From Classical Metaphysics to Medical Informatics

Barry Smith

http://ontology.buffalo.edu/smith

Forthcoming in *Humboldt Kosmos*

An item in the *Wall Street Journal* of 13.11.2001 describes how computers might be used as a means of alerting health officials to problems caused by bioterrorism. The author notes that biological attacks will likely be marked by statistical spikes in the purchase of over-the-counter remedies for specific common ailments, and the data pertaining to such purchases can be made available immediately via the computerized inventory systems used by drugstores.

There is a problem however. For while the technology for running databases has reached an impressive state of maturity, the classification systems upon which this technology is based are the products of myriad *ad hoc* decisions stretching back to the early days of database design. This means that the data in drugstore computers exists in a variety of different forms, reflecting mutually incompatible ways of partitioning the universe of pharmaceutical products. To resolve these incompatibilities is here a relatively simple matter. In many other such cases, however, the inconsistencies resulting from incompatible classifications are leading to ever more intractable problems wherever attempts are made to integrate data from different sources – problems reminiscent of the old fable of the Tower of Babel.

Initially the problem of database integration was resolved in case by case fashion. Pairs of databases were cross-calibrated by hand, rather as if one were translating from French into Hebrew. As the numbers and complexity of database systems increased, however, the idea arose of streamlining these efforts by constructing one single benchmark taxonomy into which all of the various classification systems would need to be translated only once. By serving as a *lingua franca* for database integration this benchmark taxonomy would ensure that all databases calibrated in its terms would be automatically compatible with each other.

Interestingly, now, the proposed central classification system was called by information scientists an *ontology*, and it was quickly recognized that work on its construction would have more than a few echoes of the metaphysics of old. This is because many of the difficulties faced by information scientists in building an ontology turn out to be identical to problems with which philosophers have grappled since Aristotle's day. They are problems relating to universals and particulars, properties and relations, events and processes. How, in a world of continuous differences, do category boundaries arise? How can we account for the identity of an individual over time when the individual is gaining and losing parts? Is a class or category anything more than the totality of its instances?

***

The underlying premise of the new information systems ontology was that it would be possible to construct a classification system so general that all databases could be reformulated in its terms. The potential advantages of ontology thus conceived are obvious. If all databases can be made compatible, then the prospect arises of merging all of the separate database resources in such a way as to create a single knowledge base of a scale hitherto unimagined, thus fulfilling the ancient dream of a Great Encyclopedia comprehending the entirety of human knowledge.

Unfortunately however, as experience has shown, the construction of a benchmark ontology proved to be a much more complex task than was originally envisaged. This ontology must be simple enough that it can be programmed into our computers, yet it must be comprehensive enough that it can allow the expression of terms derived from all competing systems of classification. In the face of such difficulties the information systems community has responded with a series of partial ontologies, each resting on a different pragmatically motivated choice about the way an ontology should be built. Ironically, therefore, the very Tower of

Babel conditions which the ontological project was initially designed to address have been recreated within ontology itself.

***

The Institute for Formal Ontology and Medical Information Science (http://ifomis.de) in the University of Leipzig represents a new approach to solving the problems of ontology. This Institute, which I have founded together with my Leipzig colleagues Barbara Heller and Heinrich Herre, seeks a return to the original idea of a common reference ontology. In contrast to previous efforts, however, which awarded a prime role to the *practical* factor of programmability, IFOMIS will start from the idea that the project of developing a common reference ontology can profit from the *theories* developed by philosophers over two millennia of ontological research.

The IFOMIS ontology will be marked further by the factor of realism. Thus where existing information systems ontology has been based primarily on the strategy of knowledge representation, that is to say on the attempt to represent the *concepts* used by the practitioners within a given domain, IFOMIS will seek to develop a comprehensive theory of the divisions and interrelations between the entities *on the side of reality itself*, a theory which can do justice to the fact that the very same reality may be sliced in different ways when addressed from different sides.

The work of IFOMIS will not be exclusively philosophical. It will address also applications in the domain of medical informatics. IFOMIS will thus employ not only philosophers but also information scientists and medical specialists, drawing on existing research in Leipzig under the auspices of the Competence Network for Malignant Lymphoma and the Institute for Medical Informatics, Statistics and Epidemiology (http://imise.de).

The domain of medicine has been selected for application purposes not only in light of its intrinsic significance but also because of the ontological challenges which it presents. Medicine calls for an ontology which can allow the simultaneous application of distinct perspectives (of, for example, doctor and patient, of pharmacologist and geneticist) to one and the same reality. This is because the entities in such a complex domain can sustain classifications reflecting causally relevant distinctions *at more than one level of granularity*. The medical ontology of IFOMIS must for example have the resources to sustain not merely an anatomical ontology at the level of organs within the structure of the human body, but also cell, protein, gene and molecule ontologies at successively finer resolutions. It must sustain also classifications of *processes* at different resolutions, including the chemical and biological processes taking place inside the body.

Most importantly, the IFOMIS medical ontology will need to comprehend the various types of entities involved in those complex processes we call *clinical trials*. A clinical trial is a controlled experiment in which the effectiveness of a given therapy is measured in systematic fashion in relation to pre-selected groups of patients. As principal testing ground for its methods, IFOMIS will seek to develop a medical domain ontology that is expressive enough to represent the structures of all the standard types of trials. This ontology should comprehend classification systems for therapies, patient populations and outcomes. It should yield standards not only for the representation of trial data but also for the preparation of clinical protocols and of the guidelines which specify procedures for diagnosis and treatment.

IFOMIS thus has a unique opportunity to put philosophical theories to the test empirically. Its ultimate goal is to provide a complete general ontology of the whole of reality. Its proximate goal is to demonstrate how ontological methods can lead to improvements in the domain of clinical trial management and thus to contribute to the wider effort of achieving improvements in the reliability, efficiency and economic delivery of health care through the rigorous application of the clinical trial methodology.