



Contribution ID: 86

Type: **Oral Presentation**

## **Proteomics portal: an e-Science Gateway for Swiss Bioinformatics**

*Monday, 11 April 2011 14:30 (30 minutes)*

### **Overview**

Bioinformatics is a science that routinely makes use of large computing infrastructures including Grids to understand and analyze the data and to prepare for the next wave of instruments producing even more data. Not only the data volume but also its complexity has increased. The combination of complex analysis tools and large-scale infrastructure make it difficult for the scientists to analyze the data without expert help. There is a strong need to establish automated processes and workflows that can be executed easily without any knowledge of the complexity of the underlying infrastructure. Thanks to a joint project between SystemsX.ch, ETH Zurich and MTA SZTAKI a new web-portal for easy to use automated proteomics analysis was built using the P-Grade Portal technology developed by MTA SZTAKI. The portal has been set up to be able to execute programs on local clusters of ETH and on the national distributed grid, the Swiss Multi-Science Computing Grid based on the ARC middleware.

### **Impact**

Data analysis in the domain of Proteomics has made a lot of advances in the recent years with a very large number of new emerging tools to identify and quantify proteins and peptides through mass spectrometry and liquid chromatography experiments. The problem is that many tools do not adhere to standard data formats, although standards do exist.

Through the Swiss Proteomics Gateway the usage of these tools is automated and very much simplified for the end-user. Experts take care of data transformation into the usable formats, chaining the tools to get the most relevant information and inclusion of the most recent methods. With the gateway scientists will be able to share their analysis workflows and the best methods to look at the data. More importantly, they can automate the workflows for repeated analysis with changed parameters, which was a manual, slow and very error prone process in the past. Developers of algorithms can easily interface with the existing services and test new ideas on real data. Collaborations between lab scientists and algorithm developers is very much facilitated, nationally and internationally; the scientists can focus on their core scientific discoveries as opposed to spend time in the details of data transformations. Due to these advantages the whole Swiss research community interested in proteomics analysis can set up workflows and run applications on the local and national DCIs. There is a strong synergy with the medical sciences, ie. between the Proteomics gateway community and the medical gateway to be later developed by AMC (Amsterdam Medical Centre). A strong interest exists to extend the gateway to be used to analyse other types of data, ie. imaging, microscopy or genomics datasets. The concepts are identical, but of course the data types and algorithms would need to be adapted. By the use of P-GRADE portal technology it is possible to link this effort with the efforts of AMC, giving this development a European dimension.

## Description of the work

The P-Grade Portal provides a web-based User Interface where the users can develop/manage applications on various types of DCIs. The portal is able to submit jobs using g-Lite or ARC-based middlewares as well as on LSF or PBS-based clusters, a capability that was developed for this project. The flexibility as well as the workflow and parallelization capabilities coming with the user-friendly interface were the reasons to choose P-Grade as the baseline technology for this specialized proteomics e-science gateway. To start with, three different types of workflows were developed based on commonly used proteomics tools and applications, set up to be executed on the large clusters. Then corresponding portlets were developed providing easy-to-use web interfaces tailored to the needs of the end-users. These portlets hide the complexity of the workflows and the DCIs allowing the end-users to focus just on the parameters that are important to their own research. The three initial portlets are interfaces to the Transproteomic Pipeline , a workflow to perform label-free quantification and a quality control metric calculation. But also much more complex workflows have been implemented by more advanced users of the platform for dedicated research problems and several new portlets are in the process of being finalized.

There was also a clear need to make the usage of the portal's security components as easy as possible with no shortcuts taken in terms of security. This was achieved by the integration of the Swiss national AAI infrastructure based on Shibboleth2 into the portal and providing access to the Grid automatically using the SLCS service to generate user certificates based on the user's AAI login to the portal. This way the users only have to log in once and the portal can submit jobs to the Grid with no additional steps necessary as in previous portal instances where users needed to generate a proxy certificate outside of the portal and upload it to a myproxy server.

## URL

<https://www.imsbportal.ethz.ch>

## Conclusions

Due to the recent joint development between the ETH Zurich, SystemsX.ch and MTA SZTAKI, a new web-based portal is available to researchers in proteomics. The complexity of the inner component systems are hidden, therefore the end-users can focus on their science, only parameterizing and executing common proteomics analysis workflows by the push of a button. This increases their productivity very efficiently.

**Primary author:** Mr BALASKO, Akos (MTA SZTAKI)

**Co-authors:** Dr MALSTROEM, Lars (ETH Zurich); Dr KACSUK, Peter (MTA SZTAKI); Dr KUNSZT, Peter (SystemX.ch); Mr FARKAS, Zoltan (MTA SZTAKI)

**Presenter:** Mr BALASKO, Akos (MTA SZTAKI)

**Session Classification:** User Environments

**Track Classification:** User Environments - Portals