Comments on the paper Biomedical Ontologies: Toward Scientific Debate, by Maojo et al.

Mor Peleg

Dept. of Information Systems, University of Haifa, Israel

Maojo and co-authors discuss computational biomedical ontologies from different perspectives, including their philosophical foundations, and in particular issues of emergence, the ability (or lack of ability) of ontologies to generate new discoveries and theories, and to evolve. Pointing to a new direction of research in biomedical ontologies, the authors raise open questions related to visual reasoning and spatial ontologies.

Emerging novel biological knowledge and its integration with current scientific views is not always easily represented in ontologies, even when the new knowledge is deduced by scientists and is not expected to be discovered purely by reasoning with ontologies. Maojo and co-authors note that different ontologies that cover the same domain but view it from different perspectives (e.g., of molecular biologists, physiologists, and clinicians) may address different biological levels using a different level of granularity and different context. Indeed, the different communities of researchers and practitioners have different points of view that change over time, as the scientific knowledge about a domain grows. For example, the cell cycle stages, such as interphase and its phases, G1 (first gap), S (synthesis), and G2 (second gap), and the mitosis stage with its phases, prophase, prometaphase, metaphase, anaphase, and telophase, were defined based on cytological observations made by examining dividing cells through a microscope. For example, at the onset of prophase, chromatin condenses together into a chromosome and at metaphase the centromeres of the chromosomes convene along the metaphase plate. However, ontologies need to consistently represent the dynamic knowledge about high-level biological processes in the context of their molecular-level sub-processes [1]. For example, they should be able to relate the interphase G1 phase to processes of protein synthesis, organelles production, and cell growth. Furthermore, ontologies can also be used to provide even more detailed molecular definitions. Such definitions could relate the cell cycle checkpoint between phase G1 and phase S, in which the cell duplicates its DNA, to the specific molecular-level processes and their participating molecules (e.g., cell cycle kinases CDK4/6-cyclin D and CDK2-cyclin E) and molecular complexes (for details see http://www.biocarta.com/pathfiles/h_g1pathway.asp). Peleg et al. [2] provide another example where a bio-ontology was used to specify the process of protein translation in a high-level representation that is further nested into a detailed specification of tRNA mutations that cause abnormal protein translation resulting in clinical disease phenotypes.

The authors reiterate Noble's criticism of the Gene Ontology (GO), which excludes physiology and most aspects of evolutionary biology and therefore cannot adequately address modeling phenomena across different biological levels. It is important to note, however, that bio-ontologies have different

purposes; GO [3] is meant to provide a common vocabulary used to annotate gene products according to the biological processes at which they participate, the function that they exhibit, and their cellular localization. In this way, GO can facilitate data aggregation (e.g., retrieval of all mouse genes that participate in lipid metabolism) [4] and data integration [5], where data from different biological data bases is retrieved and joined based on common GO annotations. Other bio-ontologies have been used to formulate queries for accessing and integrating data from multiple biological databases [6] or for modeling and simulating biological processes in the context of their participating molecules and their resulting disease phenotypes [1, 2, 7].

Another interesting question raised by the authors is whether computational ontologies could do more than provide a means of representing descriptive knowledge of a field in a computationally useful way, or in other words could they go beyond "what you put in is what you get out". The answer to this question, in my view, is affirmative. Existing ontologies already provide not just descriptions but can be the basis for planning care pathways [8], providing explanations for decision-support recommendations [9], and enable simulation-based prediction of system behavior over time [7]. Furthermore, by representing libraries that contain definitions of different types of exceptions and exception-handling mechanisms [10], even unanticipated exceptions that occur during execution of clinical-guideline based decision-support systems could be handled by generic exception-handlers. Similarly, using description-logics-based ontologies, default behavior could be defined and executed for controlling the access to electronic medical records when specific organizational policies have not been defined [11].

Moving on to the second topic addressed by the paper by Maojo and co-authors, the focus on visual reasoning and spatial ontologies is an important emerging need for biomedical ontologies. Geometric shapes are important means for biologists for conceptualizing and comprehending biomolecular structures as well as for understanding how structure relates to function and to process. Biologists are used to creating diagrams to convey how biomolecular complexes act in biological processes e.g., the replication complex that replicates DNA, the protein translation mechanisms, etc. Biologists find it more intuitive to comprehend such diagrams in contrast with more abstract representations of biological process models. To make computational biological process models more intuitive for biologists, tools such as Cell Illustrator [12] allow biological scientists (users) can intuitively model and simulate complex dynamic interactions and processes in biopathways comprising of hundreds of entities within and among cells, using intuitive icons that correspond to classes in the biological pathway ontologies named Cell System Ontology.

Maojo and co-authors note that when considering shapes, fundamental concepts that arise are connectedness and adjacency. These distinctions are also important when considering cellular organelles; in the BioWorkflow ontology [1], cellular components are classified into membranes (e.g., nuclear membrane), spaces (e.g., cytoplasm, lysosome lumen), membrane bound compartments (e.g.,

membrane bound-organelles, such as nucleus, chloroplast), non-membrane bound cellular compartments (such as cytostome and nucleolus), and cellular structures (e.g., centriole).

There are many additional topics that are important in the context of biomedical ontologies and have not been covered by this paper. One of those topics is the use of computational methods to discover knowledge in data [13-15] and in biomedical publications [16, 17] in order to populate ontologies. Such methods make the task of ontology construction more scalable.

[1] Peleg M, Yeh I, Altman RB. Modeling biological processes using Workflow and Petri Net models. Bioinformatics 2002;18(6):825-837.
[2] Peleg M, Gabashvili IS, Altman RB. Qualitative models of molecular function: linking genetic polymorphisms of tRNA to their functional sequelae. Proc IEEE 2002;90(12):1875-1886.
[3] The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. Nature Genetics 2000;25(May 2000):25-29.
[4] Blake JA, Bult CJ. Beyond the data deluge: Data integration and bio-ontologies. J Biomed Inform 2006;39(2):314-20.
[5] Pe'rez-Rey D, Maojo V, Garci'a-Remesal M, Alonso-Calvo R, Billhardt H, Martin-Sa'nchez F, et al. ONTOFUSION: ontology-based integration of genomic and clinical databases. Comput Biol Med 2006;36(7-8):712-30.
[6] Baker PG, Goble CA, Bechhofer S, Paton NW, Stevens R, Brass A. An ontology for bioinformatics applications. Bioinformatics 1999;15(6):510-520.
[7] Peleg M, Rubin D, Altman RB. Using Petri Net Tools to Study Properties and Dynamics of Biological Systems. J Am Med Inform Assoc 2005;12(2):181-199.
[8] Shahar Y, Miksch S, Johnson P. The Asgaard Project: A Task-Specific Framework for the Application and Critiquing of Time-Oriented Clinical Guidelines. Artificial Intelligence in Medicine 1998;14(1-2):29-51.
[9] Shankar RD, Martins SB, Tu SW, Goldstein MK, Musen MA. Building an explanation function for a hypertension decision-support system. In: Stud Health Technol Inform; 2001. p. 538-42.
[10] Grando A, Peleg M, Glasspool D. A goal-oriented framework for specifying clinical guidelines and handling medical errors. J Biomed Inform 2009;43(2):287-99.
[11] Peleg M, Beimel D. Using closed-world reasoning with OWL to implement policy-based classification. Proc. Third International Conference on Model-Based Systems Engineering; Fairfax, Virginia; 2010.
[12] Nagasaki M, Saito A, Jeong E, Li C, Kojima K, Ikeda E, et al. Cell Illustrator 4.0: A computational platform for systems biology. In Silico Biology 2010;10(1-2):5-26.
[13] Rzhetsky A, Wajngurt D, Park N, Zheng T. Probing genetic overlap among complex human phenotypes. Proc Natl Acad Sci U S A 2007;104(28):11694-9.
[14] Peleg M, Asbeh N, Kuflik T, Schertz M. Onto-clust - a methodology for combining clustering analysis and ontologi-cal methods for identifying groups of comorbidities for developmental disorders. J Biomed Inform 2007;42(1):165-75.
[15] Uzuner O, Goldstein I, Luo Y, Kohane I. Identifying Patient Smoking Status from Medical Discharge Records. J Am Med Inform Assoc 2008;15(1):14-24.
[16] Rzhetsky A, Seringhaus M, Gerstein M. Seeking a New Biology through Text Mining. Cell 2008;134(1):9-13.
[17] Coulet A, Shah N, Garten Y, Musen M, Russ B Altman, Journal of Biomedical Informatics, In Press, Uncorrected Proof, Available online 17 August 2010. Using text to build semantic networks for pharmacogenomics. J Biomed Inform 2010;43(6):1009-19.