 3 ontologies covering a domain of interest at roughly the same scope and level of granularity such as the physiology of mammalian electrolyte balance or the assembly of peptides into functional multimeric macromolecular receptor complexes, how would a bioinformatics application developer determine which one to use? Even more importantly, if users pick at random from amongst the 2 or more ontologies covering the same domain, who will maintain the maps and software required to make deductions or inferences across the annotated data repositories which use these different ontologies to cover the same domain?

Another one of my expectations for using ontologies in biomedical informatics is as the data representation gets richer and more expressive, the nature of the software each application developer needs create can focus more on application-specific issues. Community tools capable of parsing the expressivity (reasoners) can provide more of the "smarts", so that the custom tools don't need to hard code it - and can exclusively focus on the application specific algorithms.

This partly goes back to a point I made earlier in this thread. Using ontologies to power broadly scoped, federated inferencing brings with it a distinct set of requirements that differ from those of applications focused on providing decision support built on a more narrowly focussed data warehouse. In that case, it can be perfectly acceptable for developers to pick the ontology they like the best from the several covering the domain(s) they require, so long as there is

no expectation the resulting knowledge repository will be easily shared with other informatics projects.

These may all be expectations so narrowly linked to biomedical informatics they do not hold sway in a more generic OOR. That's one of the issues I'm hoping to better understand by participating in this discussion and the upcoming summit.

Thanks again to all for the stimulating and detailed dissection of this important topic.

Cheers,
Bill

BS

>>> Both reviewers and developers will be further
>>> motivated to participate in this process because
>>> they can thereby directly influence the standard
>>> set of ontology resources which will be available
>>> to them, thereby also motivating the creation of:
>>> related ontology-based software, useful bodies of
>>> data annotated in terms of these resources, etc.
>>>

HL

>> Again, in my opinion also possible in an open reviewing system.

You can't have it both ways -- on the one hand you claim multiple benefits from having lots of reviewers, lots of perspectives, people can find precisely the ontology which will suit their purposes from the wisdom of crowds, etc.
Now, you say that lone reviewers lost in these crowds will be able 'to directly influence the standard set of ontology resources which will be available to them'.