A COMMUNITY FOR OPEN ACCESS MOLECULAR SCIENCE

Stefano Evangelisti, Institut de la Recherche sur le systemes atomiques et moleculaires complexes, Universitè Paul Sabatier, Toulouse (FR)

Sonja Herres Pawlis, Rheinisch-Westfälische Technische Hochschule, Department of Chemistry, Aachen (DE)

Fabrizio Esposito, CNR-NANOTEC, Consiglio Nazionale delle Ricerche, Bari (IT)

Antonio Laganà, Leonardo Pacifici, Dipartimento di Chimica, Biologia e Biotecnologie, Università di Perugia, Perugia (IT)

Gabor Terstyanszky, Center for Parallel Computing, University of Westminster, London (UK)

Abstract

The development of a European Open Science Cloud will enable scientific communities to share and reuse research data across disciplines and borders. The Chemistry, Molecular and Materials Science and Technologies Virtual Research Community is structuring itself for that purpose through a community application for HPC compute time and the assemblage of a funding proposal to the next March 2016 Horizon 2020 calls.

INTRODUCTION: THE OPEN SCIENCE

The present strategy of Europe with respect to publicly funded research and innovation aims at the creation and dissemination of scholarly knowledge towards a multitude of stakeholders through the “development of a European Open Science Cloud that will enable sharing and reuse of research data across disciplines and borders, taking into account relevant legal, security and privacy aspects"[1]. The most important characteristics of this approach (called OPEN SCIENCE) are:

One-stop-shop: a single access support providing a common overall view of apps, tools, services and data helping researchers to select and access them regardless of the related source.

Comprehensive open access to repositories of data, information and scientific publications: reuse of harmonized and comparable outcomes of publicly funded

ISSN: 2279-8773

research for end-users thanks to a policy supporting European-level Open Science, to guarantee interoperability among e-infrastructures coupled with harmonised policies to protect data integrity and privacy where this is an important concern for the involved scientific domain.

Research-centricity: full engagement of researchers and research communities (including those from the private sector) in the design and development of services responsive to innovation needs.

Free (or credit for work done and pay per use) access to shared computing resources: different types of computing, storage, and analytics capabilities need to be available. Computing is still too local (resources and tools are not always available to the public), not connected to the data infrastructures. This implies transparent, distributed or centralised access.

Interoperability of tools, open source tools and software: tools that enable Open Science and reference tools that allow methodology preservation by adopting standard procedures and policies that include common standards, tools, and services, integrated between the different e-Infrastructures.

In this context, the engagement of user communities and Long Tail of Science should get suitably organized so as to enhance stronger interaction among the members and the adoption of common standards and services.

The most important services needed for the Open Science Cloud should be:

1. Hardware platforms combining HPC (High Performance Computing) and HTC (High Throughput Computing)

2. Computing storage analytics, unified data management and data sets access

3. Collaborative environments to share/discover/access digital resources (e.g. software, data, scientific publications)

4. Forum for sharing expertise, experience and knowledge

5. Technologies on which user defined applications can be designed, built, extended and provided

6. Scalable computing tools and applications

7. Registries of existing services/software/data

8. Intellectual Property Rights management to integrate science with business

ISSN: 2279-8773

9.Tools to link published information to tools used to obtain published results

10. Tools make results and experiments comparable

A PROPOSAL FOR THE CHEMISTRY, MOLECULAR AND MATERIALS SCIENCE AND TECHNOLOGIES COMMUNITY

For this purpose the European DCC (Division of Computational Chemistry) of EUCHEMS jointly with the COMPCHEM VO (Virtual Organization) is assembling a new H2020 funding proposal aimed at building a CMMST (Chemistry, Molecular and Materials Science and Technologies) VRC (Virtual Research Community). The key objective of such VRC is to assemble in a bottom up fashion a distributed computing environment enabling synergistic research in which members of the community provide and utilize as-a-service support for the use of their data and programs in terms of credits redeemable as a better share of the community resources (compute time, hardware and software tools and interfaces, links to workflows for multi-scale applications, research funds, collaborative projects, etc.). Credits will be assigned using a community agreed METRICS for evaluating QoS (Quality of Service), QoU (Quality of User) and RR (Resources Ranking) of the activities offered as-a-service using the GriF (Grid Framework) and GCreS (Grid Credit System) tools developed by the VRC itself [2,3].

To this end during the last EUCO 10 DCC Conference held in Fulda (August 31 – September 3, 2015) the CMMST VRC and the COMPCHEM VO have jointly decided to establish a steering committee (made by the research groups of the authors of the present paper) and circulate a questionnaire inquiring about the use of science gateways, workflows, repositories of applications, workflows and data. The questionnaire enquiries also about the typology of the researches carried out by the community members, of the societal challenges addressed and the cooperative initiatives activated with the SMEs. It inquiries also in more detail about the number of active developers of new applications and of users of existing packages (either passive or adaptation for different purposes) as well as about the type of computation (HTC, HPC, web service, ..), of computing infrastructure (cloud, cluster, grid, supercomputer,..), of data infrastructure (file system, database, data source, ..), of data size (input, intermediate, output, ..), of user interfaces (Command line Interface, GUI web page, GUI portal, …), of security requirements (access control, authorization, data confidentiality).

On the ground of the past experience two main types of compute resources will be considered for the CMMST-VRE: HPC of large scale facilities and HTC of distributed networked computers. In order to corroborate the proposal with a qualified use of HPC it has also been decided to submit a community application to PRACE for a

grant of supercomputer time (see the enclosed annex 1, 2 and 3 files). For the HTC the proposed approach is to integrate in the proposal the facilities offered by already structured DCI resource providers (EGI.eu NGIs, research centres, research groups, individual users, etc.) to be rewarded (by means of a quality evaluation of the services provided) on the funds of the project.

As to the middleware, an agreed common environment will be adopted by involving as members of the project the related providers, leveraging on existing centres of competence and relying on providers already equipped with the necessary expertise and infrastructures.

Human resources will be provided by the institutions partners of the proposal. Such synergistic model will canalize the activities of the community members (by means of incentives and depending on their wish) to higher levels of involvement as detailed in the table.

Membership Level	Short Description
1. User	*Passive*: Runs programs implemented by other VO members.
	Active: Implements at least one program for personal usage.
2. Software Provider	*Passive*: Implements at least one program for use by other members.
	Active: Manages at least one implemented program for collaborative usage.
3. Infrastructure Provider	*Passive*: Confers to the infrastructure at least a small cluster of processors.
	Active: Contributes to deploy and manage the infrastructure.
4. Manager (Stakeholder)	Takes part to the development and the management of the VO.

Leveraging on the past experience of the COMPCHEM VO, the use of the resources can be managed, monitored, and turned to account using the GriF tool [4,5]. GRIF is a Framework made of two Java servers (YC and YR the Consumer and the Registry servers) and one Java client (YP the Provider). The entry-points to the computational platforms are the User Interfaces which are able to capture, out of the data supplied by the monitoring sensors of the Distributed Computing Infrastructure, the information needed to properly manage the relevant computational applications and articulate them in sequential, concurrent or alternative quality paths by adopting a Service Oriented Architecture (SOA) and Web Services. This allows at the same time the guided search of the compute resources on the DCI and the Evaluation of the QoU. The computational services provided, are analyzed and used

ISSN: 2279-8773

to compose the submission, the monitoring, and the results recollection of molecular science simulations.

Leveraging on the past experience of the COMPCHEM VO, the use of the Grid Credit System GCreS tool [6] to reward both the QOU (the Quality of the users exploiting the compute resources and applications) and the QoS (the Quality of the services provided by the users to the other members of the community) it will be possible to assign to the users the right amount of Credits (according to mechanisms agreed by the members of the community). Such credits, redeemable in terms of a preferential utilization of the community resources (selection of compute systems, DCI services, low and high-level capabilities, memory size, cpu/wall time, storage capacity plus financial resources) will not only foster collaboration among the members but will also entitle related researchers to participate to the multi-competence teams which apply for the most challenging bids. This will result for the whole community in an enhancement of the competition among different teams (the so called competitive collaboration).

REFERENCES

1] Community Workshop on the Open Science Cloud: Shaping the Open Science Cloud of the Future, https://indico.egi.eu/indico/sessionDisplay.py?sessionId=62&confId=2544#20151113 and https://documents.egi.eu/public/RetrieveFile?docid=2637&version=1&filename=OSC_Position_Paper.pdf

2] A. Laganà, C. Manuali, L. Pacifici, S. Rampino, A. Costantini, A new bottom up approach to the CMMST-VRE, VIRT&L-COMM.7.2015.8

3] A draft of the materials used for writing the final version of the proposal is available in A. Laganà, Research and Innovation actions. Chemistry, Molecular & Materials Sciences and Technologies Virtual Research Environment (CMMST-VRE), VIRT&L-COMM.6.2014.1

4] C. Manuali, S. Rampino, A. Lagana', GRIF: A Grid Framework for a Web Service Approach to Reactive Scattering, Comp. Phys. Comm. 181, 1179-1185 (2010)

5] C. Manuali, A. Lagana', GRIF: A New Collaborative Framework for a Web Service Approach to Grid Empowered Calculations, Future Generation of Computer Systems, 27(3), 315-318 (2011)

6] C. Manuali, A. Lagana', A Grid Credit System Empowering Virtual Research Communities Sustainability, Lecture Notes Computer Science, 6784, 397-411 (2011)