

Comment

Life sentences: Ontology recapitulates philology

Sydney Brenner

Address: The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037-1099, USA.

Published: 19 March 2002

Genome Biology 2002, **3**(4):comment1006.1–1006.2

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2002/3/4/comment/1006>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

A few years ago, at a meeting at Dana Point in Southern California, I mistook the number of the room in which our breakfast was to be served and found myself in room full of strangers. I can't remember whether they were the Veterinarians or the Veterans of Southern California (VSOC), but all were very large men wearing very large placards on their chests suspended around their necks with imitation gold chains and bearing the message "HI! I'M CHUCK" or BILL or HANK. With my failing eyesight, I appreciated the two inch high lettering because I did not have to go close up to read the names with a monocle. Unfortunately, our own meeting supplied us with more modest tags, carrying our name and affiliation in small print, and I felt most embarrassed among the VSOC men not to have a sign round my neck acknowledging "HI! I'M SYD".

This way of introducing oneself is typically American. In England, I always said "My name is Sydney Brenner" and in old Mittel Europe I would probably have clicked my heels, bowed and merely said "Brenner". But, then, what's in a name? I have always thought that there is a difference between who you are and what you are called, and that objects are not the same as their names.

I was reminded of this a few months ago, when I met somebody who told me that the coming thing in the post-genomic era is the new science of Ontology. When I asked him what he meant by this, he said it had to do with how we name things in biology and directed me to a paper "Creating the Genome Ontology Resource: Design and Implementation" written by a number of websites and printed in *Genome Research* **11**: 1425, 2001. I urge everybody who has a lot of time to waste to go and read it.

I discovered that an ontology is a structured vocabulary in the form of a directed acyclic graph such that each term is descended from its parent by some defined relationship such as "part of". It is a network where the children can have many parents and, in turn, be parents themselves. The objectives of the Gene Ontology Consortium are to define these structured

hierarchical vocabularies, to describe biological objects using these terms, and to provide computing tools to manipulate these ontologies and connect them to databases.

These aims are laudable. Everybody should know what they are talking about and should use the same language, and computers and databases need to be taught to say the same thing. I doubt the paper's claims that this will solve the problems generated by the endless growth of biological data and I suspect that the best that gene ontology will do is give us a common language in which to express our confusion. My aim is to get out of the Tower of Babel and go somewhere else, rather than try to find a common language to govern it. The connection between Babel and babble is more than a coincidence.

Going back to my VOSC friends' placards, we can now see they were a cheat. The proclamation "I'M CHUCK" told me nothing about the immense biological object carrying it, and it might just as well have said "MY NAME IS CHUCK" and, perhaps in smaller print, "AND WHO I AM IS MY BUSINESS".

The great challenge in biological research today is how to turn data into knowledge. I have met people who think data is knowledge but these people are then striving for a means of turning knowledge into understanding. Knowledge and science are related words and to know, I believe, is to understand. Before rushing to convert genomics to 'genamics' and finding that it is another dead end, we should consider evacuating the Tower. We need a theoretical framework in which to embed biological data so that the endless stream of data, filled with the flotsam and jetsam of evolution, can be sifted and abstracted.

Very simply, the network we should be interested in is not the network of names but the network of the objects themselves. The language of these objects is not the Oxford Dictionary of Molecular Biology - the Ontology Consortium's main source - but that of molecular recognition, the language of molecular biology itself. Objects carry their own names in the form of

the dispositions of nucleotides and amino acids in chemical space, either as linear sequences or on the surfaces of three-dimensional structures. The objects have their own names: they are chemical names written in the language of DNA sequences and the arrangements of amino acids on protein surfaces. It is the interactions between these objects that create the processes that produce outcomes for cells, organs and organism.

This is the real vocabulary that we need to master. It is the language of molecular biology - call it 'mobish' if you like - where fluency needs to be achieved. The bard gave us "What's in a name?" But who was the bard anyway? We know his name was William Shakespeare but was he really William Shakespeare or was he somebody else whose name was Francis Bacon?

This article is reprinted with permission from The Scientist 16(6):14, March 18 2002. The original version can be viewed online at <http://www.the-scientist.com/>