

# Effectiveness and efficiency of computational scientific methods are not enough

Improving usability, reproducibility and sustainability of scientific methods via science gateways

Computational scientific methods tackle an increasing breadth and diversity of topics – analysing data on a large scale and accessing high-performance computing infrastructures, cutting-edge hardware and/or instruments. In the last decade novel technologies such as next-gen sequencing or the Square Kilometre Array telescope, the world largest radio telescope, have evolved, which allow creating data in exascale dimension. While the availability of this data salvage to find answers for research questions, which would not have been feasible before, maybe even not feasible to ask before, the amount of data creates new challenges, which obviously need novel computational solutions. Such new solutions have to consider integrative approaches, which are not only considering the effectiveness and efficiency of data processing but improve

usability, reproducibility and sustainability, especially tailored to the target user communities. The goal of science gateways, also called virtual research environments or virtual laboratories, are following exactly this goal to provide an easy-to-use end-to-end solution hiding the complex underlying infrastructure. They support researchers with intuitive user interfaces to focus on their research question instead of becoming acquainted with technological details.

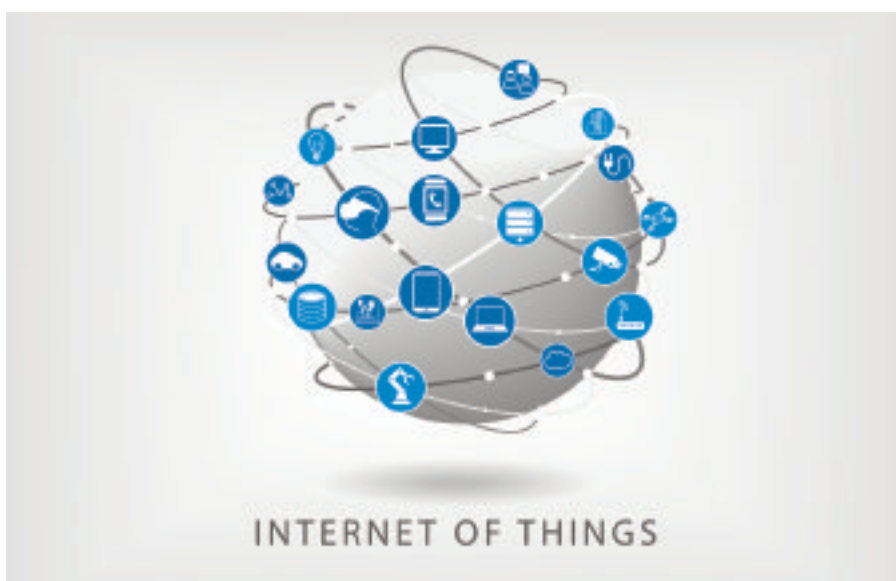
## Exciting fast evolving computational landscape

The computational landscape has never so fast evolved like in the last decade. On the hardware side from graphical processing units to sensors to smartphones, which possess computational capabilities, which are equivalent to capabilities of high-performance architectures only ten years

ago. The Internet has revolutionised our lives for the last 30 years and the Internet of Things (IoT) expand its vision to a network of physical devices, vehicles, buildings and other items embedded with electronics, software, sensors, and network connectivity that enable these objects to collect and exchange data. A further trend in the last 10 years include agile web frameworks allowing for much more user-friendly user interfaces in web browsers than ever before. Thus, users can apply rich-featured applications in browsers without the need to install or maintain desktop applications on their computers. Web-based libraries and application programming interfaces set the stage for modular, well-designed frameworks behind the scenes. The Cloud paradigm and containerisation approaches address the need for flexibility for accessing computational and data resources in the scale fitting to the users' and applications' needs.

## Multidisciplinary teams

Researchers are experts in their fields and to make their scientific discoveries faster and easier it is crucial to support them not only with effective and efficient computational methods but also with easy-to-use solutions, thus, science gateways. Science gateways lower the hurdle of employing the available complex architectures and infrastructures while simulating and model the often complex underlying theories. Multidisciplinary teams are necessary for creating a science



## PROFILE

gateway for a specific research domain and the major challenge is the breadth of topics associated with science gateways. The challenges are manifold, from intuitive user interfaces and security features to efficient data and workflow management, and parallelisation of applications employing parallel and distributed computing architectures. Besides researchers from the target domain and computational scientists and/or developers, calling on experts for specific aspects, such as librarianship, statistics and machine learning, are essential, since it is not practicable for single persons to become an expert in every aspect of a science gateway.

### **Research without geographical boundaries**

Research communities collaborating on distinct topics are often geographically distributed and not located at one campus, in one country or on one continent. A key challenge is to integrate multiple disparate community clouds across the world into a single research infrastructure, where the user effectively does not have to worry about selecting the most appropriate service for them. Service selection should be transparent to the user – and they should be able to simply use the tools they need, share and access the data that they want, in as simple, rapid and limitation free way as possible. Users may want to share data not before they published or patented it. Hence, privacy and security needs have to be met in such an infrastructure but it should not hamper the sharing capabilities if they are desired. Thus, a science gateway for a research community is ideally a single point of entry irrespective of the location of a researcher. Science gateway frameworks are designed to exploit diverse underlying computing and data infra-

structures – campus-wide, national or international ones from clusters to Grid to Cloud. Such “pluggable” frameworks ease the development of a science gateway and the access to infrastructures and the main challenge does not lay on technical side, but on policy. The computing and data resources are mostly funded via local or national mechanisms and agencies and the exploitation of such resources is also often limited to such boundaries. Policies need to be aligned independent of geographical boundaries to allow for topic-related research within a community.

**“The Internet has revolutionised our lives for the last 30 years and the Internet of Things (IoT) expand its vision to a network of physical devices, vehicles, buildings and other items embedded with electronics, software, sensors, and network connectivity that enable these objects to collect and exchange data.”**

### **Reproducibility of science**

Researchers – independent to which community they belong – aim at sharing results and publishing them. Ideally, the results can be easily reproduced and science gateways provide a promising vehicle to allow for reproducibility of computational created data. Reproducibility is a cornerstone of science but computational methods used in science gateways are among others dependent on operating systems, tools in diverse versions and local or distributed data. Studies on shared computational methods present that only 20% are reproducible and reusable out of the box. For solving such problems in science gateways, the different sharing possibilities have to be analysed and – where necessary – tools, data and workflows have to be provided in diverse infra-

structures and via various job and data management systems to address such challenges.

### **Sustainability of scientific methods**

A further challenge for science gateways and computational scientific methods in general is their sustainability. Projects in academia depend on funding and often involve PhD students on the development side. More centralised research programmer teams, who can provide more experience and contribute to sustainability of solutions, are rather rare at universities and there is still a lack of incentives for interested developers to stay in academia. One of the future challenges for science gateways will be to increase the sustainability and getting less dependent on successful proposals. The US National Science Foundation has recognised the importance of this topic for research and has funded a Science Gateway Community Institute to support not only teams in developing science gateways but also to help communities to find a way to sustain their favourite science gateway for conducting their research.



**Sandra Gesing**  
**Research Assistant Professor**  
University of Notre Dame  
Tel: +1 574 298 0690  
sandra.gesing@nd.edu  
www.sandra-gesing.com  
www.twitter.com/sandragesing