

NCI Center for Bioinformatics
Informatics Seminar Series

**Formal Principles for Biomedical
Ontologies**

—
2:00 until 4:00 PM
February 9, 2004

6130 Executive Blvd. (EPN), Conf. Room H

—

Barry Smith

Julian Park Professor of Philosophy, University at Buffalo and
Director, Institute for Formal Ontology and Medical Information
Science, University of Leipzig

The automatic integration of information resources in the life sciences is one of the most challenging problems facing biomedical informatics today. Ontologies have played an important role in realizing this goal, not least by providing controlled vocabularies which seek to make it possible to search heterogeneous databases secure in the knowledge that the same terms will also represent the same entities. One of the most impressive and influential developments in this respect is the Gene Ontology, which is rapidly acquiring the status of a *de facto* standard in attempts to develop controlled vocabularies for shared use across different biological domains. The tremendous investment of time and effort by the GO Consortium has already brought considerable benefits to a range of different types of biological and biomedical research.

An examination of the structure of GO reveals, however, a number of problems which are characteristic of much work in bioinformatics and which are destined to raise increasingly serious obstacles to the automatic integration of biomedical information in the future. This is because on the one hand, as the GO Consortium itself accepts, with increases in the size and scope of GO it will 'be increasingly difficult to maintain the semantic consistency we desire without software tools that perform consistency checks and controlled updates.' Yet on the other hand much of the information that GO contains is not capable of being accessed or manipulated by software tools.

We will show that these problems can be avoided through adherence to formal organizing principles drawn from philosophical ontology, principles which represent best practices in classification and definition. We show how their neglect is associated in

systematic ways with certain kinds of coding errors, and we conclude by sketching the potential benefits of a formally more adequate regime of classification and definition in the future.

Following the presentation there will be an open discussion.