Data Integration in myGrid with Taverna

Duncan Hull
University of Manchester, School of Computer Science, UK

Abstract

Many areas of life sciences research involve integrating terabytes of heterogeneous, distributed and autonomous data available on the Web. The myGrid project has addressed these challenging problems by developing and applying novel grid and semantic web services technology to life science data integration, particularly genome annotation and microarray analysis.

myGrid and Taverna

Data integration is a routine part of many different areas of life science research. To tackle this issue, the myGrid project[1] has developed workflow-based middleware for bioinformaticians which uses semantic web technology for man aging data, metadata, provenance and enabling resource discovery. The flagship product of the project is the Taverna[2] workbench [1, 2]. Taverna uses a registry of publicly available life science services described in RDF and the Web Ontology Language (OWL). This registry includes a wide range of services provided by third-parties around the world including the DNA DataBank of Japan, PDBJ, KEGG, BioMOBY, EBI (Europe), NCBI (USA), Wellcome Trust Sanger Institute and many others. Arbitrary services can be composed into complex workflows to execute many different analysis pipelines over distributed resources (Figure 1).

References

[1] T. Oinn, M. Addis, J. Ferris, D. Marvin, M. Greenwood, T. Carver, M. R. Pocock, A. Wipat, and P. Li. Taverna: a tool for the composition and enactment of bioinformatics workflows. Bioinformatics, 20(17), 2004.

[2] T. Oinn, M. Greenwood, M. Addis, J. Ferris, K. Glover, C. Goble, D. Hull, D. Marvin, P. Li, P. Lord, M. R. Pocock, M. Senger, A.Wipat, and C.Wroe. Taverna: Lessons in creating a workflow environment for the Life Sciences. Concurrency and Computation: Practice and Experience, 2006. Accepted for publication.

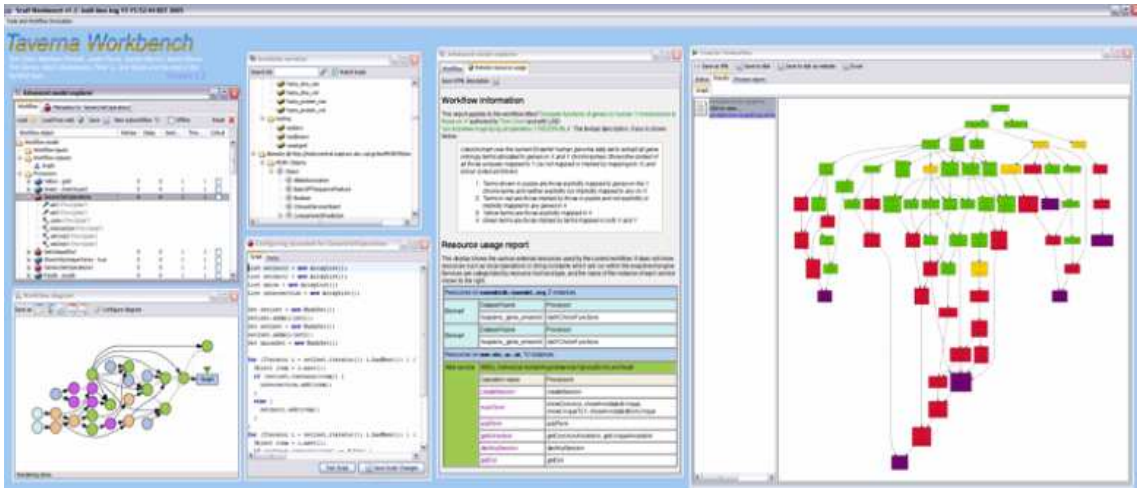

Figure 1 The workbench in action running an example workflow (http://taverna.sourceforge.net/)

[1] http://www.mygrid.org.uk
[2] http://taverna.sourceforge.net