Ondex – A life science data integration framework

Jan Taubert

Centre for Mathematical and Computational Biology
Rothamsted Research
West Common
Harpenden, AL5 2JQ, United Kingdom
jan.taubert@rothamsted.ac.uk

Abstract: Over the last decade biological research has changed completely. The reductionism approach of studying only a few biological entities at a time in the past is being replaced by the study of the biological system as a whole today. Systems Biology seeks to understand how complex biological systems work by looking at all parts of biological systems, how they interact with each other and form the complete whole. This process requires that existing biological knowledge (data) is made available to support on the one hand the analysis of experimental results and on the other hand the construction and enrichment of models for Systems Biology.

Effective integration of biological knowledge from databases scattered around the internet and other information resources (for example experimental data) is recognized as a pre-requisite for many aspects of Systems Biology research. However, systems for the integration of biological knowledge have to overcome several challenges. For example, biological data sources may contain similar or overlapping coverage and the user of such systems is faced with the challenge of generating a consensus data set or selecting the "best" data source. Furthermore, there are many technical challenges to data integration, like different access methods to databases, different data formats, different naming conventions and erroneous or missing data.

To address these challenges and enable effective integration of biological knowledge in support of Systems Biology research, the ONDEX system which is presented here was created. The ONDEX system [1-5] (http://www.ondex.org) provides an integrated view across biological data sources with the aim to enable the user to gain a better understanding of biology from integrated knowledge. ONDEX is supported by BBSRC (http://www.bbsrc.ac.uk/) as part of the System Approaches to Biological Research initiative (SABR) and is now mainly being developed at Rothamsted Research, Manchester University and Newcastle University. The first ONDEX prototype was developed at University of Bielefeld.

Literaturverzeichnis

[1]   A. Lysenko, M. M. Hindle, J. Taubert, M. Saqi, and C. J. Rawlings, "Data integration for plant genomics--exemplars from the integration of Arabidopsis thaliana databases," *Brief Bioinform,* vol. 10, pp. 676-93, Nov 2009.

[2]   J. Köhler, J. Baumbach, J. Taubert, M. Specht, A. Skusa, A. Ruegg, C. Rawlings, P. Verrier, and S. Philippi, "Graph-based analysis and visualization of experimental results with ONDEX," *Bioinformatics,* vol. 22, pp. 1383-90, Jun 1 2006.

[3]   R. Pesch, A. Lysenko, M. Hindle, K. Hassani-Pak, R. Thiele, C. Rawlings, J. Köhler, and J. Taubert, "Graph-based sequence annotation using a data integration approach," *Journal of Integrative Bioinformatics,* vol. 5, 2008.

[4]   J. Taubert, M. Hindle, A. Lysenko, J. Weile, J. Köhler, and C. J. Rawlings, "Linking Life Sciences Data Using Graph-Based Mapping," presented at the Proceedings of the 6th International Workshop on Data Integration in the Life Sciences, Manchester, UK, 2009.

[5]   J. Taubert, K. P. Sieren, M. Hindle, B. Hoekman, R. Winnenburg, S. Philippi, C. Rawlings, and J. Köhler, "The OXL format for the exchange of integrated datasets," *Journal of Integrative Bioinformatics,* vol. 4, 2007.