

# **International Symposium on Grids & Clouds 2019 (ISGC 2019)**

**Sunday, 31 March 2019 - Friday, 5 April 2019**

**Academia Sinica**

## **Scientific Programme**

## Main Theme: Efficient and safe processing of FAIR Open Data

To achieve the full potential of Open Data and Open Science, scientists should be able to focus on their area of interest and be shielded from the internal complexity of e-infrastructures and the needs to manually deal with the different data formats, input and output constraints of used tools, the authentication and access control and any other technical or technological obstacles that are still part of the current data processing and analysis environments. This is emphasized by the FAIR concept -- data must be Findable, Accessible, Interoperable and Re-usable. New approaches are emerging, hiding the complexity of the underlying computing and data fabrics, exposing just integrated views through scientific portals and gateways, notebooks and other virtual computing environments that promise to enhance the efficiency of work with advanced and complex contemporary e-infrastructures. On the other hand, possible privacy issues related to the Data collected and new, easy-to-deploy analysis methods like those applying deep neural networks, remind us of the need for proper security tools and environments and create new challenges for dealing with private and sensitive data and derived information.

The goal of ISGC 2019 is to create a face-to-face venue where individual communities and national representatives can present and share their contributions to the global puzzle and contribute thus to the solution of global challenges. We cordially invite and welcome your participation!

## Physics (including HEP) and Engineering Applications

Submissions should report on experience with physics and engineering applications that exploit grid and cloud computing services, applications that are planned or under development, or application tools and methodologies. Topics of interest include: (1) End-user data analysis; (2) Management of distributed data; (3) Applications level monitoring; (4) Performance analysis and system tuning; (5) Workload scheduling; (6) Management of an experimental collaboration as a virtual organization; (7) Comparison between grid and other distributed computing paradigms as enablers of physics data handling and analysis; (8) Expectations for the evolution of computing models drawn from recent experience handling extremely large and geographically diverse datasets.

## Biomedicine & Life Sciences Applications

During the last decade, research in Biomedicine and Life Sciences has dramatically changed thanks to the continuous developments in High Performance Computing and highly Distributed Computing Infrastructures such as grids and clouds, but also in big-data solutions to deal with the explosion in genomic data. This track aims at discussing problems, solutions and application examples related to this area of research, with a particular focus on non-technical end users. Submissions should concentrate on practical applications and solutions in the fields of Biomedicine and Life Sciences, such as Drug discovery, Structural biology, Bioinformatics, Medical imaging, Public health applications / infrastructures, High throughput (grid and cloud-based) data processing/analysis, Distributed data computing and services, and Big data management issues. Submissions should ideally highlight how the availability and use of Big Data has enabled new processes for or dramatically evolved the scope of their research.

## **Earth & Environmental Sciences & Biodiversity Applications**

Natural and Environmental sciences are placing an increasing emphasis on the understanding of the Earth as a single, highly complex, coupled system with living and dead organisms. It is well accepted, for example, that the feedbacks involving oceanic and atmospheric processes can have major consequences for the long-term development of the climate system, which in turn affects biodiversity, natural hazards and can control the development of the cryosphere and lithosphere. Natural disaster mitigation is one of the most critical regional issues in Asia. Despite the diversity of environmental sciences, many projects share the same significant challenges. These include the collection of data from multiple distributed sensors (potentially in very remote locations), the management of large low-level data sets, the requirement for metadata fully specifying how, when and where the data were collected, and the post-processing of those low-level data into higher-level data products which need to be presented to scientific users in a concise and intuitive form. This session would in particular address how these challenges are being handled with the aids of e-Science paradigm.

## **Humanities, Arts, and Social Sciences (HASS) Applications**

Disciplines across the Humanities, Arts and Social Sciences (HASS) have critically engaged with technological innovations such as grid- and cloud computing, and, most recently, various data analytic technologies. The increasing availability of 'born digital' data has led to an increasing interest in analysis methods such as natural language processing, social network analysis, machine learning and text mining. These developments pose challenges as well as opening up opportunities and members of the HASS Disciplines across the Humanities, Arts and Social Sciences (HASS) have critically engaged with technological innovations across various data formats. The increasing availability of data, ranging from social media text data to consumer big data has led to an increasing interest in analysis methods such as natural language processing, social network analysis, machine learning and text mining. These developments pose challenges as well as opening up opportunities and members of the HASS community have been at the forefront of discussions about the impact that novel forms of data, novel computational infrastructures and novel analytical methods have for the pursuit of science endeavours and our understanding of what science is and can be. The ISGC 2019 HASS track invites papers and presentations covering applications demonstrating the opportunities of new technologies or critically engaging with their methodological implications in the Humanities, Arts and Social Sciences. Innovative application of analytical tools for survey data, social media data, and government (open) data are welcomed. We also invite contributions that critically reflect on the following subjects: (1) the impact that ubiquitous and mobile access to information and communication technologies have for society more generally, especially around topics such as smart cities, civic engagement, and digital journalism; (2) philosophical and methodological reflections on the development of the techniques and the approaches by which data scientists use to pursue knowledge.

## **Virtual Research Environment (including Middleware, tools, services, workflow, ... etc.)**

Virtual Research Environments (VRE) provide an intuitive, easy-to-use and secure access to federated computing resources for solving scientific problems, trying to hide the complexity of the underlying infrastructure, the heterogeneity of the resources, and the interconnecting middleware. Behind the scenes, VREs comprise tools, middleware and portal technologies, workflow automation

as well as security solutions for layered and multifaceted applications. Topics of interest include but are not limited to: (1) Real-world experiences building and/or using VREs to gain new scientific knowledge; (2) Middleware technologies, tools, services beyond the state-of-the-art for VREs; (3) Innovative technologies to enable VREs on arbitrary devices, including Internet-of-Things; and (4) One-step-ahead workflow integration and automation in VREs.

## **Data Management & Big Data**

The rapid growth of the data available to scientists and scholars – in terms of Velocity and Variety as well as sheer Volume – is transforming research across disciplines. Increasingly these data sets are generated not just through experiments, but as a byproduct of our day-to-day digital lives. This track explores the consequences of this growth, and encourages submissions relating to two aspects in particular: firstly, the conceptual models and analytical techniques required to process data at scale; secondly, approaches and tools for managing and creating these digital assets throughout their lifecycle.

## **Network, Security, Infrastructure & Operations**

Networking and the connected e-Infrastructures are becoming ubiquitous. Ensuring the smooth operation and integrity of the services for research communities in a rapidly changing environment are key challenges. This track focuses on the current state of the art and recent advances in these areas: networking, infrastructure, operations, and security. The scope of this track includes advances in high-performance networking (software defined networks, community private networks, the IPv4 to IPv6 transition, cross-domain provisioning), the connected data and compute infrastructures (storage and compute systems architectures, improving service and site reliability, interoperability between infrastructures, data centre models), monitoring tools and metrics, service management (ITIL and SLAs), and infrastructure/systems operations and management. Also included here are issues related to the integrity, reliability, and security of services and data: developments in security middleware, operational security, security policy, federated identity management, and community management. Submissions should address solutions in at least one of these areas.

## **Infrastructure Clouds and Virtualisation**

This track will focus on the use of cloud computing, mainly but not exclusively Infrastructure-as-a-Service (IaaS) and virtualization technologies in large-scale distributed computing environments in science and technology. We solicit papers describing underlying virtualization and "cloud" technology, scientific applications and case studies related to using such technology in large scale infrastructure as well as solutions overcoming challenges and leveraging opportunities in this setting. Of particular interest are results exploring usability of virtualization and infrastructure clouds from the perspective of scientific applications, the performance, reliability and fault-tolerance of solutions used, data management issues. Papers dealing with the cost, price, and cloud markets, with security and privacy, as well as portability and standards, are also most welcome.

## **Science Gateways, Volunteer Computing, Shared Resources and Long Tail Science**

This track welcomes contributions dealing with technologies, concepts and applications that support management of and easy access to very large distributed systems, desktop grids and resources provided through volunteer (unguaranteed) computing. Special focus will be on support of the long tail of science, making the ad hoc provided resources available to small teams or even individual researchers. Science gateways and other kinds of portals, specific interfaces to connect and use the systems, but also new ways to contribute and to combine volunteered and institutional computing resources are expected. The topics cover new technologies of related software frameworks, recent application developments, as well as infrastructure operation and user support techniques. Special focus will be on the (1) Interoperability with other and integration in other e-infrastructures, esp. via Science gateways and other kinds of portals (2) Data management and (3) Quality of service in such environments (4) Novel uses of volunteer computing and Desktop Grid (5) Best practices and (social) impacts.

## **Supercomputing, High Throughput Computing, Accelerator Technologies, and their Integration**

There is a growing availability of powerful computing resources using a combination of general purpose, accelerators, GPGPUs and many-core processors, available through public grids (e.g., EGI and OSG) and public/private clouds (e.g., Amazon EC2), as well as through coordinated access to supercomputing resources (e.g. PRACE). Using, accessing, aggregating and managing these High Performance and High Throughput Computing (HPTC) infrastructures, whose components are under control by different resource providers, is still quite challenging. This session solicits recent research and development achievements and best practices in exploiting these computing resources available around the world. The topics of interest include, but are not limited to the followings: (1) Use of virtualization techniques (including containers) to support access to and portability across different heterogeneous (HPTC) systems (2) Delivery of and access to heterogeneous HPTC resources through grid and cloud computing (as a Service models) (3) Experiences, use cases and best practices on the development and operation of large-scale HPTC applications (4) Integration and interoperability to support coordinated federated use of different HPTC e-infrastructures (5) Robustness and reliability of HPTC applications and systems over a long-time scale (6) Performance on HPTC resources of applications developed for more traditional architectures.