



Contribution ID: 124

Type: **not specified**

## **InSilicoLab: A User Workspace Implementing E-science Principles**

*Thursday, 29 March 2012 11:40 (25 minutes)*

### **Description of the Work**

The InSilicoLab portal is a powerful tool, providing means for conducting research to scientists from different domains of science. At the same time, one of its key assumptions is to provide a layer of interaction with the user that would be as much domain-specific as possible.

Combining these, seemingly contradictory, postulates is possible due to a generic, layered architecture of the tool. The topmost (closest to the user) layer of this architecture is entirely concerned with objects and processes from the researchers today's work. However, it also serves as an entrance to the mechanisms hidden in lower layers.

At the opposite side of the researcher's computation, lay the computational and storage resources of an e-Infrastructure.

These two distant layers are joined in the InSilicoLab system by an intermediate layer (called mediation layer). This layer is responsible for transforming a domain-specific scenario of a scientific computation to an execution on the computational infrastructure. This is done with help of a specialised components functioning in this layer. These are:

- Experiment logic, that organises the user's computations into a workflow;
- Automatic parallelisation;
- Execution engine - responsible for conducting and monitoring all the operations defined by the experiment logic;
- Storage structure, that facilitates managing raw files;
- Data model - used for storing data (other than raw files) produced and used by the researcher;
- Metadata description;
- Provenance tracking, that records all the transformations and usage of data relevant to the user and their experimentation process.

The latter functionality allows also repeatability and retraceability of the researcher's computation and composition of larger workflows (in form of DAGs) that join many computations in a repeatable set of actions.

Defining the functionality and providing implementation of the described mediation layer is the main achievement of the work on InSilicoLab.

### **Conclusions**

The implementation of e-science principles in the contemporary research, especially in the case of computationally intensive and conceptually complex in silico experiments, seems an important and natural path for the development of tools for science.

An approach that is presented in this paper, in the shape of the InSilicoLab portal, allows separation of the domain-specific concepts that are close to the researcher from the issues related to the infrastructure they want to use for their computations. This makes the scientists' work much more efficient, as they may focus

only on the information and activities relevant to their research, instead of learning the technical details of the underlying resources.

However, the development of such tools as InSilicoLab cannot be successful without collaboration with the scientist who would use them, and their validation and feedback afterwards.

## Impact

The InSilicoLab portal may serve as a workspace for scientists from different science domains. Making it accessible through Web interface enables researchers from any place in the world, to collaborate on a common research task, represented by an in silico experiment; and to use available large computing power to realise this task. In this way, the InSilicoLab system implements the basic principles of e-science.

The tool was already validated and used by researchers from the domains of computational chemistry and astrophysics.

For the chemists, it enabled performing, so called, conformation scans as well as simple access to chemistry software - Gaussian, GAMESS and TURBOMOLE installed on the grid infrastructure.

In the astrophysics domain, the portal is used for performing Monte Carlo simulations with use of custom software developed especially for this purpose and installed on the grid. The use of InSilicoLab allows performing parameter study to decide on specific parameters of telescopes that will be built for the Cherenkov Telescope Array (CTA) observatory. Such a study is necessary for the design of individual telescopes included in the array. A built-in LFC catalogue manager enables also sharing large input files used by the CTA community for their research.

In both aforementioned domains, stress was put on managing the data obtained in the course of experimentation, which tends to be very complex. The InSilicoLab environment provides a workspace that allows to organise all the researchers' data properly, as well as to use it for further processing. The computation processes can be tracked organised as well to enable even easier experiment management.

## URL

<http://insilicolab.grid.cyfronet.pl>

## Overview (For the conference guide)

E-science is a challenging vision that still has not been fully achieved in contemporary research. Among its principles, collaboration between globally dispersed groups of scientists and use of large-scale computing resources and large data collections are usually pointed as most important. Modern infrastructures, like grids, already meet many of these requirements. However, for full implementation of the e-science vision, tools that go further - towards support of collaboration environments capable of extracting knowledge from data obtained with use of the infrastructures - are needed.

InSilicoLab is an application portal, that supports these aspects of research by facilitating access to computational software deployed on grids and management of data and processes involved in scientific computations. The portal provides mechanisms to track the processes that lead to valuable data, as well as for sharing these data and processes with fellow scientists enabling their collaborative work.

**Primary author:** KOCOT, Joanna (ACC CYFRONET AGH)

**Co-authors:** HAREŹLAK, Daniel (ACC CYFRONET AGH); NOGA, Klemens (ACC CYFRONET AGH); STERZEL, Mariusz (ACC CYFRONET AGH); SZEPIENIEC, Tomasz (ACC CYFRONET AGH)

**Presenter:** KOCOT, Joanna (ACC CYFRONET AGH)

**Session Classification:** Portals & Gateways