

# A Uniform Approach to Data and Workflow Integration for the Life Sciences

Lucas Zamboulis<sup>1,2,3</sup>, Nigel Martin<sup>1,3</sup>, Alexandra Poulouvassilis<sup>1,3</sup>

<sup>1</sup>School of Computer Science and Information Systems, Birkbeck, Univ. of London

<sup>2</sup>Department of Biochemistry and Molecular Biology, University College London

<sup>3</sup>London Knowledge Lab

## Abstract

Data analysis in the life sciences requires the resolution of conflicts arising from the heterogeneity of data resources and from incompatibilities between the inputs and outputs of services used in the analysis of the resources. The work presented here uses a uniform approach to develop a proteomics grid using e-science technologies to support in silico data analysis. In particular, we present our approach for the integration of data resources and for the reconciliation of services within two bioinformatics projects, ISPIDER and BioMap.

In the ISPIDER project, the data integration architecture supports the combined use of software tools for grid data access (OGSA-DAI), grid distributed querying (OGSA-DQP) and data integration (AutoMed) for the *virtual integration* of a number of distributed, autonomous and rapidly evolving proteomics resources. The BioMap project supports the *materialised integration* of structured and semi-structured functional genomics resources using XML as the unifying data model.

*Workflow integration* is addressed in the ISPIDER project, by using a data integration system (AutoMed) with a scientific workflow tool (Taverna) for the semi-automatic reconciliation of services.