

# InSilicoLab: a Domain-Specific Science Gateway

Joanna Kocot, Tomasz Szepieniec, Mariusz Sterzel, Daniel Haręźlak, Maciej Golik,  
Tomasz Twaróg, Piotr Wójcik

ACC Cyfronet AGH, ul. Nawojki 11, 30-950 Krakow  
emails: {j.kocot,t.szepieniec,m.sterzel,d.harezlak,m.golik,  
t.twarog}@cyfronet.pl,piotr.iwo.wojcik@gmail.com

**Keywords:** in silico, scientific computation, data management, Gaussian, GAMESS, CTA, scientific gateway, portal

## 1. Introduction

The contemporary computational research becomes more and more focused on solving domain problems, instead of on the computational aspects of this process. Following this trend, tools like science gateways are built, in order to strip the e-science computations from the technical aspects of the computation itself. A role of the scientific gateway is to present the user with an interface (be it a Web-based portal or even a command-line tool) that is adjusted to the domain and hides away the actual process of submitting and running an application. More advanced gateways often offer also instruments for collaboration between globally dispersed groups of scientists using large-scale computing resources and large data collections.

InSilicoLab [1] is an application portal based on the science gateway concept, that supports the user- and domain-centric aspects of research, by facilitating the access to computational software deployed on grids and management of the data and processes involved in such scientific computations. To meet the researchers' needs, the portal provides mechanisms to track the processes that lead to obtaining valuable data.

## 2. Description of a problem solution

To provide a comprehensive and functional science gateway, we had to define common requirements and challenges that this solution should address. We focused on some aspects of problems common to many scientific domains, forming the following challenges:

- Data-centric approach: the computation is focused on data which are modified, produced or may serve as input to another computation. The flow of data is usually described by a common pattern and can be reproduced and repeated.
- Text or binary result files – the computation results are delivered as schema-free text files or binary files. Therefore, it is difficult to skim through them to find interesting positions – further processing is needed in order to enable that.
- Parameter sweep studies – running a computation which is a scan through certain parameter values with result comparison for each value
- Prototyping: a trial computation run – ending before the final result is achieved or run on a reduced set of data, in order to test if the chosen parameters are promising or valid. The ones considered useful judging from the trial's results are used for full-time computations.
- Complex input specification – the application input specification is well-defined and repeatable, therefore, the input may be generated from a form of a wizard.
- Reusing the outputs of computations as inputs to others.

InSilicoLab portal addresses all of these challenges by providing a problem solving environment, where the scientists are able to access all their data produced or used in different

experiments and simulations. New simulations are run with use of domain-customized forms. Their results are analyzed and summaries are displayed as soon as the computation finishes, leaving the original raw data files on the storage, free to download through the portal. The experiments can be repeated and the results of some of them can be used as inputs to the others with a single click.

A generic architecture of the solution enables support for integration of experiments from many scientific domains, leaving the interface and the data processing processes specific to a specific domain.

### 3. Results

InSilicoLab currently features experiments from two scientific domains:

- Astrophysics – the experiments were validated by several users representing the Cherenkov Telescope Array (CTA) [2] project, and are still improved and extended,
- Computational Chemistry – the computational chemistry experiments in InSilicoLab include specification and parallelization of program runs for popular quantum chemistry packages – Gaussian [3] and GAMESS [4] and TURBOMOLE [5], as well as management and transformations of trajectory files.

Applications from other two domains of science, namely, Astronomy (the Piernik [6] package) and High Energy Physics (Fast- and Full-Sim used by the SuperB [7] project), were initially tested for ability to be integrated in InSilicoLab with positive result.

### 4. Conclusions and future work

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.

The InSilicoLab portal is accessible to all the members of Gaussian VO, vo.plgrid.pl and vo.cta.in2p3.fr at <http://insilicolab.grid.cyfronet.pl/> (stable release for chemistry applications) and <http://ctaportal.grid.cyfronet.pl/> (beta-stage release for the CTA project). The access to the portal is granted on the base of user certificate installed in the browser and respective Virtual Organization membership.

### References

1. J. Kocot, T.Szepieniec, D. Harezlak, K. Noga, M. Sterzel: InSilicoLab – Managing Complexity of Chemistry Computations; in PL-Grid: Building a National Distributed e-Infrastructure.
2. The Cherenkov Telescope Array project: <https://www.cta-observatory.org/>
3. M. J. Frisch et al., Gaussian 09; Gaussian, Inc., Wallingford CT, 2009.
4. M. W. Schmidt, K. K. Baldrige, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery: General Atomic and Molecular Electronic Structure System; in: J. Comput. Chem., 14, 1347-1363(1993).
5. TURBOMOLE – Program Package for ab initio Electronic Structure Calculations. <http://www.turbomole.com/>
6. Piernik MHD Homepage <http://piernik.astr1.umk.pl/>
7. SuperB Homepage <http://superb.infn.it/home>