

Introduction

Review of the selected proceedings of the Fifth International Workshop on Data Integration in the Life Sciences 2008

Amos Bairoch¹, Sarah Cohen-Boulakia^{*2} and Christine Froidevaux²

Address: ¹Swiss Institute of Bioinformatics (SIB) and Department of Structural Biology and Bioinformatics, University of Geneva, Switzerland and ²Laboratoire de Recherche en Informatique (LRI), CNRS UMR 8623, Université Paris-Sud 11, France

Email: Amos Bairoch - Amos.Bairoch@isb-sib.ch; Sarah Cohen-Boulakia* - cohen@lri.fr; Christine Froidevaux - chris@lri.fr

* Corresponding author

from Fifth International Workshop on Data Integration in the Life Sciences 2008 Evry, France. 25–27 June 2008

Published: 22 July 2008

BMC Bioinformatics 2008, **9**(Suppl 8):S1 doi:10.1186/1471-2105-9-S8-S1

This article is available from: <http://www.biomedcentral.com/1471-2105/9/S8/S1>

© 2008 Bairoch et al; licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background

For several years now, there has been an exponential growth of the amount of life science data (e.g., sequenced complete genomes, 3D structures, DNA chips, Mass spectroscopy data) most of which are generated by high throughput experiments. This exponential corpus of data is stored and made available through a large number of databases and resources over the web, but unfortunately still with a high degree of semantic heterogeneity and varying levels of quality. These data must be combined together and processed by bioinformatics tools deployed on powerful and efficient platforms to permit the uncovering of patterns, similarities and in general to help in the process of discovery. Analysing complex, voluminous, and heterogeneous data and guiding the analysis of data are thus of paramount importance and necessitate the involvement of data integration techniques.

DILS 2008 venue

DILS 2008 was the fifth in an international workshop series that aims at fostering discussion, exchange, and innovation in research and development in the area of data integration for the life sciences. The DILS 2008 workshop was held at the University of Evry, in what is known as the 'Genomic Valley' at the heart of the Ile-de-France region, in France. Each previous DILS workshop attracted around 100 researchers from all over the world and has seen an increase of submitted papers over the preceding one. This year was not an exception and the number of submitted papers increased to 54. The 18 papers selected

for presentation at DILS 2008 by the Program Committee cover a wide spectrum of theoretical and practical issues including data annotation, semantic web for the life sciences, and data mining on integrated biological data. 16 of them have been published in the Volume 5103 of *Lecture Note in Bioinformatics of Springer-Verlag*, the two remaining papers have been chosen for publication in this supplement to *BMC Bioinformatics*.

Summary of the selected contributions

The two papers selected for *BMC Bioinformatics* are extended and improved versions of the best papers accepted to DILS 2008. In the following paragraphs, we briefly review them.

The research paper by Jaeger *et al.* [1] addresses the challenging problem of functional annotation of proteins. The methods they designed and developed identify conserved protein interaction graphs and predict missing protein functions from orthologs. Their contribution is two-fold. On the one hand, the procedure they have developed has shown its ability to retrieve more than 80% of the GO annotations for UniProtKb/Swiss-Prot proteins with highly conserved orthologs. On the other hand, new GO annotations have been predicted on a subset of proteins. Results have been validated by biological experts.

The system paper by Jenkinson *et al.* [2] presents the latest updates of the Distributed Annotating System (DAS), increasingly used in the life science community. Exten-

sions presented include dealing with various data types and providing an ontology for protein features. These new functionalities make the latest release of DAS able to span several areas, from genomic sequences to protein interactions.

Workshop Program

In addition to the 18 presented papers, DILS 2008 featured three keynote talks by Olivier Bodenreider, National Library of Medicine, NIH, USA; Peter Karp, SRI International, USA; and Norman Paton, University of Manchester, UK. DILS 2008 also included a tutorial on Bio-ontologies and a session dedicated to updates of biomolecular resources of world-wide importance: the UniProt knowledgebase and the EBI proteomics services.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors wrote and approved the manuscript.

Acknowledgements

DILS 2008 was kindly sponsored by the University of Paris-Sud 11, Microsoft Research who also made available their conference management system, the ENFIN network of Excellence, and the following institutes: IMGT, CEA, SIB, IBISC, and CNRS (LRI and GDR BIM). We are very grateful to the University of Evry for hosting and supporting DILS, the MAISEL school for providing rooms for students, and the Genopole-Evry for its help in the local organization.

As editors of this volume, we thank all the authors who submitted papers, the Program Committee members and the external reviewers for their excellent work. Special thanks go to the local organizers, webmasters, publicity and sponsorship chairs: Patrick Amar, Marie-Dominique Devignes, Nicole Lefevre-Villain, Frédéric Lemoine, Isabelle Mougenot, Bastien Rance, Malika Smail, and Fariza Tahi. Finally, we are grateful for the cooperation of Ms Peters from Biomedcentral in putting this volume together.

This article has been published as part of *BMC Bioinformatics* Volume 9 Supplement 8, 2008: Selected proceedings of the Fifth International Workshop on Data Integration in the Life Sciences 2008. The full contents of the supplement are available online at <http://www.biomedcentral.com/1471-2105/9?issue=S8>.

References

1. Jaeger S, Gaudan S, Leser U, Rebholz-Schuhmann D: **Integrating protein-protein interactions and text mining for protein function prediction.** *BMC Bioinformatics* 2008, **9(Suppl 8)**:S2.
2. Jenkinson AM, Albrecht M, Birney E, Blankenburg H, Down T, Finn RD, Hermjakob H, Hubbard TJ, Jimenez RC, Jones P, Kähäri A, Kulesha E, Macias JR, Reeves GA, Prli(A: **Integrating biological data – the Distributed Annotation System.** *BMC Bioinformatics* 2008, **9(Suppl 8)**:S3.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

