

DON'T KNOW MUCH ABOUT PHILOSOPHY: THE CONFUSION OVER BIO-ONTOLOGIES

Mark A. Musen, M.D., Ph.D.

*The National Center for Biomedical Ontology
Stanford University
251 Campus Drive, X-215
Stanford, CA 94305 USA*

Abstract:

For the past decade, there has been increasing interest in ontologies in the biomedical community. As interest has peaked, so has the confusion. The confusion stems from the multiple knowledge-representation languages used to encode ontologies (e.g., frame-based systems, Semantic Web standards such as RDF(S) and OWL, and languages created specifically by the bioinformatics community, such as OBO), where each language has explicit strengths and weaknesses. Biomedical scientists use ontologies for multiple purposes, from annotation of experimental data, to natural-language processing, to data integration, to construction of decision-support systems. Each of these purposes imposes different requirements concerning which entities ontologies should encode and how those entities should be encoded. Although the biomedical informatics community remains excited about ontologies, exactly what an ontology is and how it should be represented within a computer are points about which, with considerable questioning, we can see little uniformity of opinion. The confusion will persist until we can understand that different developers have very different requirements for ontologies, and therefore those developers will make very different assumptions about how ontologies should

be created and structured. We will review those assumptions and the corresponding implications for ontology construction.

Our National Center for Biomedical Ontology (<http://bioontology.org>) is one of the seven national centers for biomedical computing formed under the NIH Roadmap. The Center takes a broad perspective on what ontologies are and how they should be developed and put to use. Our goal, simply put, is to help to eliminate much of the current confusion. The Center recognizes the importance of ontologies for use in a wide range of biomedical applications, and is developing new technology to make all relevant ontologies widely accessible, searchable, alignable, and useable within software systems. Ultimately, the Center will support the publication of biomedical ontologies online, much as we publish scientific knowledge in print media. The advent of biomedical knowledge that is widely available in machine-processable form will alter the way that we think about science and perform scientific experiments. The biomedical community soon will enter an era in which scientific knowledge will become more accessible, more useable, and more precise, and in which new methods will be needed to support a radically different kind of scientific publishing.