Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Editorial**

Your database is obsolete: The promise of contextual bioinformatics

Martin Gollery*

Center for Bioinformatics, University of Nevada at Reno, Department of Biochemistry, MS330, USA;
E-mail: marty.gollery@gmail.com; Phone: +775 784 7042; * Corresponding author

received February 14, 2007; published online February 19, 2007

Editorial Message:

The latter half of the 1990's was known for a tremendous rise in the number and variety of nucleotide and protein sequence databases. As sequencing technologies became ever more powerful and inexpensive, hundreds of databases were developed for various specific purposes and for numerous model organisms.

Unfortunately, each of these databases contains not only information that is specific to a particular organism, but also schemas, interfaces and nomenclatures that are also typically unique. As a result a great deal of coding, typically hand done in Perl, has been performed to extract correlations between one repository and the next. The end result has been that Bioinformaticists have had to contend with a virtual Tower of Babel and far too much time has been spent with trying to decipher the data repositories versus trying to decipher biological meaning.

Contextual Bioinformatics promises to alleviate the need for all these correlation tables and Perl scripts by utilizing Semantic Web technologies to provide a unified view of this data depending on the context with which the user prefers.

For example, in the traditional Bioinformatics method someone studying horse hemoglobin might look up the sequence in GenBank and then might turned to a gene ontology database to find the function, the location and the processes that it may involve. Next, our intrepid researcher might browse over to the PDB database to find the structure and then travel to a microarray database to find whether this protein is up regulated or down regulated under a certain set of conditions. In this example, the same gene is represented by a different identifier in each of these steps.

Contextual Bioinformatics is a way of looking at all these types of data depending on the context that is of interest at that time. The underlying framework is based on a semantic Web architecture which captures all the information of all the older style, relational databases and eliminates the confusion that is created through a structured methodology. When one system refers to a particular horse hemoglobin gene as 'ABC' and the next refers to the same data as '321'. The sources for this data will remain diverse, but what will change is the nature of the storage. Rather than the consolidation of data on one large server, the information will be distributed at various locations around the world. The information that is viewed by the researcher will be less a function of the design goals of the database administrator and more a function of the desires of the end-user, who will be the one to set the context for data display.

The conversion from relational databases to semantic Web technologies will not happen overnight. People are used to their favorite databases and will continue to visit them for many years in the future. This is as it should be, because it will take some time before semantic Web technologies are able to produce the types of views that Bioinformaticists want and need to do their work. Jumping too quickly into any new technology, no matter how useful, can cause a backlash effect. Just as SQL databases involve a learning curve on the first projects, so does the implementation of contextual bioinformatics, and early adopters ought to plan for additional time for their rollouts.

New databases will continue to crop up. I generated one myself only recently, because one of my collaborators had generated a few thousand EST's for a particular plant. So I went back and did it the old way, using an SQL database simply to get it online quickly. This was enough to meet the needs of my collaborator, and most importantly, it met his timeframe. If this project should happen to grow and to branch out, it will be time to redesign the process, because the day of the monolithic database is over. The Genome Database is dead, long live the new database!

Acknowledgement:

Martin Gollery is supported by NIH-NCRR grant P20 RR-016464. (For more information, see www.ContextualBioinformatics.com)

Editorial: M. Gollery

Citation: Gollery, Bioinformation 1(9): 356 (2007)