

# Putting biomedical ontologies to work

Barry SMITH<sup>ab</sup> and Mathias BROCHHAUSEN<sup>a</sup>

<sup>a</sup> *Institute of Formal Ontology and Medical Information Science, Saarland University, Germany*

<sup>b</sup> *Department of Philosophy and New York State Center of Excellence in Bioinformatics and Life Sciences, University at Buffalo, USA*

*h*

**Abstract.** *Objectives:* Biomedical ontologies exist to serve integration of clinical and experimental data, and it is critical to their success that they be put to widespread use in the annotation of data. How, then, can ontologies achieve the sort of user-friendliness, reliability, cost-effectiveness, and breadth of coverage that is necessary to ensure extensive usage? *Methods:* Our focus here is on two different sets of answers to these questions that have been proposed, on the one hand in medicine, by the SNOMED CT community, and on the other hand in biology, by the OBO Foundry. We address more specifically the issue as to how adherence to certain development principles can advance the usability and effectiveness of an ontology or terminology resource, for example by allowing more accurate maintenance, more reliable application, and more efficient interoperation with other ontologies and information resources. *Results:* SNOMED CT and the OBO Foundry differ considerably in their general approach. Nevertheless, a general trend towards more formal rigor and cross-domain interoperability can be seen in both and we argue that this trend should be accepted by all similar initiatives in the future. *Conclusions:* Future efforts in ontology development have to address the need for harmonization and integration of ontologies across disciplinary borders, and for this, coherent formalization of ontologies is a prerequisite.

**Key Words:** Biomedical Ontologies, Ontology Harmonization, Quality Assurance, SNOMED CT

## Introduction

We can distinguish a number of influences which have played a role in the development of terminology resources for application in biomedicine:

1. the influence of library science and of dictionary and thesaurus makers, illustrated most conspicuously by MeSH, the indexing resource maintained by the National Library of Medicine [1] (which was itself inspired by LCSH, the Library of Congress Subject Headings);
2. the influence of database design and conceptual modeling, illustrated for example

by the HL7 initiative [2];

3. the influence of biological science, illustrated by the Gene Ontology (GO) [3] and by the other ontologies within the Open Biomedical Ontologies (OBO) Foundry initiative [4, 5];

4. the influence of advances towards greater formal rigor, illustrated for example by current developments within SNOMED CT and within the framework of the Semantic Web [6].

It is the recent advances made under headings 3. and 4. which are addressed in what follows. Progressively, the new ontologies and terminology resources now being developed are beginning to be distinguished from their predecessors by factors such as:

- a concern for the interoperability of ontologies developed for the representations of distinct but related domains (which implies also a concern for consistency of content),
- the attempt to create and validate coherent strategies for quality assurance of ontologies on the basis of user feedback and empirical testing,
- the pursuit of coordinated strategies for update and maintenance in light of scientific advance,
- an increasing concern with formal rigor, supported by increasingly sophisticated software tools for ontology maintenance, validation and interoperation,
- an increasing concern with biological accuracy, and thus with the reality to which the representational units in an ontology relate,

and, closely connected thereto,

- recognition of the need more clearly to distinguish data, information and other representational artifacts from the entities represented on the side of the organism – thus for example to distinguish disease from diagnosis.

We have addressed some of these factors already in [7], where we focused on obstacles to the harmonization of ontology and terminology resources in the eHealth domain, drawing conclusions specifically in relation to the integration of clinical and laboratory data on cancer research. Since biomedical ontologies exist to serve integration of data, their conditions of success can usefully be compared to those of a telephone network. Crucial to the latter is *numbers of subscribers*: a network with a small subscriber base is a failure, however excellent its technology might be. For an ontology, as for a telephone network, increasing numbers of subscribers/users bring increasing degrees of success. In the case of ontologies this means that, with every new body of data that is annotated using a given ontology, there is an increase in value not only of that ontology but of all the other ontologies with which it interoperates. Here we focus on two specific strategies for advancing usability and usefulness of ontology and terminology resources, strategies incorporated, respectively, in the SNOMED CT vocabulary and in the OBO Foundry initiative.

### **1. Increasing concern with formal rigor**

One strategy consists in increasing the attention that is paid to the formal-logical

properties of biomedical terminologies and related artifacts. Certainly there is still skepticism in some circles as to whether enhanced formal rigor is an asset in medical ontologies. Some parties still argue that medical knowledge is too dependent on subjective experience and local traditions to allow for the creation of unified scientifically based terminologies or ontology platforms. They see medicine as ‘art’ (or ‘folklore’) and not as ‘science’. With the growth of molecular and genomic medicine arising from the fusion of medicine with experimental biology, such arguments are beginning to be recognized as being outdated.

As medicine becomes biomedical science, the resultant growing need to integrate data deriving from different disciplines at different granularity levels implies also that medical terminology becomes affected by an increasing concern with matters of consistency and logical rigor, as illustrated for example in the work of the Semantic Web Healthcare and Life Sciences Interest Group [8] and in the development of description logic infrastructures for medical vocabularies such as GALEN [9] (a pioneer in this respect), the National Cancer Institute Thesaurus [10] and—of principal importance to us here—the most recent versions of the SNOMED vocabulary [11].

The same forces are in operation also with regard to biology, too, where the growth in importance of bioinformatics, and the vast expansion in the wealth of biological data available to researchers (much of it in open-access forms on the Web), have led to considerable effort in the development of formally more rigorous ontology resources, and also of associated software tools as are encapsulated for example in the work of the National Center for Biomedical Ontology [12].

The success of the Gene Ontology has meant that many biologists favor the OBO format (formerly the GO format) [13] as the logico-linguistic idiom within which to develop ontologies. The OBO format, too, is being subjected to increasing attention from the point of view of formal rigor. The aim is to find ways to harvest new possibilities of algorithmic reasoning [6] of a sort that will provide both quality assurance of the ontologies themselves and also enhanced computational support for biomedical research and clinical care. The OBO ontology repository provides 53 biological and biomedical domain ontologies, 49 of which are developed in the OBO format, accompanied by large bodies of data annotated in their terms, including over 11 million publicly accessible annotations relating gene products (proteins and functional RNAs) to terms in the GO [14].

Outside biology, it is the OWL (Web Ontology Language) is the favored (W3C standard) logico-linguistic idiom of the ontology community [15], and OWL’s nice computational qualities make it an attractive target for software developers. On the other hand the ontology content developed in native OWL format is still too often immature as compared, for example, to the GO. Thus OWL developers have still paid too little attention to uniformity in their treatment for example of basic ontological relations such as *part\_of*, and their work is marked by frequent examples of use-mention and related confusions (as in *Paris is\_a city* and *has\_temperature 29°* and *has\_synonym Parigi*). Thanks to the creation of OBO-OWL converters, however, OBO ontology content is now available for use in OWL-based applications [16, 17]. The OBO ontologies and associated annotations are accordingly now serving as an important channel for the expansion and qualitative enhancement of the Semantic Web in the life science domain.

## 2. Increasing concern with reality

A second strategy in ontology development involves paying more consistent attention to the biological and clinical reality on the side of the patient. Ontology is still primarily an affair of computer scientists and knowledge engineers and still often embraces some form of the conceptualist view, which sees ontologies as representations of ‘concepts’ (which means, roughly, units of knowledge or of meaning) in the mind of human beings. Ontology work is accordingly directed overwhelmingly at what is called ‘knowledge modeling’ and is focused on the separate concept systems developed by separate local groups of researchers. Principal topics of interest are the study of automatic methods for the merging and mapping of the separate ontologies thereby created, and creation of ontology repositories based on computational rather than on biological or clinical features of the collected ontologies. Now, however, as ontologies are being increasingly developed not by computer scientists but by natural scientists and clinical researchers for their own domain-specific needs, a concern for the accurate and globally consistent representation of biomedical reality is becoming manifest, with a consequent decrease in focus on ontologies as independently developed computer artifacts in need of retrospective merging and mapping.

This new focus on biological reality on the part of life science ontologists – a move away from the earlier concern with data or information considered for its own sake towards what we shall call the ‘realist’ paradigm – is illustrated notably by recent revisions of the SNOMED CT vocabulary [11, 18, 19], and in the development of the OBO Foundry initiative. The latter is based on a distinction between three levels of reality:

- *mental representations* (ideas or thoughts in our minds representing specific portions of reality)
- *representational artifacts* (including ontologies, textbooks, and so forth), which we develop to make our mental representations concretely accessible to others
- *reality itself*, which serves as the target of the mental and physical representations especially in the scientific domain.

The realist holds that success in ontology development depends on keeping clear the distinction between these three levels [20], and this means recognizing that the reality which our representations are developed to represent exists independently of these representations themselves.

Realists, in contrast to conceptualists, define an ontology as a representation of the types or classes or kinds of entities (“types” in the following) existing in a specific domain of reality and of the relations between them. They see ontologies as representational artifacts developed in order to support scientific investigations or other endeavors, for example in the domain of clinical diagnosis and treatment, focused on identifying general laws.

Types are the real invariants or patterns in the world apprehended by the specific sciences through experimental research. Types are *instantiated* at different places and times by different particulars [20]. Types are designated by general terms such as ‘dog’ or ‘diabetes’. ‘Dog’ is the name of the type that is instantiated by my dog Fido and by

your dog Rover.

The tendency towards greater realism in the case of SNOMED CT can be seen in its recent deactivation of concepts involving the qualifier “not otherwise specified” (NOS), such as “Brain injury NOS (disorder)” (262686008). Already Cimino [21] had pointed out that qualifiers like “NOS” cause problems. These problems can be explained from a realist perspective by adverting to the fact that *there is no such entity* as a ‘brain injury not otherwise specified’ (whether as type or as instance). If there were such an entity at a given time  $t$ , and if some specification were to be added to the record of this entity at some later time  $t'$ , then we would either have to conclude that the original entity was thereby destroyed by the mere act of recording new information about it, or that it was non-identical to the putative successor entity created through this act of recording. A good axiom of realist ontology, however, is that the world does not change in reflection of the fact that we change the ways we talk about it.

As argued in [22], NOS and similar terms do not capture any mind-independent reality. Rather, they are confusingly formulated representations of a state of *knowledge* about reality. Of course, for biomedical information systems, keeping track of such states of knowledge is indispensable for liability and other purposes. But even though both sorts of information must be captured, if coding schemes are to support algorithmic reasoning in ways valuable to biomedical research in the future, then we believe that a clear distinction must be drawn between the two.

SNOMED CT is taking some of the necessary steps in this direction, but it still includes terms such as ‘unknown living organism’ (SNOMED: 89088004) and ‘presumed viral agent’ (SNOMED: 106551006), which it treats as designating special sorts of organisms. Terms of this sort are, again, roundabout representations of states of knowledge. Similarly, SNOMED CT contains a number of terms, such as ‘abscess’, which are included twice, first as clinical finding and then as morphological abnormality:

128477000 Abscess (disorder)

44132006 Abscess (morphologic abnormality)

Problems then arise for those using SNOMED CT as a coding resource, because there is no clear distinction between the referents of these terms on the side of reality [23, 24].

The Gene Ontology and its sister ontologies within the OBO repository, too, initially manifested non-realist components of these sorts, which led to a number of structural problems in GO, documented for example in [25, 26] Since the launching of the OBO Foundry initiative in 2003, however, these ontologies have been subjected to a series of reforms designed to ensure that the ontologies conform incrementally to the goal of biological accuracy and thus to the realist paradigm.

Each ontology included in the OBO Foundry consists of structured representations of the types existing in a given domain of reality which are intended to be correct when viewed in light of our best current scientific understanding of reality. Each such ontology is thus itself analogous to a scientific theory; it has a unified subject-matter, which consists of entities existing independently of the ontology, and it seeks to optimize descriptive or representational adequacy to this subject matter to the maximal degree that is compatible with the constraints of formal rigor and computational

usefulness.

A simple example of the sorts of reforms to which the OBO Foundry ontologies have been subjected is the removal of assertions of the form:

1. biological process *is\_a* Gene Ontology

which hitherto existed alongside more biologically accurate assertions such as

2. cell development *is\_a* biological process.

The presence of assertions of the former sort reflects a use of the '*is\_a*' relation to mean something like 'is a terminological component of', which is inconsistent with its principal use, in the GO and elsewhere, as meaning 'is a subtype of'. Use-mention confusions such as 1. cause problems where the GO becomes integrated with other ontologies relating, for example, to information artifacts. In the presence of

3. Gene Ontology *has\_file\_format* RDF-XML,

for example, and from the transitivity of *is\_a*, we could infer from 1. and 2. that

4. cell development *has\_file\_format* RDF-XML,

which is biological nonsense.

Such problems with the Gene Ontology and its sister ontologies are being systematically eliminated through application of the principle that all assertions within the ontology should employ relations defined in accordance with the pattern of definitions laid down in the OBO Relation Ontology [27, 28]. Application of this rule ensures that the meanings of relation terms are the same in all ontologies, thus facilitating logically sound integration of the ontologies for purposes of formal reasoning.

The move in the direction of greater realism is illustrated also by the proposed incorporation into the OBO Foundry of the Foundational Model of Anatomy, an ontology, built from the start on realist principles, comprising representations of the types and relationships necessary for representation of the structure of the mammalian, and especially the human, body in a computable form that is also understandable by humans [29, 30].

The FMA strategy for the formulation of definitions, which has been adopted for use throughout the OBO Foundry, is based on principles first laid down by Aristotle. Briefly the definition of a representational unit 'A' in an ontology referring to a *species* (type) A should be formulated by reference to A's *is\_a* parent B within the same ontology (the corresponding *genus*) and to some *differentiating characteristic* C that is possessed by all and only the As. This results in Aristotelian definitions (also called 'genus-species definitions') of the form:

an A =def. a B which Cs,

Examples from the FMA are:

a *cell* =def. an *anatomical structure* which consists of *cytoplasm* surrounded by a *plasma membrane*

a *plasma membrane* =def. a *cardinal cell part* that surrounds the *cytoplasm*

When this strategy is followed, careful attention to definitions not only helps foster consistent use of terms, but also contributes to ensuring the soundness of the ontology's *is\_a* hierarchy. At the same time careful attention to the creation of this hierarchy helps to ensure correctness of the associated definitions. In addition, each definition that is composed in accordance with this rule is enriched by the definitions of all of its parents within the hierarchy.

### 3. Modularity of Biomedical Ontologies

SNOMED CT and the OBO Foundry embrace alternative strategies with respect to the issue of modularization, strategies which differ in both their theoretical and their practical consequences.

The most ambitious initiative to address the problem of terminological unification in biomedicine is currently being mounted by the International Health Terminology Standards Development Organization (IHTSDO) [31], which is seeking to establish the SNOMED CT vocabulary as an international master terminology providing comprehensive coverage of the entire domain of medicine in a multiplicity of languages, starting out from the basis of the English-language SNOMED CT vocabulary, which already comprehends more than 315,000 concepts. This vocabulary already exists in partial versions in certain other languages, and is built around a description logic backbone [32], albeit not one of the W3C standard OWL description logics. In addition, the SNOMED CT vocabulary is mapped to other important existing standards, and full inclusion of the WHO's International Classification of Disease (ICD) 10 is on the way.

A major advantage of SNOMED CT is its comprehensive reach, which is secured through some 21 hierarchies, including for instance Clinical Findings, Procedure, Body Structure, Organism, Substance, Social Context, Observable Entity, Staging and Scales, Pharmaceutical / Biological Product, and Record Artifact.

By providing a common structure and terminology, the SNOMED CT vocabulary ultimately aims at providing a single data source for review. Furthermore, the use of such a common vocabulary promises benefits of less redundant data and easier opportunities for longitudinal studies and meta-analyses and for ensuring consistency of data across the lifetime of the patient and from one healthcare institution to the next. Employing one single comprehensive source offers the advantage of allowing the use of common tools and techniques as well as common training. This promises significant rewards, for example for clinical decision support and for ensuring more adequate public health data; unfortunately, however, it is on the other side difficult to motivate the large investments needed to ensure such prospective rewards, especially in light of certain well-publicized failures [33].

The fact that so few healthcare institutions have embraced SNOMED CT for clinical coding seems to suggest that incentives are still missing for the considerable

investments which would be needed to harvest these benefits. The strategy for compensating for the high cost of implementing and running a sophisticated broad-coverage terminology such as SNOMED CT in the hospital environment is still an open issue. Thus, for example, SNOMED CT was not created to support hospital billing needs, still a pre-eminent consideration of hospital management that is catered for currently by simpler terminological resources (primarily ICD) for which there is already an established user base.

One additional obstacle to the more widespread use of SNOMED CT is that it is marked by a number of internal structural problems, some examples of which have been mentioned above. In addition to problematic redundancies, there are also gaps in the terminology, a lack of compositional structure, and shortfalls in consistency arising from the application of different development principles in different parts of the vocabulary as it has grown, sometimes via bulk incorporation of alien matter, over time. There is also evidence that these flaws hinder efficient application in coding [23, 24]. These quality-related issues are, as we have noted, being addressed. An objection of a different sort concerns the very foundations of the strategy of providing a single broad-coverage master terminology.

The OBO Foundry, in contrast thereto, advocates a modular approach. Its strategy, (<http://obofoundry.org>) [5], is to create a suite of interoperable ontologies covering different (non-overlapping) domains of biomedical reality on the basis of an evolving set of *common principles for ontology development*. Only those ontologies whose developers commit to adhering to these principles are admitted into the Foundry. The principles themselves are designed both to bring about the harmonization of existing domain ontologies and also to support the creation of new ontologies by providing an evolving set of simple guidelines and of tested models of good practice for those new to the field. Such guidelines are extremely important if the interoperability of ontologies for neighboring domains is to be preserved as new areas are identified where ontologies need to be created *ab initio*.

The goal of SNOMED CT, in effect, is to develop one single ontology covering the entirety of the medical domain. The goal of the OBO Foundry, in contrast, is to bring about a situation in which the coverage of terminology resources is advanced in stepwise fashion across ever broader domains of biomedical reality while the advantages of consistency and the lessons learned from earlier successes (and failures) are preserved. To this end the Foundry has developed a methodology by means of which ontologies covering different domains can be incrementally developed in a coordinated fashion that allows exploitation of the benefits of division of labor and pooling of expertise.

The participants in the OBO Foundry initiative hold that use of a minimal upper ontology framework is a basic presupposition of harmonization. Such an ontology can incorporate certain basic principles of best practice in ontology development and help to avoid common mistakes in something like the way in which the use of basic mathematical theories serves to bring about these same effects in physics. Advantages will accrue from using an upper level ontology, however, only if the principles employed in its development are both well founded and thoroughly tested. Re-use of an upper level ontology then fosters harmonization and avoids time-consuming re-development of those meta-level structures which are needed by domain scientists to



organize their ontology resources. It embodies principles of which these domain scientists will likely have an imperfect grasp.

The Basic Formal Ontology (BFO), which serves as upper level kernel of OBO Foundry ontologies, rests on a basic distinction (shared also by the SUO and DOLCE upper level ontologies) between *continuants* and *occurrents*. The former are entities in reality that *endure* (continue to exist) through time. They persist self-identically even while undergoing changes of various sorts. The latter *occur*, which means that they have, in addition to their spatial dimensions, also a fourth, temporal dimension. Occurrents (for example processes) unfold through a period of time in such a way that they can be divided into temporal parts or phases. They have a beginning, a middle, and an end. Continuants, in contrast (for example organisms), exist in full at any time at which they exist at all, while at the same time gaining and losing parts in the course of development and growth [34, 35].

Using an upper level ontology can foster harmonization by providing a uniform and coherent approach to reality representation at the topmost level of organization. It is at the lower levels, however, that we will find those terms which predominate in practical uses of the ontology. General criteria of the sort embodied in an upper level ontology provide useful assistance when organizing these lower level terms [36], but the use of such an ontology should not diminish the utility of an information resource to its ultimate users.

The OBO Foundry has addressed also the issue of motivation. Ontology development, when carried out manually, is an expensive affair because it requires the investment of considerable time on the part of domain experts who have a commitment to the scientific accuracy of the venture. Experience thus far suggests that, in at least a significant portion of life science domains, experts will participate in the Foundry initiative and will manifest this commitment because—as with comparable open source efforts in the software domain [37]—they are rewarded for their participation through the fact that, thanks to the division of labour and authority brought about by modularity, they will come to play a direct role in shaping the resources that they will need in their own work in the future [38].

## Conclusion

Both IHTSDO and the OBO Foundry strive to find a way to provide a complete representation of clinical reality. IHTSDO has the advantage of a large unitary organization with considerable financial resources and a systematic plan. The OBO Foundry, which – aiming to be future proof in an era when medicine is ineluctably becoming transformed into biomedicine – treats the representation of clinical reality as part of the larger aim of creating a suite of interoperable ontologies covering the entire domain of the life sciences, has the advantage of a strategy for the incentivization of experts and a large established user base, albeit still primarily in domains of biology. It is too early to judge whether either of these strategies will provide the solution to the pressing problems at hand, but these problems—of user-friendliness, reliability, cost-effectiveness, and breadth of coverage—must be resolved if effective integration of biomedical data is to be secured.

## Acknowledgements

Smith's contribution to this work was funded in part by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant 1 U 54 HG004028 (National Center for Biomedical Ontology). Information on the National Centers for Biomedical Computing can be found at <http://nihroadmap.nih.gov/bioinformatics>.

## References

- [1] Medical Subject Headings [Internet]. Bethesda (MD): U.S. National Library of Medicine. The U.S. National Library of Medicine. Available from: <http://www.nlm.nih.gov/mesh/>. Last accessed 17 January 2010.
- [2] Health Level Seven International, Ann Arbor (MI). Available from: <http://www.hl7.org/>. Last accessed 17 January 2010.
- [3] The Gene Ontology Consortium. 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 32:D258-D261.
- [4] <http://www.geneontology.org/>. Last accessed 17 January 2010.
- [5] Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology* 2007 Nov 25: 1251-1255. Available from <http://obofoundry.org/>. Last accessed 17 January 2010.
- [6] Ruttenberg A, Clark T, Bug W, Samwald M, Bodenreider O, Chen H, et al. Advancing translational research with the Semantic Web. *BMC Bioinformatics* 2007;8(Suppl 3):S2.
- [7] Smith B, Brochhausen M. Establishing and Harmonizing Ontologies in an Interdisciplinary Health Care and Clinical Research Environment. In: Blobel B, Pharo P, Nerlich M, editors. *eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge*. Amsterdam: IOS Press; 2008. p. 219-234.
- [8] Charter of the Semantic Web Health Care and Life Sciences (HCLS) Interest Group. World Wide Web (W3C) Consortium. Available at: <http://www.w3.org/2008/05/HCLSIGCharter>. Last accessed 17 January 2010.
- [9] Open GALEN Mission Statement. Available from: <http://www.opengalen.org/>. Last accessed 17 January 2010.
- [10] Ceusters W, Smith B, Goldberg L, A Terminological and Ontological Analysis of the NCI Thesaurus, *Methods of Information in Medicine*, 44 (2005), 498-507. The NCI Thesaurus is available from: <http://nciterns.nci.nih.gov/ncitbrowser/>. Last accessed 17 January 2010.
- [11] International Health Terminology Standards Development Organisation, SNOMED CT: Systematized Nomenclature of Medicine-Clinical Terms. Available from: <http://www.ihtsdo.org/snomed-ct/>. Last accessed 17 January 2010.
- [12] Rubin DL, Noy NF, Richter JB, et al. The National Center for Biomedical Ontology: Advancing Biomedicine through Structured Organization of Scientific Knowledge. *OMICS: A Journal of Integrative Biology*, 2006; 10(2): 185-198. Available from: <http://bioontology.org/>. Last accessed 17 January 2010.
- [13] Gene Ontology Consortium. File Format Guide. Available from: <http://www.geneontology.org/GO.format.shtml>. Last accessed 17 January 2010.
- [14] Gene Ontology Consortium. Gene Ontology Annotation Database. European Bioinformatics Institute. Available from: <http://www.ebi.ac.uk/GOA/>. Last accessed 17 January 2010.
- [15] OWL 2 Web Ontology Language. W3C Recommendation. 27 October 2009. Available from: <http://www.w3.org/TR/owl2-syntax/>. Last accessed 17 January 2010.
- [16] OBO-OWL RESTful Conversion API. Berkeley (CA): Berkeley Bioinformatics Ontology Project. Available from: <http://www.berkeleybop.org/obo-conv.cgi/>. Last accessed 17 January 2010.
- [17] Moreira DA, Musen MA. OBO to OWL: A Protégé tab to read/save OBO ontologies. *Bioinformatics* 2007 Jul; 23(14):1826-70.
- [18] Ceusters W, Spackman KA, Smith B. Would SNOMED CT benefit from realism-based ontology evolution? *AMIA Annu Symp Proc.* 2007 Oct; 11:105-9.
- [19] Schulz S, Suntisrivaraporn B, Baader F. SNOMED CT's Problem List: Ontologists' and Logicians' Therapy Suggestions. *Stud Health Technol Inform*, 2007;129:802-6.

- [20] Smith B, Kusnierczyk W, Schober D, Ceusters W. Towards a reference terminology for ontology research and development in the biomedical domain. KR-MED 2006, biomedical ontology in action. Baltimore MD; 2006.
- [21] Cimino JJ. Desiderata for controlled medical vocabularies in the Twenty-First Century. *Methods Inf Med*; 1998 Nov;37(4-5):394-403.
- [22] Bodenreider O, Smith B, Burgun A. The Ontology-Epistemology Divide: A Case Study in Medical Terminology. In: Varzi A, Vieu, L, editors. *Proc. of the Int. Conf. on Formal Ontology and Information Systems (FOIS 2004)*, Turin, 4-6. November 2004; Amsterdam: IOS-Press; 2004. p. 185-195.
- [23] Andrews JE, Richesson RL, Krischer J. Variation of SNOMED CT coding of clinical research concepts among coding experts. *J Am Med Inform Assoc*. 2007 Jul-Aug;14(4):497-506.
- [24] Chiang MF, Hwang JC, Yu AC, Casper DS, Cimino JJ, Starren J. Reliability of SNOMED-CT coding by three physicians using two terminology browsers. *AMIA Annu Symp Proc*. 2006:131-5.
- [25] Smith B, Williams J, Schulze-Kremer S. The Ontology of the Gene Ontology. *AMIA Annu Symp Proc*. 2003:609-13.
- [26] Smith B, Köhler J, Kumar A. On the Application of Formal Principles to Life Science Data: A Case Study in the Gene Ontology. In: *Proceedings of DILS 2004 (Data Integration in the Life Sciences)*, (Lecture Notes in Bioinformatics 2994), Berlin: Springer; 2004, p. 79–94.
- [27] OBO Relation Ontology. Available from: <http://www.obofoundry.org/ro/>.
- [28] Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, et al. Relations in Biomedical Ontologies. *Genome Biol*. 2005;6(5):R46.
- [29] The Foundational Model of Anatomy [Internet]. Seattle (WA): Structural Informatics Group, University of Washington; c2008-10 [cited 2010 Jan 17]. Structural Informatics Group, University of Washington; [1 screen]. Available from: <http://sig.biostr.washington.edu/projects/fm/>.
- [30] Rosse C, Mejino JV Jr. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform*. 2003 Dec;36(6):478-500.
- [31] International Health Terminology Standards Development Organisation [Internet]. Copenhagen (Denmark): International Health Terminology Standards Development Organisation; c2010 [cited 2010 Jan 17]. Available from: <http://www.ihtsdo.org/>.
- [32] Donnelly K. Multilingual documentation and classification. In: Blobel B, Pharow P, Nerlich M, editors. *eHealth: Combining Health Telematics, Telemedicine, Biomedical Engineering and Bioinformatics to the Edge*. Amsterdam: IOS Press; 2008. p. 235-44.
- [33] Rector A. Terminologies, Ontologies, & SNOMED [Internet]. Freiburg (Germany): Institute of Medical Biometry and Medical Informatics; c2006-10 [updated 2006 Oct 24; cited 2010 Jan 17]. *Health Informatics World Wide* [67 pages]. Available from: [www.hiww.org/smcs2006/talks/Rector.ppt](http://www.hiww.org/smcs2006/talks/Rector.ppt).
- [35] Basic Formal Ontology [Internet]. Saarbrücken (Germany); c2006-10 [updated 2009 Nov 5; cited 2010 Jan 17] Institute for Formal Ontology and Medical Information Science [about 2screens]. Available from: <http://www.ifomis.org/bfo/>.
- [36] Grenon P, Smith B, Goldberg L. Biodynamic Ontology: Applying BFO in the Biomedical Domain. In: Pisanelli DM, editor. *Ontologies in Medicine*. Amsterdam: IOS Press; 2004, p. 20-38.
- [37] Smith B. From Concepts to Clinical Reality: An Essay on the Benchmarking of Biomedical Terminologies. *J Biomed Inform*. 2006 Jun;39(3):288-98.
- [38] Weber S. *The Success of Open Source*. Cambridge, MA: Harvard University Press; 2004.
- [39] Smith B. Ontology (Science). In: Eschenbach C, Grüninger M, editors. *Proceeding of the 2008 Conference on Formal Ontology in Information Systems: Proceedings of the Fifth international Conference (FOIS 2008) Frontiers in Artificial Intelligence and Applications*, vol. 183. Amsterdam: IOS Press; 2008. p. 21-35.