

Zeit: Friday, February 25, 2005, 2:00 pm c.t.

Place: Humboldt University, Institute of Computer Science

(Humboldt-Kabinett), Rudower Chaussee 25, 12489 Berlin

Rolf Apweiler

The European Bioinformatics Institute, Wellcome Trust Genome Campus, UK

Standardisation and Integration of Proteomics and Genomics Data

The increasing sophistication to store, manipulate and communicate information has transformed the way life scientists work. The information of life - DNA coding for complex proteins involved in intricate biological processes - has become accessible during the last decades; fortuitously an era when computer hardware and methodology has seen a comparable revolution.

The way scientists deal with data has been completely transformed. It is possible to collect, analyse, communicate and share huge amounts of information rapidly and accurately. Molecular biology, driven by the need to deal with large volumes of information, was quick to build large collections of shared scientific information. Substantial international efforts now support databases of nucleotide sequences, protein sequences and protein structures. Aside from these major projects, numerous other shared information repositories developed, and the research in proteomics will lead to additional databases storing large amounts of biological data in different, often quite specialised databases.

However, while there is a vast amount of valuable information, it often exists as islands, with little interconnection, it can be ill defined and difficult to use, and there is little to help the user distinguish between high quality and low quality information. Improvements in exploiting this information depends on database integration as well as rigorous usage of controlled vocabularies and common standards to describe the data, and it is on such efforts like InterPro, UniProt, GeneOntology, and Integr8 projects that my talk will concentrate.

Guests are welcome!

For further information, please visit http://www.bcbio.de