-- Editorial --

Science Gateways Applying Clouds, Grids and HPC

With IWSG 2014 we featured the sixth edition of the workshop series IWSG (International Workshop on Science Gateways). As international platform the workshop brought together researchers and scientists from different scientific domains, along with science gateways developers, to discuss problems and solutions in the area, to identify new issues, to shape future directions for research, foster the exchange of ideas, standards and common requirements and push towards the wider adoption of science gateways in e-Science.

Science gateways are a community-specific set of tools, applications, and data collections that are integrated together via a web portal or a desktop application, providing access to resources and services on major computing infrastructures such as clouds, clusters, desktop and service grids and supercomputing resources. Science gateways offer the potential to open the utilisation of such infrastructures to wider audiences by providing a customised and easy-to-use user interface to access large computational and data resources. The complexity of the underlying infrastructure can be completely hidden from the end-users by a suitably tailored interface. As interest in science gateways has accelerated in the past few years, an increasing number of new user communities can utilise computing resources in a convenient manner.

The main goal of science gateways is to increase the usability of applications and data management for distributed communities. An important factor for creating successful science gateways is the collaboration between users, science gateway developers and science gateway providers. The research challenges are manifold, and concern different layers of a science gateway infrastructure – from intuitive user interfaces through security features to efficient data, job and workflow management employing parallel and distributed architectures. Novel developments in each of the layers are reflected in novel developments for science gateways: leading-edge web technologies can be applied for user interfaces, apps on smartphones extend the variety of user interfaces and job and workflow management are now supported in cloud infrastructures. Additionally, there is a trend to provide sophisticated data and metadata management.

Workflows themselves also attract greater interest and are becoming a central part of many portals. Workflows allow for the formal representation of a scientific process thus supporting sharing and joint research activities. The integration of workflows into e-Science portals and ensuing growth in complexities are a very active field of research and are a key topic of research and discussion in the community.

IWSG 2014 attracted over 30 participants from nine different countries and 25 submissions, which consists of 17 full papers and eight abstracts. Accepted full papers resulted in 25-minutes presentations and abstracts in lightning talks of 15 minutes. While the 15 accepted full papers are published in these proceedings, the abstracts will be published on the website. Additionally, the workshop featured two high-profile keynotes and an outstanding panel discussion on “Science gateways in the cloud era” performed by research leaders from academia and industry.
The topics of the proceedings include fundamental enhancements in general features of science gateways like extended developments of the security of science gateways as well as novel developments to support science gateway developers, new approaches in workflow management and use cases from diverse research communities.

The paper by Hajnal et al. describes a solution for a very important problem in DCIs. The need for an effective and integrated data management system to support processing at workflows and portal level. Data Avenue services integrate data management support in science gateways and is increasingly becoming a key functionality of the gUSE/WS-PGRADE suite of tools for the development of e-Science portals. Rion et al. present the further developed MyProxy gateway extending the features of the original MyProxy gateway for demands raised by resource providers and a strong trend of science gateways moving to the web. The gateway provides a RESTful API and solves problems like exposing the passwords to third parties and applies the standard OAuth2. Interesting new possibilities are explored by Vitello et al. who describe a mobile application connected to a workflow-enabled framework performing visualization and data filtering. Providing access to large distributed computing infrastructures from the convenience of a mobile device provides not only flexibility and easy access to scientists but also facilitates collaboration and information sharing between researchers. Balasko et al. present a solution for large metadata process for agricultural communities. The DRUPAL content management system is used as it is adopted by several of their web-based information providers while orchestration is achieved by workflow-based orchestration.

Smaller local user communities often have specific requirements and prefer using their own custom-made tools that fulfill their expectations better than more generic widely available community gateways. However, developing a custom solution for such local user community could be expensive and time consuming when compared to the number of users potentially supported. Jaghoori et al. explore how reusing, customizing and tailoring existing gateways speeds up the development process and results in a suitable solution within a reduced timeframe. The paper by Kondikoppa et al. focuses on the implementation of the Replica Exchange Statistical Temperature Molecular Dynamics application as a Map-Reduce application executed on HPC resources. The challenges posed by fine-tuning and scaling Hadoop over a cluster is tackled by different runtime analysis of two experiments on biological systems. Because of support for large-scale capacity over multiple DCIs, map-reduced based RESTMD offers an interesting example of the next generation applications.

Cloud computing infrastructures are gaining more and more emphasis as the distributed computing infrastructure behind science gateways. Fiehe et al. describe the experiences and results of the European EASI-CLOUDS project that develops a platform for a convenient service delivery, and illustrate the capabilities of the solution via the example of a medical service cloud. The paper by Caballer et al. is concerned with the deployment of scientific virtual infrastructures on the cloud as Infrastructure as a Service (IaaS). The platform has been applied for two key areas in bioinformatics: the execution of the BLAST biomedical application on Hadoop clusters and the process for single-node infrastructures with a very specific installation and configuration to perform NGS sequencing. Even though the use cases are very specific, the platform has been proven to increase the usability and performance for applications in general. Apache Airavata is introduced by Pierce et al. and forms an API to support science gateway developers with building blocks for accessing diverse computing and data infrastructures. As platform as a service it allows developers to focus on the specific demand of a community, while offering flexible and modular basic features as open-source API.

The paper by Massimino et al. introduces ACID, the Astronomical & Physics Cloud Interactive Desktop that was developed by the Astrophysical Observatory of Catania within the SCI-BUS European project. ACID enables both on-line interactive and off-line workflow submission-based analysis of astrophysical data, and fully integrated into the WS-PGRADE/gUSE science gateway framework. The paper by Gordienko et al. describes the IMP Science Gateway Portal, a gUSE/WS-PGRADE-based portal for the study of simulations of complex behavior of various
nanostructures. WS-PGRADE workflows are used to orchestrate the various complex scientific simulations, gUSE capabilities of interfacing to different DCIs is used to match optimal resources to different job profiles.

The paper "Expansion of Quantum Chemical Metadata for Workflows in the MoSGrid Science Gateway". present a quantum chemical use case in the workflow-enabled MoSGrid science gateway to create workflows and metadata for two commonly used quantum chemical applications. The applications resulted in diverse values while processing the same methods and energy functions defined by the users. The differences has been clarified and corrected via close collaboration with the developers of the quantum chemistry applications but the occurrence also clearly shows the requirement for a detailed analysis to adapt workflows for diverse tools implemented based on the same theory. A further paper in the context of the MoSGrid project "Meta-Metaworkflows for Combining Quantum Chemistry and Molecular Dynamics in the MoSGrid Science Gateway" describes the results in the area of computational chemistry using interdisciplinary workflows inside the community regarding quantum chemistry and molecular dynamics. An additional abstraction layer for structuring workflows into meta-workflows allows for further workflow interoperability and enhanced re-usability of building blocks of workflows in diverse research domains. Pierantoni et al. follow a similar approach and offer a novel method for the use of workflows for heliophysics to tackle the unique characteristics of that community. The need for re-usability of complex workflows suggests a two-layer architecture were sub-workflows (highly re-usable atomic components) are combined in higher-level meta-workflows that tackle scientific challenges. As members of the community may use different workflow systems, their preferences are catered for by extending this two-layer architecture with the capabilities of executing workflows on different languages.

Bozic et al. also make widely utilized existing tools available via a web-based application service that is integrated to WS-PGRADE as a portlet. The solution provides intuitive user interface for the eSBMTools package that is a UNICORE-based grid application for native structure based modelling. Targeted bioinformatics, biophysics and structural biology community end users can access the tools via the MoSGrid Portal.

*Sandra Gesing, Tamas Kiss, and Gabriele Pierantoni*