

## Poster II-56

### **A Mediation Framework for a Transparent Access to Biological Data Sources**

***MEDIAGRID Project—A Multidisciplinary Project Supported by the French Minister of Research (ACI-GRID)***

**Contact: Christine Collet**

***Laboratoire LSR IMAG, Saint-Martin d'Herès, France***

Nowadays, data are distributed over the net and are stored in data sources with different formats and data models. This heterogeneity makes the querying task difficult because users have to cultivate their knowledge in terms of formats, query languages and data models. Mediation systems have been proposed to provide access to data distributed all over the net and stored in sources with different formats and data models. Such systems generally provide a global view of the underlying systems and hide several functions on data, queries and transactions for mediating the underlying systems.

Biological data sources are highly heterogeneous in terms of content types (sequences, structures, analysis results, bibliography, etc) and data storage models. Several mediation systems for biology sources have been proposed. All provide a minimal infrastructure for mediating heterogeneous and distributed sources, but none of them is able to represent local sources capability and availability. They are not able to represent characteristics of sources such as functional dependencies or semantic equivalence, or to optimise queries dynamically. All these aspects are essential to react in case of local sources unavailability and to provide alternative or semantic-equivalent query plans. The MEDIAGRID project has as objective to better take up these challenges. The mediation framework we propose will allow to build systems able to (i) consider sources containing weakly structured data which are generated by applications and stored as HTML or XML files and generate mediation queries over these sources; (ii) authorise partial results for queries in case of data sources unavailability; and (iii) be efficient even if queries are complex and/or net traffic slows down. The talk will focus on two aspects of the framework: (i) mediation queries generation (ii) iterative and dynamic query evaluation. It will also show how such functionalities are interesting for biological mediation systems.