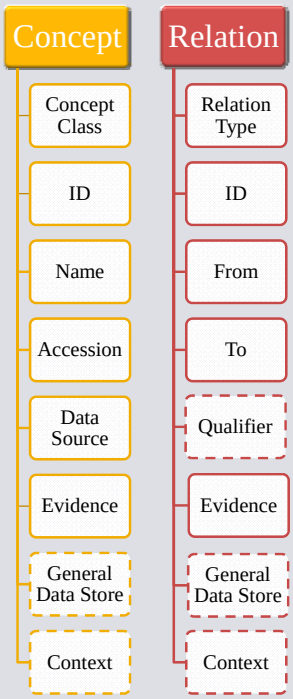
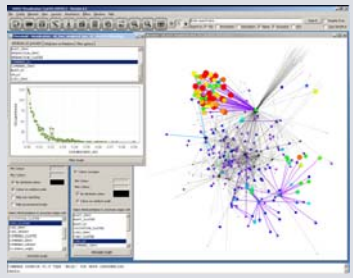
Ondex: data integration and visualisation for the Semantic Web

Catherine Canevet[1], Andrea Splendiani[1], Steve Kuo[1,3], Robert Stevens[2], Phil Lord[3], Chris Rawlings[1]

[1] Centre for Mathematical and Computational Biology, Rothamsted Research, Harpenden, UK
[2] School of Computer Science, University of Manchester, Manchester, UK
[3] School of Computing Science, University of Newcastle, Newcastle, UK

Many systems approaches to biology need to identify, integrate and analyse information that is captured in a myriad of databases which use a wide variety of different formats and access methods. The Ondex data integration platform [1] (www.ondex.org) enables data from diverse biological data sets to be linked together, integrated, analysed and visualised using graph-based techniques. At the basis of Ondex is a graph data structure where entities and properties are associated to classes [2]. This data structure is closely related to the data model of RDF and supports a limited representation of ontologies.

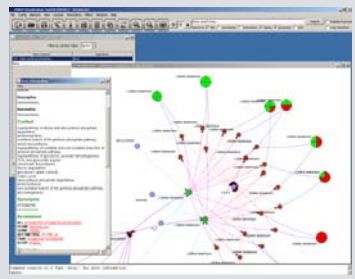




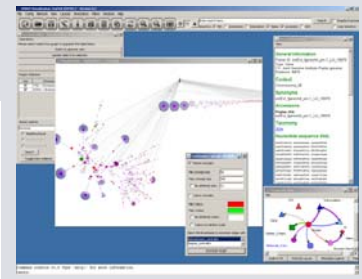
Concept	
Concept Class	Protein
ID	1
Name	ARF1
	Auxin response factor 1
Accession	UniProtKB:Q8L7G0
	TAIR:AT1G59750
	RefSeq:NP_176184
Data Source	UniProtKB
Evidence	IMPD
GDS	taxID 3702
	Sequence MAASNHSSG…
Context	Auxin signaling pathway

Relation	
Relation Type	Has_function
ID	1
From	1
To	2
Qualifier	3
Evidence	IMPD

Concept		
Concept Class	Publication	
ID	3	
Name	16176952	
Accession	NLM:16176952	
Data Source	UniProtKB	
Evidence	IMPD	
GDS	Title	Auxin response …
	Abstract	In plants, both …
	Journal	Development
	Year	2005



Concept	
Concept Class	Biological Process
ID	2
Name	Leaf senescence
Accession	GO:0010150
Data Source	UniProtKB
Evidence	IMPD



In the context of the SABR project (http://www.ondex.org/sabr.html), we are investigating ways to create a mapping between the Ondex data structure and the RDF/OWL. Our objective is to allow Ondex to query, visualise and analyse Semantic Web based knowledge bases [3]. The challenges are to maintain acceptable performance when translating between the RDF and Ondex data models. Once this mapping is implemented, Ondex could also be used to build workflows for the curation and management of such knowledge bases. As a reference use case, we have started the investigation of how Ondex can be used to interact with BioGateway [4], which implements a Sparql endpoint. Apart from creating a mapping that would make BioGateway accessible in Ondex, we are defining a set of guidelines to harmonise the development of Ondex modules to ensure a common representation compliant with this mapping [5].

1. Köhler, J., Baumbach, J., Taubert, J., Specht, M., Skusa, A., Ruegg, A., Rawlings, C., Verrier, P., Philippi, S.: Graph-based analysis and visualization of experimental results with Ondex. Bioinformatics 22 (11):1383-1390 (2006).
2. Taubert, J., Sieren, K.P., Hindle, M., Hoekman, B., Winnenburg, R., Philippi, S., Rawlings, C., Köhler, J.: The OXL format for the exchange of integrated datasets. Journal of Integrative Bioinformatics 4, 62 (2007).
3. Antezana, E., Kuiper, M., Mironov, V.: Biological knowledge management: the emerging rold of the Semantic Web technologies. Briefings in Bioinformatics 10, 392--407 (2009).
4. Antezana, E., Blondé, W., Egaña, M., Rutherford, A., Stevens, R., De Baets, B., Mironov, V., Kuiper, M.: BioGateway: a semantic systems biology tool for the life sciences. BMC Bioinformatics 10 (Suppl 10):S11 (2009).
5. Splendiani, A., Kuiper, M., Rawlings, C.: Toward a new paradigm for user interaction on the Semantic Web to support life sciences investigation. To appear in Proceedings of ISWC-SWUI (CEUR) (2009).

Contact: catherine.canevet@bbsrc.ac.uk **Website:** www.ondex.org

Acknowledgements: This work was supported by BBSRC Grants BBS/B/13640 & BB/F006039/1.