

Science Gateway Workshops 2015 Special Issue Conference Publications

**Sandra Gesing · Nancy Wilkins-Diehr ·
Michelle Barker · Gabriele Pierantoni**

Received: 2 November 2016 / Accepted: 2 November 2016 / Published online: 14 December 2016
© Springer Science+Business Media Dordrecht 2016

The special issue combines extended versions of submissions to the International Workshop on Science Gateways (Budapest, June, 2016), the Gateway Computing Environments workshop (Boulder, CO, September, 2016) and the International Workshop on Science Gateways – Australia (Brisbane, October). These events serve to foster the world-wide community of those involved in developing science gateways – advanced web portals used by the scientists and scholars to conduct their work. Four main areas have been addressed in the submissions: domain gateways, gateways adapted for diverse research infrastructures, data management in gateways and gateway workflows.

S. Gesing (✉)
University of Notre Dame, 123 Information Technology
Center, Notre Dame, IN 46556, USA
e-mail: sandra.gesing@nd.edu

N. Wilkins-Diehr
San Diego Supercomputer Center, University of California,
San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
e-mail: wilkinsn@sdsc.edu

M. Barker
National eResearch Collaboration Tools and Resources,
Parkville VIC 3010, Canberra, Australia
e-mail: michelle.barker@nectar.org.au

G. Pierantoni
University of Westminster, 115 New Cavendish Street
London W1W 6UW, Marylebone, UK
e-mail: G.Pierantoni@westminster.ac.uk

These workshops are among the few occasions that bring together researchers that are domain expert in different scientific fields or the humanities with technology experts that specialize in building and designing the technology (gateways and portals) This combination of expertise and point of views allows a fertile discussion on both usability and technology. Furthermore, it allows technology providers to better understand the shortcomings of the developed systems and the domain experts to better understand the solutions underlying the portals they commonly use. Both sides show an increase in the variety and sophistication of the various solutions: science gateways now span from specific solutions to generic services that serve entire scientific communities while the underlying technology witnesses a great richness in the implementation of various components including portal technologies, workflows, and tools for collaboration and access to distributed system including grids, clouds and single clusters. In addition to work in the development of enabling technologies and specific domain solutions, there is a growing body of work to study the theoretical aspects of workflows that will help the development of more effective future systems.

Several domain-specific science gateways are featured in this special issue. In their submission “Advancing a Gateway Infrastructure for Wind Turbine Data Analysis” [1], the authors describe the need for a data gateway (VAVID) to monitor and process sensor information from wind turbines. Numerical

simulation and sensor measurement combine and activities include development of improved techniques for data compression, new methods for data management and interactive data visualization. Sensor data comes from up to 600 wind turbines, each of which produce 100 MB per hour. Care is taken to design an interface useable by engineers and it is now possible to monitor for ice on the rotor blades to address safety concerns. The VAVID algorithms, workflows and infrastructure are all described in detail, with particular attention paid to data provenance. Scalability, both as it relates to algorithms and storage systems is an item for future work.

In “Beyond a Disease Registry: An Integrated Virtual Environment for Adrenal Cancer Research” [2] the authors describe the proliferation of web-based databases that provide phenotypic and genotypic information – disease registries. They describe extensions to this work that could include bio-banking, bio-sample data tracking, support for analysis of clinical trials, data transfer to and from these trials and search and analytical capabilities. The contribution describes just such a system, the ENS@T-CANCER VRE, a virtual research environment that supports research on four types of adrenal tumors. Also highlights are its use in a large international clinical trial. The gateway has galvanized previously fragmented registry efforts.

“The essential components of a successful Galaxy service” [3] describes efforts to launch a production scale Galaxy service for a large research organization. Gateways such as Galaxy allow bench biologists in genomics to analyze large amounts data without unix command line computing familiarity. The Commonwealth Scientific and Industrial Research Organisation (CSIRO) (<http://www.csiro.au>) is Australia’s national science agency. The bioscience community is spread across 50 sites throughout the country. The Galaxy implementation serves the needs of all of these researchers. A pilot was launched in 2012 with a move to production in 2013. The submission describes several components, from an initial survey, to the IT deployment, user support, training, collaborative tools and social media and scaling the services to HPC. This contribution provides a very nice, full-featured outline that others can follow to achieve a successful Galaxy implementation.

Research infrastructures are essential for supporting users with access to computation and data to

conduct their research. In “Using Science Gateways for Bridging the Differences between Research Infrastructures” [4] the authors analyze challenges in general to use science gateways for diverse research infrastructures, e.g., policies and available tools. They describe the different layers of science gateway architectures, which have to be considered for porting a science gateway. The general analysis is underpinned via the challenges faced and actions performed for porting the MoSGrid science gateway, which has been originally developed for the German National Grid Infrastructure and its use with XSEDE and PRACE infrastructures.

Piontek et al. [5] present latest achievements of their work to lower entry barriers for new users and remove limitations in science gateways exploiting HPC, Grid and Cloud infrastructures. The paper goes into detail for a number of mature and successful science gateways serving diverse research domains and built on their application tools: the QCG (Quality in Cloud and Grid) middleware, in particular Vine Toolkit, QCGPortal and QCG-Icon. The deployed software solutions have reached recently a critical mass of active users in the PLGrid e-infrastructure in Poland and they also discuss how changing technologies, visual design and user experience could impact the way science gateways should be re-designed in general.

The paper “From lesson learned to the refactoring of the DRIHM science gateway for hydro-meteorological research” [6] goes into detail for a workflow-enabled science gateway tailored for full hydro-meteorological simulations. This multidisciplinary target area requires a workflow composed by complex and heterogeneous model engines and an extensive set of configuration parameters. The first version of the science gateway was developed during the DRIHM project using the gUSE science gateway toolkit. The authors present the lessons learned, which guided subsequently a refactoring process. They describe besides the technological aspects the emerging need of a strong interplay between ICT and other domain-specific communities, together with coherent policies in the management of data, computational resources and software components that represent the ecosystem of science gateways.

The importance of data management and analysis of data on a large scale in science gateways is especially reflected in two papers. Gugnani et al.

[7] investigate how workflow systems and science gateways can be extended with Big Data processing capabilities. They suggest a generic approach based on infrastructure-aware workflows and they describe an implemented proof of based on the WS-PGRADE/gUSE science gateway framework and its integration with the Hadoop parallel data processing solution based on the MapReduce paradigm in the cloud. The provided analysis demonstrates that the methods described to integrate Big Data processing with workflows and science gateways work well in different cloud infrastructures and application scenarios, and can be used to create massively parallel applications for scientific analysis of Big Data.

The paper “Privacy Preserving Geo-Linkage in the Big Urban Data Era” [8] is concerned with Big Data technologies and the possibilities arising from open data initiatives on urban data. The authors present the Australian Urban Research Infrastructure Network (AURIN - www.aurin.org.au) project, which provides a single gateway with access to over 2000 data sets from over 70 major and typically definitive data-driven organizations across federal and State government, across industry and across academia. They highlight a novel geo-privacy supporting solution that provides seamless and secure access to individual data. The solution meets all of the associated privacy and confidentiality concerns, without obliging the users to obtain ethical approval or any other hurdles that are normally put in place on access to and use of sensitive data. This provides a paradigm shift in secure access to sensitive data with geospatial content.

Workflows tools comprise part of the generic side of science gateways development, enabling the creation of building blocks that can be re-used across the science gateways with divergent community needs. There are five papers in this issue that examine issues related to workflows for science gateways. Farkas et al. [9] provide an example of new developments in this field, detailing a solution to integrate workflow-oriented science gateways with multi-cloud systems. Their solution provides a generic method that could be adopted by other gateways, and is built on both the principles of integrating science gateways with multi-cloud systems, and analysis of existing systems. Their integration of WS-PGRADE/gUSE and Cloud-Broker Platform provides a concrete example of a complete cloud-enabled science gateway platform for

a diverse set of use-cases, with different customization methodologies, with the availability to use both academic and commercial cloud resources.

The second paper in this chapter, “A Formal Approach to Support Interoperability in Scientific Meta-workflows” by Arshad et al. [10], contributes to the definition and analysis of meta-workflow approaches for complex scientific meta-workflows. The authors examine how workflow interoperability plays a role in achieving the objective of reusing existing workflows from potentially different workflow systems to achieve more complex and experimentation minimizing workflow design and testing efforts. The paper presents a proof of concept from astrophysics which successfully demonstrates the use of the concepts of meta-workflows and workflow interoperability within a scientific simulation platform, and identifies a need for a standard to describe scientific workflows to facilitate workflow sharing.

The paper written by Susana Sánchez-Expósito et al. [11] addresses issues related to Distributed Computing Infrastructures, focusing on how to efficiently exploit the application of these for analysing and visualising large datasets to deal with the data deluge that astrophysics is facing. With science gateways now being enriched with advanced tools that not only enable the scientists to build their experiments but also to optimize their adaptation to different infrastructures, this work presents a method to build this kind of tools in a manner that will enable astronomers to be unaware of the technical complexity behind these.

The work of Karoczkai, Kertesz and Kacsuk in “A Meta-Brokering Framework for Science Gateways” [12] focuses on a more specific problem in resource management for science gateways developed for the gUSE family. Their metabrokering framework solves this issue through resource priority services, and the concept and services can be re-used for science gateways supporting the BES standard. The benefits of this approach are demonstrated through the extension of WS-PGRADE/gUSE with this concept.

Finally, Kacsuk et al. [13] present “Infrastructure aware scientific workflows and infrastructure aware workflow managers in science gateways”, to offer a solution to an important problem of interest to a broad community of researchers. The paper that investigates the interoperability problem for workflows running on heterogeneous Grid infrastructures, and

presents a solution in the form of infrastructure-aware workflows and workflow managers. This is practically implemented by a new service called Occopus, and integrated in the WS-PGRADE/gUSE portal, as a demonstration how to apply the proposed solution in real-life situations.

Acknowledgments The editors are once again pleased with the quality and variety of submissions to the science gateway workshops and the interest in and effort taken to submit extended papers for this special issue. We are grateful to both authors and reviewers for their invaluable help in making this possible. Special thanks go also to the editor of the Journal of Grid Computing and the submission contact at Springer.

References

1. Aguilera, A., Grunzke, R., Habich, D., Luong, J., Schollbach, D., Markwardt, U., Garcke, J.: Advancing a Gateway Infrastructure for Wind Turbine Data Analysis. *J Grid Computing* (2016). doi:[10.1007/s10723-016-9376-9](https://doi.org/10.1007/s10723-016-9376-9)
2. Sinnott, R.O., Beuschlein, F., Effendy, J., Eisenhofer, G., Gloeckner, S., Stell, A.: Beyond a Disease Registry: An Integrated Virtual Environment for Adrenal Cancer Research. *J Grid Computing* (2016). doi:[10.1007/s10723-016-9375-x](https://doi.org/10.1007/s10723-016-9375-x)
3. McGrath, A., McMahon, S., Li, S., Ludbey, J., Ho, T.: The Essential Components of a Successful Galaxy Service. *J Grid Computing* (2016). doi:[10.1007/s10723-016-9379-6](https://doi.org/10.1007/s10723-016-9379-6)
4. Gesing, S., Krüger, J., Grunzke, R., Herres-Pawlis, S., Hoffmann, A.: Using Science Gateways for Bridging the Differences between Research Infrastructures. *J Grid Computing* (2016). doi:[10.1007/s10723-016-9385-8](https://doi.org/10.1007/s10723-016-9385-8)
5. Piontek, T., Bosak, B., Ciznicki, M., Grabowski, P., Kopta, P., Kulczewski, M., Szejnfeld, D., Kurowski, K.: Development of Science Gateways Using QCG — Lessons Learned from the Deployment on Large Scale Distributed and HPC Infrastructures. *J Grid Computing* (2016). doi:[10.1007/s10723-016-9384-9](https://doi.org/10.1007/s10723-016-9384-9)
6. D'Agostino, D., Danovaro, E., Clematis, A., Roverelli, L., Zereik, G., Galizia, A.: From Lesson Learned to the Refactoring of the DRIHM Science Gateway for Hydro-meteorological Research. *J Grid Computing* (2016). doi:[10.1007/s10723-016-9377-8](https://doi.org/10.1007/s10723-016-9377-8)
7. Gugnani, S., Blanco, C., Kiss, T., Terstyanszky, G.: Extending Science Gateway Frameworks to Support Big Data Applications in the Cloud. *J Grid Computing* (2016). doi:[10.1007/s10723-016-9369-8](https://doi.org/10.1007/s10723-016-9369-8)
8. Sinnott, R.O., Bayliss, C., Bromage, A., Galang, G., Gong, Y., Greenwood, P., Jayaputera, G., Marques, D., Morandini, L., Nogoorani, G., Pursultani, H., Sarwar, M., Voorsluys, W., Widjaja, I.: Privacy Preserving Geo-Linkage in the Big Urban Data Era. *J Grid Computing* (2016). doi:[10.1007/s10723-016-9372-0](https://doi.org/10.1007/s10723-016-9372-0)
9. Farkas, Z., Kacsuk, P., Hajnal, A.: Enabling Workflow-Oriented Science Gateways to Access Multi-Cloud Systems. *J Grid Computing* (2016). doi:[10.1007/s10723-016-9388-5](https://doi.org/10.1007/s10723-016-9388-5)
10. Arshad, J., Terstyanszky, G., Kiss, T., Weingarten, N., Taffoni, G.: A Formal Approach to Support Interoperability in Scientific Meta-workflows. *J Grid Computing* (2016). doi:[10.1007/s10723-016-9383-x](https://doi.org/10.1007/s10723-016-9383-x)
11. Sánchez-Expósito, S., Martín, P., Ruiz, J.E., Verdes-Montenegro, L., Garrido, J., Sirvent, R., Ruiz Falcó, A., Badia, R.M., Lezzi, D.: Web Services as Building Blocks for Science Gateways in Astrophysics. *J Grid Computing* (2016). doi:[10.1007/s10723-016-9382-y](https://doi.org/10.1007/s10723-016-9382-y)
12. Karoczkai, K., Kertesz, A., Kacsuk, P.: A Meta-Brokering Framework for Science Gateways. *J Grid Computing* (2016). doi:[10.1007/s10723-016-9378-7](https://doi.org/10.1007/s10723-016-9378-7)
13. Kacsuk, P., Kecskemeti, G., Kertesz, A., Nemeth, Z., Kovács, J., Farkas, Z.: Infrastructure Aware Scientific Workflows and Infrastructure Aware Workflow Managers in Science Gateways. *J Grid Computing* (2016). doi:[10.1007/s10723-016-9380-0](https://doi.org/10.1007/s10723-016-9380-0)