TOWARDS A QRS EUROPEAN OPEN SCIENCE CLOUD COMMUNITY?

A. Laganà and C. Manuali, Department of Chemistry, Biology and Biotechnologies, University of Perugia, Perugia, Italy

1 – INTRODUCTION

As editors of the on-line Open Journal System (OJS) Magazine VIRT&L-COMM developed within the European Grid Infrastructure (EGI) project, we (AL and CM) are happy to host some extended papers originating from contributions presented at the last Quantum Reactive Scattering (QRS) workshop organized by G. Lendvay in Balatonföldvár, Hungary on September 2022 **[1]**. As suggested by the title, the aim of the present paper is to stimulate a discussion on the possible evolution of the QRS group into an EGI Virtual Research Community (VRC) and its possible integration into the European Open Science Cloud (EOSC) environment **[2]**. This would imply the enhancement of the production of shared flexible workflows for the proper collaborative handling of related jobs, as well as the simplification of the extraction of the most significant outputs for use in the cloud while progressively hiding the resources needed to run them. The ultimate goal of the paper is, therefore, to stimulate the singling out of the mechanisms enhancing the sustainability of the potential QRS VRC by exalting relevant collaborative computational processes.

2 – GRID COMMUNITIES FOR THE EUROPEAN OPEN SCIENCE CLOUD

In the Grid, geographically distributed resources (such as storage devices, data sources and supercomputers) are exploited and dealt with as a single unified platform. As a result, grid computing is offered as a utility provided by third parties whose associated service-oriented elaborations are transparent, open, cost-effective, reliable, scalable and flexible. Grids and Clouds are distributed computing systems sharing different objects including physical (computational power, storage devices and communication capacity) and virtual (operating systems, software and licenses, tasks, applications and services) resources. Typically, a client layer is used to display information, receive user input and communicate with the other layers. Accordingly, a transparent and network-independent middleware plays the role of mediator when connecting clients with requested and provisioned resources, when balancing peak loads between multiple resources and customers, when regulating the access to devices introducing bottle necks, when monitoring activities and gathering statistics for the management of the system and the evaluation of quality. The middleware used for this endeavour has to be reliable and always available since it provides interfaces to over- and under-lying layers, which can be used by programmers to shape the system according to their needs. Different resources can be geographically dispersed or hosted in the same data centre. Furthermore, they can be interconnected. Accordingly, regardless of the architecture of the resources, they appear to the user as a single entity. The priorities for further developing EOSC are set in the SRIA document (Strategic Research and Innovation Agenda, https://www.eosc.eu/sria) defining, as from the year 2023, the EOSC general framework of research, development and innovation activities. In particular, the procured EOSC environment should:

- be robust, secure, scalable, flexible and user-centric;
- be constantly improved and upgraded following user feedbacks and the state-of-the-art of the underlying core technologies;
- offer high quality of service management compliant with industrial standards and

ISSN: 2279-8773

- provide a superior user experience, usability and ease of use for a very large number of users, with available functionalities.

Accordingly, as spelt in the EOSC Procurement tender, market players from industry and research communities are the EOSC general framework committed to develop a set of innovative, modular, customisable and composable services while “research will be placed at the centre of the initiative” so as to ensure “academic sovereignty of research data”.

3 – THE EVOLUTION OF MOLECULAR SCIENCE GRID AND CLOUD COMMUNITIES

The key projects that led to the evolution of the European grid technologies for sciences were the EGEE (Enabling Grids for E-science, https://eu-egee-org.web.cern.ch/index.html) I, II and III ones. At the same time, the molecular science (MS) projects that established research and innovation networks suited to exploit EGEE grid technologies were those of the Chemistry domain of the COST (Cooperation in Science and Technology) European organization. They were, in particular, the Action D23 (https://www.cost.eu/actions/D23/) and the Action D37 (https://www.cost.eu/actions/D37/). As mentioned in the relevant web pages, Action D23 exploited the potentialities of meta and grid computing activities of 46 research groups from 19 countries “to build the so called European Meta Laboratories (clusters of geographically distributed Laboratories working in a coordinated way on a common project by sharing manpower, hardware and software), fostering innovative solutions for chemical applications and a new paradigm for collaborative research”. Later on, as mentioned in the final scientific report of D37 (ISBN 978-88-97228-03-05), the Chemistry community was able to implement in that Action the Grid computing infrastructures for: *Computational photochemistry and photobiology* (coordinated by Hans Lischka, University of Vienna, AT), *Quantum dynamics engines for grid empowered molecular simulators* (coordinated by Antonio Laganà, University of Perugia, IT), *E-science for learning approaches in molecular science* (coordinated by Osvaldo Gervasi, University of Perugia, IT), *Code interoperability in computational chemistry* (coordinated by Elda Rossi, CINECA, IT), *Computational chemistry workflows and data management* (coordinated by Thomas Steinke, Zuse Institute Berlin, DE) covering so far the key areas of computational chemistry and middleware development in those areas. For illustrative purposes we mention here that the amount of integrated normalized cpu hours of the Grid infrastructure utilized by the COMPCHEM Virtual Organization (VO) **[3]** in its first year of life were 5 millions provided jointly by the National Grid Infrastructures (NGI) of Italy (NGI_IT), France (NGI_FRANCE), Greece (NGI_GRNT), Spain and Portugal (NGI_Ibergrid) and Poland (NGI_PL). During the next EGI.eu project, EGI-Inspire (https://www.egi.eu/about/egi-inspire/), the COMPCHEM yearly utilization of cpu time increased by a factor of 5 and the CMMST **[4]** VRC (https://wiki.egi.eu/wiki/VT_Towards_a_CMMST_VRC) was established in May 2014 with the purpose of generalizing the use of the relevant Grid Empowered Molecular Simulator (GEMS) of ref. **[5]** to various innovative applications on Distributed Computing Infrastructures (DCI). Due to the associated high demand of High Performance Computing (HPC) by the CMMST VRC, specific efforts were paid to connect the Grid node of Perugia to the supercomputer networks of PRACE and XSEDE. GEMS is a typical as-a-service molecular simulator (high level electronic structure computing, fitting, nuclei dynamics and statistics) as a building block for the CMMST applications like:

- multi-scale accurate process simulation of smart energy carriers in combustion, energy storage, aerothermodynamics, etc.;
- computational modelling and design of materials, supra-molecular and biochemical phenomenologies, etc.;

- handling extended (huge) data bases to extract data-embedded information;
- management of distributed Knowledge to support training and education in molecular and materials sciences.

4 – QUALITY OF SERVICE AND USERS FOR VO/VRC SUSTAINABILITY

Without permanent public funding collaborative endeavours (such as the above mentioned VOs and VRCs) cannot survive long enough if no metric providing a solid quality assessment procedure suited to attract resources. This is, for example, the case of the European Chemistry Thematic Network, ECTN, https://ectn.eu/committees/virtual-education-community/echemtest/ for the release of the Individual Proficiency Certificates (IPC)s that will be discussed in some detail later. Having this is mind, some solid tools allowing a positive cost/benefit analysis were introduced a few years ago as reported in detail in ref. **[6]**. In this section, we illustrate only the main features of such progress and of the tools introduced to trigger such mechanism.
In particular we show that:
1 - the balancing of a) the relevant costs associated with the QoS (Quality of Service (namely GFD (Grid Framework Data) and GMD (Grid Middleware Data) both based on GriF/Grid Sensors) and b) the relevant credits associated with the QoU (Quality of User namely GSD (Grid Service Data) and GUD (Grid User Data) both based on Services/Users in the lhs Gcres section;
2 - their linking to the most appropriate low level capabilities of the middleware ensuring a proper implementation a) in the rhs **GriF** upper side through YR, YP and YC of the i.e. the Grid Registries, Providers and Consumers) and b) in the rhs **Grid** lower side section through the User Interface (UI).

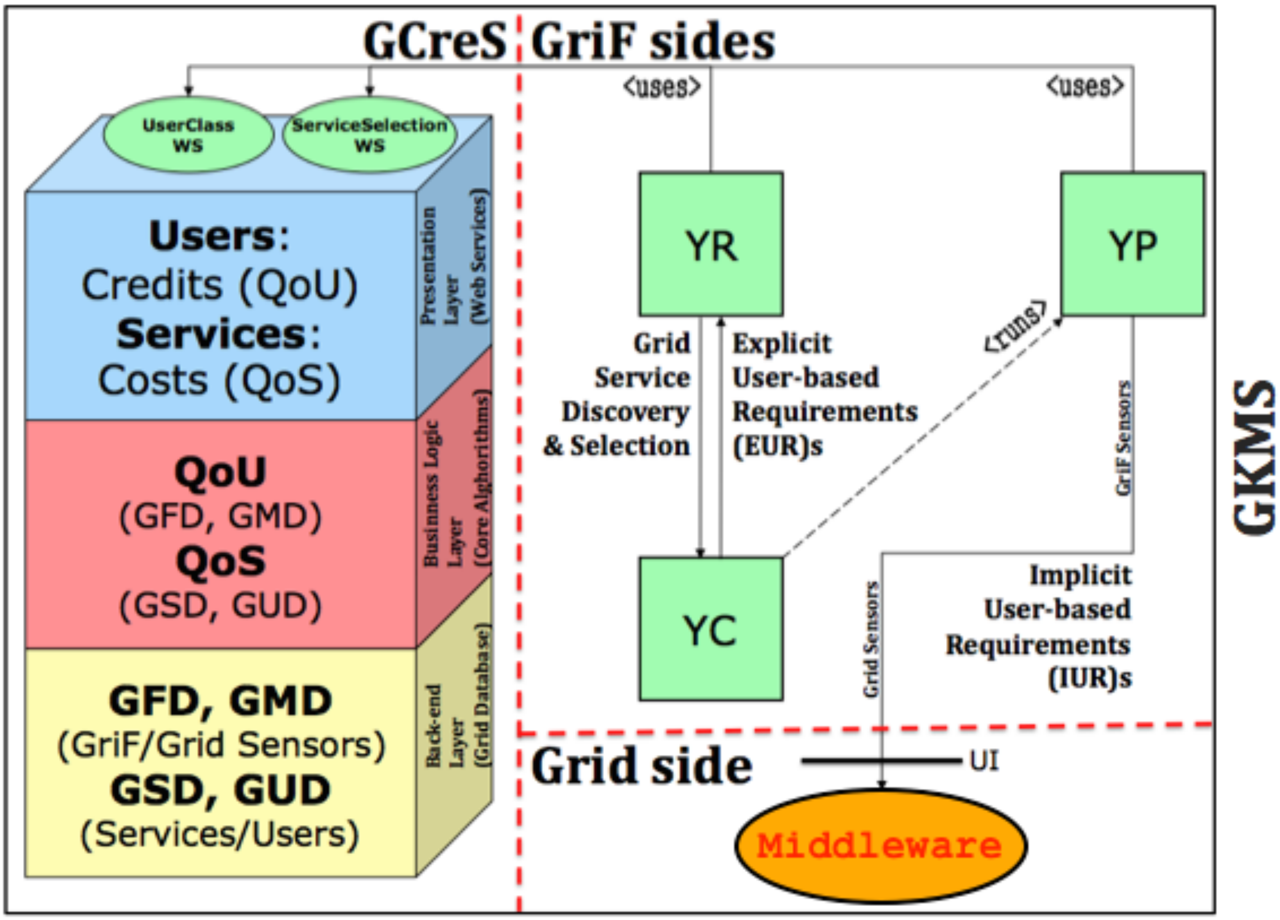

Figure 1 – Sketch of the GCres GriF flowchart to its servers and user interfaces.

Fundamental in this approach is the possibility of formulating a fair evaluation and recognition of the work carried out by the user. This has implied the development of new structured methods to operate on the Grid by evaluating both the QoS achieved in a Grid Service-based producing modality and the QoU provided by the VO/VRC users performing their computations on the Grid. An important advantage of using GriF (and therefore its database) and the new Grid sensors as architectural pillars in the modelling of GCreS is their suitability to support the introduction of QoS and QoU evaluations. GriF offers also to Grid developers new interesting perspectives like those associated with the selection (rather than pure discovery) of Grid Services and, as a consequence, the evaluation of the activities carried out on the Grid. More than that, GriF is open to a direct utilization by the user allowing so far the management of a domain-specific operation logic. This has a clear value for the development of new Workflow Design and Service Orchestration advanced features and for the establishing of collaborative operational modalities in which VO/VRC users and providers collaborate and get their activities rewarded by means of an assignment of Credits.

Basically, in GCreS users are rewarded for the work done on behalf of an organization by being assigned a certain amount of Credits to be redeemed via a preferential utilization of the resources (including the financial ones) of a VO/VRC. Such development, in addition to levering on collaboration, stimulates also a certain extent of competition among the members of a community to produce innovative Grid Services and improve the existing ones. GriF is also a means to stimulate the VO/VRC members to further step up their membership and to contribute to the infrastructure development (Hardware Provider) by conferring to the VO/VRC some of their computing resources (Passive) and by taking care of their deployment (Active) on the Grid. At later stages of the VO/VRC evolution a higher level of membership (Stakeholder) is also foreseen for members highly committed to take care of its global management.

In addition to contributing to the formulation of QoS and QoU evaluations, new higher level ways of managing Grid Services can be adopted:

1) by VO/VRC users: whio will be able to ask for Grid Services by specifying as keywords high-performance capabilities rather than memory size, cpu/wall time and storage capacity;

2) by GriF: that will be able to automatically select the most appropriate low-level capabilities related to the current Grid job (in other words, when a Grid job has to be run, GriF can make use of different system requirements in terms of memory size, cpu/wall time and storage capacity according to the class level of the related VO/VRC user owning that Grid job).

GCreS is also designed to transform the obtained QoS and QoU values into Costs and Credits (both already supported by GriF by making use, respectively, of dedicated fields of the **services** and **vo-users** tables of its database). In particular:

a) Costs to be paid by VRC users for the use of Grid Services on the basis of their QoS;

b) Credits to be awarded to the VRC users on the basis of their QoU.

In case **a** (Costs assigned to Grid Services) the concept of Service Discovery (in which when VO/VRC users search for Grid Services they receive back an unranked list of matching) is better replaced by that of Service Selection (in which the ranking is provided by QoS).

In case **b** (the award of Credits to VO/VRC users) the new concept of QoU is applied. This is, indeed, the case of the above mentioned management of the EChemtest IPCs dealt with more technical details in ref. **[7]**. Such procedure generates incomes that once reached the break-even balance is shared among the community members involved in providing the service using a Prosumer model.

5 – FROM GEMS TO A CLOUD VERSION OF THE QRS COMMUNITY?

The implementation of the components of the scheme illustrated in Fig. 1 allows the full a priori determination of the experimental observable properties of chemical processes (including, therefore, the quantum reactive scattering ones object of the activities of the QRS group). For illustrative purposes, the flow chart of GEMS appeared in the previous Virt&l-comm issue (in which the 2022 QRS workshop was announced) is replicated below in Fig. 2.

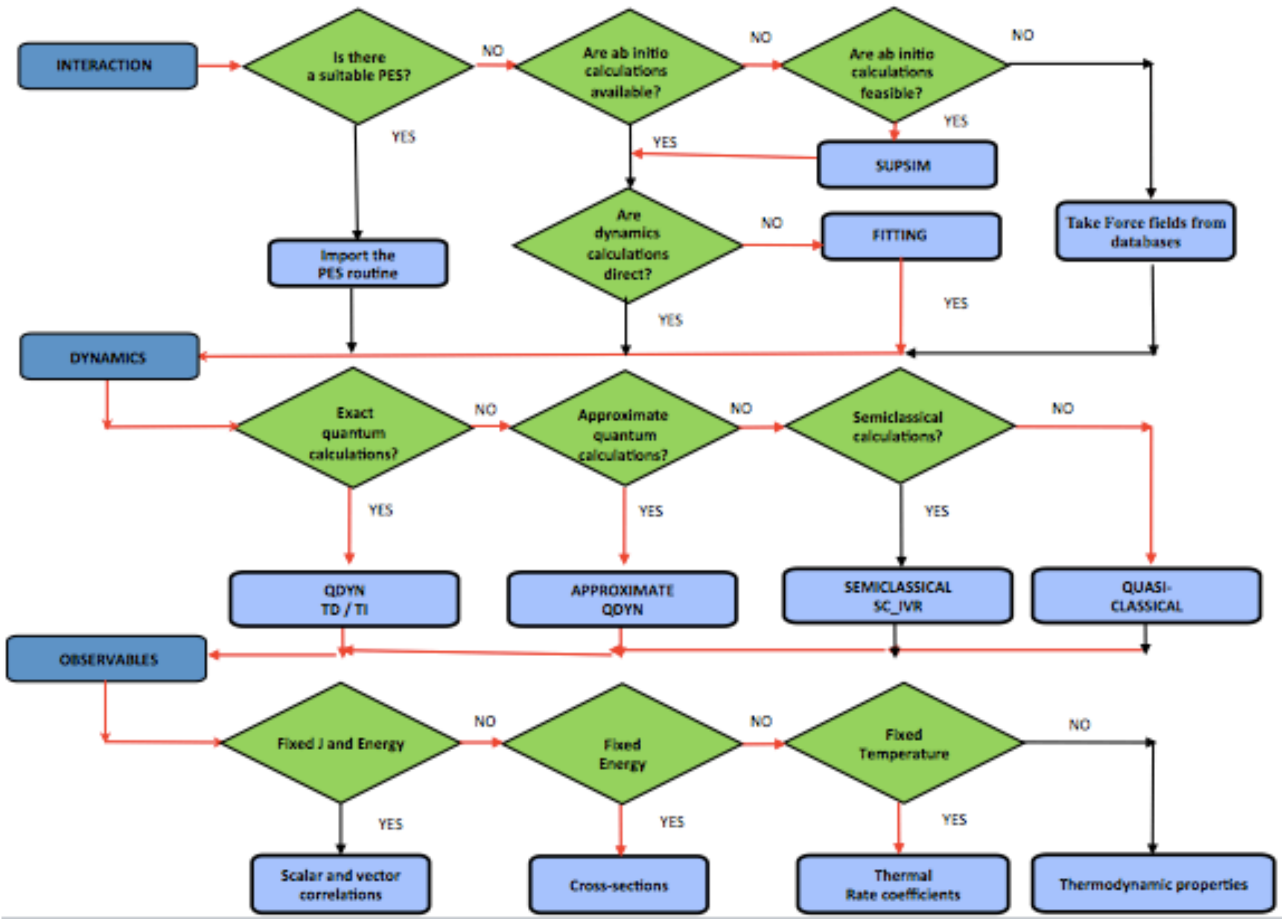

Figure 2 – The general flow chart of GEMS

It is useful to point out here that an extension of that flow chart has been already provided also for the complex quantum treatments of refs. **[8]** and **[9]** that include multi-level meta-workflows reducing the time of manual job definition and output extraction. Such workflows are executed on computing infrastructures and may require large computing and data resources. Scientific workflows hide these infrastructures and the resources needed to run them.

As clearly spelt in its portal, the QRS Workshop is a serial one initiated by David Clary in 1990 at Cambridge, UK. Next editions were held at:

1994 Cambridge, Massachusetts, USA (Yan Sun and Michael Baer),

1995 Nottingham, UK (David Clary and David Manolopoulos),
1997 Telluride, Colorado, USA (Joel Bowman),
1999 Perugia, Italy (Vincenzo Aquilanti and Antonio Laganà),
2001 Pasadena, California, USA (Aron Kuppermann),
2003 San Lorenzo de El Escorial, Spain (Javier Aoiz and Luis Bañares),
2005 Santa Cruz, California, USA (Millard Alexander and Anne McCoy),
2007 Cambridge, UK (Stuart Althorpe),
2009 Dalian, China (Dong-Hui Zhang and Ke-Li Han),
2011 Santa Fe, New Mexico, USA (Hua Guo),
2013 Bordeaux, France (Laurent Bonnet and Pascal Larregaray),
2015 Salamanca, Spain (Octavio Roncero, Tomás González-Lezana, Susana Gómez-Carrasco, Lola González-Sánchez),
2017 Trieste, Italy (Niyazi Bulut, Noelia Faginas Lago, Andrea Lombardi, Federico Palazzetti),
2019 Saitama, Japan (Toshiyuki Takayanagi),
2022 Balatonföldvár, Hungary (György Lendvay, Gabriella Lendvayné Győrik, Ákos Bencsura).
With an average per conference attendance of about 50 people over the past quarter of century, the QRS community has shown to be particularly active in gathering together scientists actively carrying out research in Gas-phase elementary reactions, Theoretical reaction dynamics, Quantum mechanical methods, Semiclassical approaches, Stereodynamics, Polyatomic reaction dynamics, Theory of non-adiabatic processes, Electronic, vibrational and rotational spectroscopy, Potential energy surfaces, Reactions at low and ultra-low temperatures, Gas-surface reactions and Theoretical methods in astrochemistry. Accordingly, the QRS community can act as a key player in promoting and producing substantial advances in the field of experiments, theory and computations on molecular processes by assembling an ad hoc EOSC VRC (as suggested in ref. **10)** aimed at:
a) producing/discovering and downloading MS data for the open collaborative and fully re-usable databases and repositories of FAIR (Findable, Accessible, Inter-operable and Re-usable) validated data enhancing TOP (Transparency and Openness) promoting good scientific practices of accountability, traceability and reproducibility;
b) running/checking/correcting (where possible) or discarding and validating data of service (a) through iterative cycles collaboratively undertaken by the members of the community for the validation of MS data by comparing results obtained by different users when adopting different computational tools;
c) annotating/curating/preserving data of service (b) for a more efficient re-use according to the standards adopted by the EOSC-Pillar initiative. The development of appropriate tools for their efficient re-use will be of invaluable help in order to further enhance research and applications by public institutions and private companies;
d) structuring/annotating/illustrating data produced and teaching schemes involved for an efficient re-use as Learning objects and massive on line courses.
The above steps will act as support actions not only to the research work of the QRS members but also to that of the Materials, Astrochemistry, Combustion, Circular economy, etc. areas by levering on the evaluation of the quality of services provided by the members of the QRS community.

REFERENCES

1] qrs.ttk.hu

2] https://digital-strategy.ec.europa.eu/en/policies/open-science-cloud

3] A. Laganà, A. Riganelli, O. Gervasi, On the structuring of the Computational Chemistry Virtual Organization COMPCHEM, Lecture Notes in Computer Scinece 3980, 665-674 (2006)

4] A. Laganà, Research and Innovation actions. Chemistry, Molecular & Materials Sciences and Technologies Virtual Research Environment (CMMST-VRE), VIRT&L-COMM.6.2014.1

5] S. Rampino, N. Faginas Lago, A. Laganà, F. Huarte-Larrañaga, An extension of the Grid empowered Molecular Simulator GEMS to quantum reactive scattering *J. Comput. Chem*. *33*, 708–714 (2012).

6] https://www.researchgate.net/publication/269398521_A_Grid_Knowledge_Management_System_aimed_at_Virtual_Research_Communities_Sustainability_based_on_Quality_Evaluation

7] C. Manuali, A. Laganà, CHEM-LEARN: the new EChemTest® accounting system, VIRT&L-COMM 22.2021.4

8] A. Laganà, G. Terstyanszky, J. Krüger, Open Molecular Science for the Open Science Cloud http://dx.doi.org/10.1007/978-3-319-62398-6_3, Lecture notes in Computer science, 10406, 29-43 (2017)

9] J. Arshad, A. Hoffmann, S. Gesing, R. Grunzke, J. Krüger, T. Kiss, S. Herres-Pawlis and G. Terstyanszky, Multi-level meta-workflows: new concept for regularly occurring tasks in quantum chemistry, Journal of Cheminformatics, volume 8, Article number: 58 (2016)

10] G. Vitillaro, A. Laganà, Mosex: molecular open science cloud services, VIRT&L-COMM 20,2020,7