Preface

Science gateways represent a community-specific set of tools, applications, and data collections that are integrated via a web portal or a desktop application, providing access to resources and services of distributed computing infrastructures (DCIs). Science gateways offer the potential to open the utilisation of DCIs to wider audiences by providing a customised and easy to use interface to the users enabling access to large computational and data resources, e.g., distributed file systems, sensor data, data from lab instruments. 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 distributed computing and distributed data resources in a convenient manner.

IWSG 2016 was the 8th event of the European workshop series IWSG (International Workshop on Science Gateways) and has taken place in Rome, Italy. The workshop forms an international platform to bring together researchers and scientists from different scientific domains, along with science gateways developers and research infrastructure providers. IWSG offers the venue for discussing problems and solutions in the area, for identifying new issues as well as for shaping future directions for research and fosters the exchange of ideas, standards and common requirements and thus, pushes towards the wider adoption of science gateways in diverse communities.

The workshop has featured three thought-provoking keynotes and the International Coalition on Science Gateways (ICSG) has been kicked off at the event. 13 full papers have been presented as 30-minutes talks, while the nine accepted abstracts resulted in corresponding lightning talks.

The submissions reflect the trends in the area of science gateways and address three main areas: sustainability of software, reproducibility of science and management of data on a large scale. The need for sustainability of science gateways and reproducibility of science leads to solutions connecting research and business and building flexible, reliable research infrastructures and service-oriented architectures. Sustainability and usability are also tackled via novel models and programming frameworks targeting the developer and user community to ease the creation of science gateways. Reproducibility is a major driver for advanced workflow development and sophisticated data management tailored for analysis and modeling of data on a large scale. These proceedings include 12 of the accepted full papers which are briefly introduced below highlighting the current trends in the research area of science gateways.

Models and programming frameworks are addressed by the following three papers.

A Model for Information and Action Flows Connecting Science Gateways to Distributed Computing Infrastructures tackles the challenge of complex and unstructured flow of action and information in science gateway infrastructures and the diversity of actors using and maintaining science gateways. The authors suggest a model for such flows considering the diversity of roles and need

for information to aid with designing and implementing more flexible and user-friendly science gateways and workflow management systems in the future. M. Assante et al. present gCube, a system enabling hybrid data infrastructures, which facilitate the dynamic definition and operation of virtual research environments. It offers a comprehensive set of data management capabilities on various types of data and a rich array of mediators to interface well-established infrastructures and information systems from various domains. gCube operates the D4Science.org infrastructure and the paper gives an overview on the gCube system. *Rosemary: A Flexible Programming Framework to Build Science Gateways* goes into detail for the Rosemary science gateway framework, whose core offers the three fundamental science gateway functions data, computing and collaboration management. Rosemary is flexible to changes in e-Infrastructures and user community requirements and provides a generic data model, RESTful API, and responsive UI that can be customized through programming to build customized science gateways. It is used for three prototypes of science gateways. The paper presents the design considerations and the features applied in the three prototypes.

Two papers go into detail for scientific use cases.

An Innovative Workspace for The Cherenkov Telescope Array presents a prototype workspace developed at INAF that aims at providing innovative solutions for the Cherenkov Telescope Array (CTA) community. Two different user interaction models, connected to an authentication and authorization infrastructure, have been implemented in this workspace. The first one is a workflow management system accessed via a science gateway and the second one is an interactive virtual desktop environment. *A Science Gateway for Biodiversity and Climate Change Research* elucidates that there is a strong need for multidisciplinary skills, tools and a large variety of heterogeneous, distributed data sources to better understand the mutual interaction between climate change and biodiversity. The EUBrazilCloudConnect project provides a user-centric research environment built on top of a federated cloud infrastructure across Europe and Brazil to serve scientific needs. One of the science gateways implemented in this project is BioClimate that focuses on climate change and biodiversity research. It provides end-users with a highly integrated environment, addressing mainly data analytics requirements.

Containerization and service-oriented architectures are the main focus of the following four papers.

Joe Stubbs et al. introduce Endofday, a workflow engine that orchestrates a directed acyclic graph (DAG) of computational science apps where the nodes of the DAG are Docker containers. The endofday engine enables users to execute entire workflows of science applications without actually installing any of the applications themselves. As an example, the authors present the Validate system, a suite of software applications for testing the accuracy and precision of Genome Wide Association methods, and illustrate how it can be run using Endofday without the need for installation. Additionally, they show how Endofday integrates with the Agave platform's application catalog. *MiCADO*

- *Towards a Microservice-based Cloud Application-level Dynamic Orchestrator* gives an introduction to the concepts of the new framework called MiCADO. It focuses on dynamically adjustable infrastructure resources and their provision through microservices in a cloud environment. MiCADO is intended to serve the variable needs of not only academic communities but also the ones from data intensive commercial applications. Daniele D'Agostino et al. present the EXTraS gateway. It offers a platform for astrophysicist to extract time dependent information from the data of the European Photon Imaging Camera which was collected in the 16 years of its operation. The gateway concept is based on the use of microservices in order to provide a largely resilient service. *Fast Access to Remote Objects 2.0 - A renewed web portal to ENEAGRID distributed computing resources* introduces Faro 2.0 a science gateway for the fast access of remote objects. Through the usage of modern technologies such as HTML5, JavaFx, CSS3 and jQuery a collection of centralized services is offered to the users. The project offers a workflow-oriented access to ENEAGRID resources.

Advanced workflow capabilities and data management is tackled by the following three papers.

Luis de la Garza et al. give an overview about the recent developments for KNIME2gUSE. The authors highlight the challenges for the conversion of abstract and concrete layers of different workflow languages into each other. In particular the usage of common tool descriptors eases the conversion of specific nodes, facilitating the creation of complex workflows with KNIME and porting them to gUSE/WS-PGRADE. *Milky Way Analysis through a Science Gateway: Workflows and Resource Monitoring* introduces the most recent developments for the VIALACTEA Science Gateway. The gateway offers a platform for the integration and analysis of all kind of infrared and radio data about our home galaxy, the milky way. The newly introduced agile software development concept led to a set of highly useful resource monitoring services. Richard Grunzke et al. gives an insightful presentation about the relevance of metadata information with respect to the management of scientific data involving multiple communities. The relevance in the context of open and reproducible science becomes apparent from this work, which in large parts is based on work accomplished within the MASi project.

Thanks to the support of the University of Notre Dame, its Global Gateway program and its Center for Research Computing, we were able to host the workshop in Notre Dame's Global Gateway just beside the Colosseum. The workshop was supported by the National Science Foundation (Award ID 1632929), EuCheMS (European Chemical Sciences) and the IEEE Technical Area on Science Gateways. We would like to thank the PC members, authors, presenters and participants. Without their work and contributions, IWSG 2016 would not have been such a success.

Sandra Gesing
Jens Krüger
Chairs of IWSG 2016

Organization

Chairs

Sandra Gesing, University of Notre Dame, US
Jens Krüger, University of Tübingen, Germany

Program Committee

David Abramson, University of Queensland, Brisbane, Australia
Viktor Achter, University of Cologne, Germany
Akos Balasko, MTA SZTAKI, Hungary
Antun Balaz, Institute of Physics Belgrade, Serbia
Roberto Barbera, University of Catania and INFN, Italy
Michelle Barker, NeCTAR, Melbourne, Australia
Ugo Becciani, INAF Catania, Italy
Andre Brinkmann, University of Mainz, Germany
Daniele D'Agostino, CNR-IMATI, Italy
Silvia D. Olabarriaga, Academic Medical Centre, Amsterdam, The Netherlands
Zoltan Farkas, MTA SZTAKI, Hungary
Geoffrey Fox, Indiana University, US
Sandra Gesing, University of Notre Dame, US
Helen Glaves, British Geological Survey, UK
Aaron Golden, Albert Einstein College of Medicine, US
Richard Grunzke, University of Dresden, Germany
Keith Jeffery, RDA, UK
Peter Kacsuk, MTA SZTAKI, Hungary
Joohyun Kim, Louisiana State University, Baton Rouge, US
Tamas Kiss, University of Westminster, London, UK
Dagmar Krefting, University of Applied Sciences, Berlin, Germany
Jens Krüger, University of Tübingen, Germany
Krzysztof Kurowski, PSNC, Poznan, Poland
Miklos Kozlovszky, MTA SZTAKI, Hungary
Daniele Lezzi, Barcelona Supercomputing Center, Spain
Robert Lovas, MTA SZTAKI, Hungary
Suresh Marru, Indiana University, US
David Meredith, STFC Daresbury Laboratory, UK
Ivan Merelli, CNR-ITB, Italy
Jarek Nabrzyski, University of Notre Dame, US
Horacio Perez-Sanchez, Universidad Catlica San Antonio de Murcia, Spain
Gabriele Pierantoni, TCD, Dublin, Ireland
Marlon Pierce, Indiana University, US
Mats Rynge, University of Southern California, US
Cevat Sener, METU, Turkey
Richard Sinnott, University of Melbourne, Australia

Gergely Sipos, EGI.eu, Amsterdam, The Netherlands
Ian Taylor, University of Cardiff, UK and University of Notre Dame, US
James Taylor, John Hopkins University, Baltimore, US
Gabor Terstyanszky, University of Westminster, UK
Chen Wang, CSIRO ICT Centre, Marsfield, Australia
Nancy Wilkins-Diehr, San Diego Supercomputing Center, US
Eric Yen, Academia Sinica Grid Computing Center, Taipei, Taiwan