EGI-InSPIRE

TOWARDS A CMMST VRC

VIRTUAL TEAM PROJECT REPORT

PROPOSAL FOR THE SETUP OF THE CMMST VRC IN EGI THAT HAS BEEN PREPARED IN THE RELATED EGI-INSPIRE VT PROJECT

https://wiki.egi.eu/wiki/Towards_a_CMMST_VRC

Date:	**18/06/2013**
Document Status:	**Draft**
Dissemination Level:	**Internal**
Document Link:	**https://documents.egi.eu/document/xxxx**



EXECUTIVE SUMMARY

Objective of the document: The present document formulates a proposal for the setup of the Chemistry, Molecular & Materials Science and Technology (CMMST) Virtual Research Community (VRC) in EGI. The proposal has been elaborated as a deliverable of the EGI-InSPIRE Virtual Team (VT) project "Towards a CMMST VRC" promoted by the members of the Virtual Organization (VO) COMPCHEM. The proposed VRC is tailored to the specific nature of the CMMST community that is a typical aggregate of several independent research laboratories gathered together by common scientific interests (not necessarily bound to a common experimental infrastructure) in the CMMST field and aims at establishing a highly collaborative ("synergistic") distributed environment.

Partners of the project: The proposal leverages on the convergence of goals and activities in building the CMMST VRC between COMPCHEM and other EGI VOs (like GAUSSIAN and CHEM.VO.IBERGRID.EU) whose members belong to some University Departments and/or Research Centres. The proposal is supported by some National Grid Infrastructures (like IGI, PL-Grid and IBERGRID), resource providers (like CINECA, CYFRONET, CESCA, CSCS), the MoSGrid and ScalaLife projects and the academic spinoff MASTER-UP. It has also been approved by the Councils of the European Chemistry Thematic Network Association and the Computational Chemistry Division of EUCHEMS.

Goals of the VRC: The specific goals of the VRC are: meet user requirements for an efficient access and use of high throughput and high performance computing resources, enable the composition of higher level complexity applications through the sharing of hardware and software, develop a new collaborative model of carrying out research grounded on a quality evaluation of the work done for the community.

Tasks and tools: Both technical and non-technical aspects related to the building of the VRC by following the evolution of technology, structure and organization of distributed computing are analysed. Related tasks are articulated as follows: assemble a set of inter-linkable applications useful to build higher level of complexity multi-scale computational procedures, exploit the tools offered by EGI to support the activities of a distributed computing community and further develop applications and tools specific of the partner VOs. Such articulation is meant to enable the selection of the provided resources (from personal systems to supercomputers) and services (from number crunching to massive data handling on heterogeneous platforms). The quality of the services provided and of the resources used will be utilized also to build a metrics on which grounding the rewarding of the work done for the community through a credit acquisition/redemption system. This will ground a sustainable Grid economy (generalizable to other communities) and affect how the CMMST VRC carries out research and impacts society.

Perspective research and developments: the enhancement of the European Grid computing platform user friendliness and interoperability will be further pursued by equipping the CMMST VRC with appropriate portals and tools, with de facto standards for large data formats and with instruments allowing the bridging of HTC-HPC (thanks also to the activities of the MoSGrid (DE) and XSEDE (US) projects). On this, the building of innovative solutions in molecular and materials research, education and technology training is planned by involving also European associations and SME networks (via the activities of the MASTER-UP spinoff).



I. DOCUMENT LOG

Version	Date	Comment	Author/Partner
0.0	18/06/2013	Draft0	A. Costantini
0.1	19/06/2013	Revision	A. Costantini
1.0	26/06/2013	Draft1	A. Laganà
1.1	30/06/2013	Revision	A. Costantini
1.2	01/07/2013	Revision	D. Cesini
1.3	05/07/2013	Revision	M. Ceotto
1.4	12/07/2013	Revision	A. Costantini
2.0	30/07/2013	Draft2	A. Costantini
2.1	02/08/2013	Revision	A. Costantini
3.0	08/08/2013	Draft3	A. Laganà
3.1	19/08/2013	Revision	A. Costantini
3.2	27/08/2013	Revision	A. Laganà
3.3	27/08/2013	Revision	A. Costantini
3.4	27/08/2013	Revision	A. Costantini
3.5	29/08/2013	Revision	S. Farantos
3.6	09/09/2013	Revision	G. Sipos
3.7	26/09/2013	Revision	A. Costantini
4.0	30/09/2013	Draft4	A. Laganà
4.1	07/10/2013	Revision	A. Costantini
4.2	07/10/2013	Revision	A. Costantini
4.3	11/10/2013	Revision	A. Laganà
4.4	21/10/2013	Revision	A. Costantini
4.5	22/10/2013	Revision	S. Herres-Pawlis
5.0	22/10/2013	Revision	A. Costantini, A. Laganà
5.1	02/11/2013	Revision	A. Costantini, A. Laganà
5.2	03/11/2013	Revision	A. Costantini, A. Laganà
5.3	8/12/2013	Revision	S. Herres-Pawlis, R. Apostolov, A Costantini, A Laganà
5.4	10/2/2014	Copy for circulation	A. Laganà

II. ACKNOWLEDGMENT

None

III. APPLICATION AREA

This document is a public report produced by the members of the "Towards a CMMST VT" EGI Virtual Team project, run under the EGI-InSPIRE NA2 virtual team framework. Further information is available at https://wiki.egi.eu/wiki/Towards_a_CMMST_VRC.

IV. TERMINOLOGY

A complete project glossary is provided at the following page: http://www.egi.eu/about/glossary/.

TABLE OF CONTENTS

1 INTRODUCTION

The VRC: Virtual Research Communities (VRC)s are groups of like-minded individuals and clusters of researchers organised by discipline or computational model. European Grid Infrastructure (EGI, http://www.egi.eu/) VRCs (http://www.egi.eu/community/vrcs/) typically have an established presence in their field and represent well-defined scientific research communities. A VRC can establish a support relationship, formalised through a Memorandum of Understanding (MoU), with the EGI collaboration. Multi-national scientific communities can draw many benefits from having a VRC partnership with EGI. For example, they can benefit from resources and support available within the National Grid Infrastructures (NGI)s, the main stakeholders of EGI.eu. They can also benefit from the workshops and forums organised by EGI, receive support on resolving specific technical issues with EGI services and become involved in the user-focussed evolution of the production infrastructure of EGI.
Proposer and partners: For the above mentioned reason, the Virtual Organization (VO) COMPCHEM (https://www3.compchem.unipg.it/compchem/), whose members are largely represented in the Virtual Team (VT) list given in Appendix B, together with two other EGI VOs (GAUSSIAN, https://indico.cern.ch/event/1504/session/131/contribution/216/material/slides/0.pdf and CHEM.VO.IBERGRID.EU, http://ibergrid.lip.pt/USP), the support, as well, of some National Grid Infrastructures (like IGI (http://www.italiangrid.it/), PL-Grid (http://www.plgrid.pl/) and IBERGRID (http://ibergrid.lip.pt/)), projects (like MoSGrid (http://www.sci-bus.eu/mosgrid) and ScalaLife (http://www.scalalife.eu/)), resource providers (like CINECA (ww.cineca.it), CYFRONET (http://www.cyfronet.krakow.pl/), BSC-CNS (http://www.bsc.es/), CSC (http://www.csc.fi/english)). Moreover, the proposed VRC relies also on strong interactions with the European Chemistry Thematic Network (ECTN, Association, the Division of Computational Chemistry (DCC, http://www.euchems.eu/divisions/computational-chemistry.html) of EUCHEMS and the academic spinoff MASTER-UP (www.master-up.it/).

The articulation of the document: Accordingly, the document is articulated in the following sections:

- Section 2: Outcomes of the milestones
- Section 3: The proposal of the CMMST-VRC;
- Section 4: Platforms, tools and services offered by EGI to the VRC;
- Section 5: Technologies and services offered by the CMMST VRC through EGI;
- Section 6: Management and sustainability;
- Section 7: Perspective research and developments.

2 OUTCOMES OF THE MILESTONES

The objectives addressed by the VT project [1] were:

- develop a plan aimed at assembling a VRC out of the already existing CMMST oriented EGI Virtual Organizations;
- exploit related applications, tools as well other resources and services that NGIs and EGI projects provide;
- identify tools, services and resources that the VRC needs to develop or bring into EGI in order to operate as a sustainable entity for the CMMST scientific community;

- elaborate a proposal for the technical, organisational and funding aspects.

The above objectives were articulated in the following M1-M5 milestones' outcomes.

2.1 M1 outcomes

M1 outcomes are those resulting from the preliminary interactions of the COMPCHEM management (A. Laganà, A. Costantini) with the VO members and with the coordinator of the GAUSSIAN VO coordinator (M. Sterzel) in close contact with the EGI User Community Support Team (G. Sipos, R. McLennan) taking care of the coordination of the technical outreach and of the collaboration with VRCs and research infrastructures. These outcomes were focused on the preparation of the Investigation Planning Document (IPD) containing a list of technical and non-technical topics (like tools used, access to computing resources, structure of the VRC, attraction and training of new users, etc…) investigated by the VT. The IPD presents an analysis of the consistency of the available application patrimony and indications on how to attract more users to CMMST (in addition to the already existing ones of the partner VOs) with a particular attention paid to the involvement of the members of the Division of Computational Chemistry (DCC) of EUCHEMS and to an extension of the activities to research based education carried out by the virtual education working group of the European Chemistry Thematic Network Association (ECTNA). Each topic was allocated either to an individual or to a team leader from the VT.

2.2 M2-M4 outcomes

M2-M4 outcomes are those resulting from the discussions held among the VO partners of the VT and in particular from the contributions of the following Institutions: University Departmental structures (like Chemistry of the Wroclaw Technology University (PL), Chemistry of the Texas Tech University (US), Chemistry of the University of Perugia (IT), Physical Chemistry of the University of the Basque Country at Vitoria (ES), Physical Chemistry of the University of Barcelona (ES), Physics of the University of Oklahoma Norman (US)) and Research Centres (like FORTH (GR), Research Centre for Natural Sciences of the Academy of Science (HU), GC3: Grid Computing Competence Centre (CH)). It has involved also the representatives of some National Grid Infrastructures (like IGI, PL-Grid and IBERGRID) and of some resource providers (like CINECA, CYFRONET, BSC-CNS, CSC). Such outcomes constitute the basic guidelines for the assemblage of the proposed VRC and provide the ground for implementing the various steps of the proposal like the singling out of the existing EGI technologies, resources and services useful for the development of the CMMST VRC, the formulation of the requirements set by the community, the implementation on the Grid of the programs and tools to be offered by the VRC through EGI. In this a particular attention has been paid to the need of going beyond pure HTC computing and of filling the lack of standard data formats in molecular sciences.

2.3 M5 outcomes

The M5 outcomes are those resulting from the work performed by the VT partners and their further interaction with projects (like MoSGrid) and academic spinoffs (like MASTER-UP). M5 outcomes, detailed in the next sections of the present document, are the final product of the VT project. They complete the previously described outcomes and are presented in a form suitable for implementation and dissemination. To this end, they single out the different levels of involvement of the participating Institutions. The strong collaborative frame of the synergistic model adopted by the CMMST VRC

implies, in fact, that some researchers (or research teams) in addition to their user activities take levels of responsibility in sharing and managing resources, tools and applications to ensure sustainability and guarantee efficiency. Other outcomes of M5 are the requirements listed below (resulting from an in depth analysis of the optimal working conditions of the proposed community) are:

a. Drive a combined selection of the available hardware and software resources among those considered to be suitable for running the proposed applications;
b. Adopt allocation mechanisms for computing resources in addition to the "grant" and "paid" modalities;
c. Develop a quality based credit economy with credits being awarded in return for activities carried out on behalf of the community either as work performed within VRC projects or as funds procurement and are redeemable in terms of a preferential use of computing and/or financial resources.

3 THE PROPOSAL OF THE CMMST-VRC

In this section the most popular models presently used for allocating computing resources to the members of the CMMST community (and their evolution towards the Grid federated distributed one) are analysed in subsection 3.1 and 3.2, respectively. The collaborative synergistic model, the main feature of the present project, that leverages on the possibility of chaining different applications in workflows (and workflows of workflows as well) is described in subsection 3.3. The evaluation of the quality of the work performed on behalf of the community and of its rewarding by means of a credit award/redeem system on which the synergistic model is founded are given in subsection 3.4.

3.1 Current scenario of computing resources allocation and usage

The evolution of computing technologies has led to the growth of both remotely accessible high-performance platforms (High Performance Computing, HPC) and local (departments, research groups, institutions, etc.) clustered computational platforms and networks. These two types of platforms adopt different models of resources allocation:

- the first (the centralized grant model) assigns the shared resources in a dedicated fashion after an ex ante evaluation (based on criteria set by the facility management) of a proposal referring to a specific call for user projects;
- the second (the local opportunistic model) assigns the available resources so as to match the requests of the registered users as they come without making any selection

though often complemented with some mitigation criteria.

The grant model is in general adopted by large scale supercomputing facilities (as well as by medium-large size computer centres) in which specialized staff is devoted to the maintenance of HW and SW. The local opportunistic model, instead, is in general adopted either by research groups or departments running their machines for internal use (with the help of some permanent or temporary technical staff).

In large scale computing facilities (like the computer centres networked in PRACE, http://www.prace-ri.eu/) the users can rely on computational resources, software and skills evolving according to both the current technological development and time assignments (the already mentioned grants) born out of a centralized vision of computing needs and strategies. In such model, however, the users develop a rather passive attitude (with respect to the hardware, software and networking evolution) and have substantial difficulties to give continuity to their work when awarded grants expire. The model has,

however, the advantage of stimulating the users to adapt continuously their codes to the most up to date computing paradigms and tools.
In local computing facilities, instead, the users (which are either responsible or co-responsible for managing the available resources and related environments), pay usually the price of a slower evolution of available computing technologies. Despite that, thanks to the fact that they play a more active role in managing the platform, users are able to better tune and customize the machine to fit their needs. Moreover, they can deliver/utilize the services in addition to getting access to computer resources at their wish (compatibly with availability and related queuing policies).

3.2 The federated opportunistic model of the Grid

Although widely adopted, both the centralized and the local opportunistic models have become in the recent past inadequate to meet alone the evolving requirements of the scientific communities. As an alternative, an innovative opportunistic model, based on the allocation of the computational tasks on federated distributed resources, was developed within a series of European projects and collaborations (DATATAG (http://datatag.web.cern.ch/datatag/), EGEE-I-II-III (http://www.egee.eu), WLCG (http://wlcg.web.cern.ch/), etc.) mainly under the pressure of the research needs of High Energy Physics. Such model, often called High Throughput Computing (HTC) or Grid Computing (GC) exploits the aggregated compute power of a large number of geographically dispersed CPUs connected over the public network through the use of appropriate middleware and tools. GC has enabled the concurrent execution of distributed programs on a large number (over hundreds of thousands) of processors mainly for applications made of decoupled or loosely coupled tasks and has allowed the production of a large volume of results in reduced time intervals.

After the above mentioned first impulse given by the High Energy Physics community, the use of GC was extended to other disciplines. After all, for researchers not belonging to a single Institution, or not being tightly bound to the management of a shared experimental infrastructure, the opportunistic model of the Grid, bears already the added value of being an invaluable complement to the use of local resources and supercomputers of the large scale computing facilities. In this model, in fact, the generic users can access, among the already implemented applications, those of their interest or can port on GC their applications for personal use (though this requires some additional skills and support). However, the use of GC leads naturally to the sharing of HW and SW and to the development of cooperative endeavours. Within the GC model, in fact, users not only can get an on demand allocation of the available computing time but they can also set common requirements, share in a bottom up fashion data and programs, complement each other expertise, establish virtual communities, etc..

As a matter of fact, within the current EGI-Inspire European project (https://www.egi.eu/about/egi-inspire/) the number of participating communities, of networked resources, software and skills as well as of inter-disciplinary collaborations has continuously increased. Moreover, in order to meet the peculiarities of the different research areas and strengthen their role, the users of the Grid have been clustered in VOs (http://www.egi.eu/community/vos/) that are now encouraged to organize themselves into VRCs (http://www.egi.eu/community/vrcs) and share tools and applications.

3.3 The proposed synergistic Grid model of the CMMST VRC

In order to cope with the multi-scale nature of most of the CMMST simulations which aim at reproducing the observable physical properties of realistic systems by starting from first principles and at integrating different expertise and treatments, the federated opportunistic model will be upgraded to the synergistic one [2] allowing an a la carte combination of the applications made available to the members of the community by EGI, CMMST and other providers (including external ones). The mentioned synergistic computing model, in fact, meets the dreams of the most recent generations of computational scientists because it allows to run highly complex applications not only by having flat access to highly heterogeneous platforms but also, for example, by exploiting the advanced features of some workflow structured computing procedures like the Grid Empowered Molecular Simulator (GEMS) [3] that has been used to the end of carrying out also virtual experiments [4].

In other words, the proposed synergistic model encompasses much more than the simple feature of aggregating a large amount of computing resources and users for massive distributed computing. It targets, in fact, the more radical objective of changing the model of using and managing research oriented computing resources by fostering:

- a synergy among complementary types of programs (or sections of programs) and expertise (including the monitoring of experiments and the access to data bases);
- a service oriented organization with a rewarded active participation of the users in different activities of the community [5];
- a coupled distributed and parallel HTC and HPC computational approach.

It is important here to emphasize that, the synergistic model grounds the rewarding of the users' contribution to the community activities on the quality evaluation of the services provided. This enhances the community sustainability by calling both for the development of tools facilitating collaborative activities and for the introduction of a metrics suited for ranking services. Within the synergistic model users can, in fact, provide and maintain stable versions of their own programs and applications software properly structured in terms of (input and output) data formats so as to be chained in GEMS and offered as services (in return for credits as we shall detail later) to other users as a kind of shared library of applications. To this end COMPCHEM is going to join its efforts with those of other partners either belonging to the CMMST community (like GAUSSIAN, MoSGrid, Scalalife and CHEM.VO.IBERGRID.EU) or external to it like the Division of Computational Chemistry of EUCHEMS (http://www.euchems.eu/divisions/computational-chemistry.html) and the VEC standing committee of the ECTN (http://ectn-assoc.cpe.fr/) Association including mainly operators of the research based higher education [6].

In order to better achieve a high level of cooperation, quarterly teleconferences for planning and managing the activities of the CMMST community will be held and an annual (or biennial) specific face-to-face meeting will be held during one of the main EGI Grid events. Such meetings will also focus on technical problems of the VRC including operations, services, support to user, communication and dissemination.

This grounds also the VRC purpose of competing for acquiring, as a general policy, computing resources from resource providers for the community (to be added to the ones available to the VRC members from other sources). These resources will be used by the community for supporting fundamental activities (like basic research and algorithms development but also production runs), sustaining collaborative projects, pursuing special innovative targets, etc. whose assignments to the members of the community will be regulated by the credit system.

3.4 Membership levels and credit system

As already mentioned, the rewarding of the users' contribution to the community activities based on quality evaluation of the services provided is a key added value of the synergistic model proposed for the CMMST VRC. It fosters, in fact, the sharing of expertise and products among the members of the community (especially to the end of tackling multi-scale problems of higher complexity for which individual competences are insufficient) in a trustable objective way once the related metrics of monitoring the activities, evaluating their quality and assigning related credits is agreed by the community.
This stimulates some members of the community to become service providers for the community. Service providers can be of different types. They can offer hardware and software, support the design and development of new algorithms and applications, assist and help the users in running existing packages, produce and validate new data, design and develop new Grid approaches, disseminate community activities, etc.

Obviously, CMMST researchers can choose to join the community just as passive users (the first sublevel of level 1 of the scheme shown below and commented in Ref. [5]) to the end of running the codes made available by the VRC for demo purpose. The reason for having this entry level (usually meant to be temporary) is to offer the opportunity to the newcomers to check their real willingness to operate on complex Distributed Computing Infrastructure (DCI) platforms. At this level, in fact, after experiencing the use of the Grid for running applications made available by someone else, a user may wish to further exploit the advantage of using the available resources to run his/her own software. To this end, he/she has to acquire the necessary skills and competences (through, for example, the community specific training activities) and step up to the level of active user. In case, however, the CMMST member wants to remain passive user, after the expiration of the trial period, he/she would have to contribute to the VRC sustainability using different means (like for example becoming a paying user). On the contrary, the involvement gained by being active user may encourage the VRC members to escalate to the level of service provider. A type of service central to the synergistic model is that of making available for a shared usage by the other members of the VRC the software ported on the Grid for personal use (software provision).

SCHEME OF THE COMPCHEM USER LEVELS

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.

The involvement reached as software provider may further motivate a VRC member to upgrade its participation to a higher level of service conditions (like the offer of a validated stable version of the code as opposed to the on-development one or the adoption of proper, even if only de facto, standard data formats).

Software providers may also guarantee additional services like software upgrade, maintenance, user support and the assemblage of all the necessary GUIs (Grid User Interfaces). Such levels of service will be used to define the quality of the service provided and to assign the related credits.

Another important type of service is the one related to the commitment to confer to the Grid resource provider additional hardware (including special devices needed for some experiments) after a negotiation with the VRC Management about the relevance of such a commitment to the strategic choices of the community. Accordingly, both software and hardware providers may escalate up their commitment to the provision of basic services, tools and related support.

Needless to say that the climbing up through all these levels of commitment will take place in steps during a certain time interval of time (with only properly qualified and motivated members of the VRC being able and wishing to do so in order to achieve the status of "VRC stakeholder" (top user level).

The belonging to a given user level is periodically revised against the number of credits acquired. Accordingly, a given member of the community might end up by belonging, at the same tine, to different levels depending on the product or the service considered (for example, the status of stakeholder does not exclude that of paying customer or paid supplier for certain services). This is particularly useful when the VRC wins a bid or becomes a funded project holder.

The mechanism through which credits are assigned and redeemed are established and regulated by the governing body of the VRC which will make use of the two already mentioned tools *GriF* [7] and *GCres* [8] which have been designed by COMPCHEM and will be further developed by the VRC as will be discussed in detail later on.

4 EGI PLATFORMS, SERVICES AND SUPPORT

In this section we list and summarily describe the platforms, the services and the support offered to the CMMST VRC by the NGIs of EGI. The section starts in subsection 4.1 with the description of two EGI platforms: the EGI HTC platform and the Federated Cloud Computing (FCC) platform. Then it continues in subsection 4.2 by describing the basic operational and other general services. All these services can be used by the CMMST VRC members to the end of performing compute and/or data intensive elaborations on the resources of the EGI NGIs.

Support activities (including dissemination) are illustrated in subsections 4.3. Related software tools and human services can be used by the CMMST VRC to coordinate the community, to facilitate communication inside the VRC, to assure continuous exchange of relevant information between the VRC and the rest of EGI, and to reach new members.

Platforms, tools and services described in this section are all operated by the NGIs. The software tools have a web interface that is accessible through some graphical portal (URLs are indicated in the text). The two platforms provide rather low level access mechanisms. These can be extended and made more user friendly by adopting science gateways and portals often built out of flexible frameworks. Some of these tools, such as portals and application gateways, have additional interfaces in the form of web gadgets and/or programming APIs which facilitate their strong integration with the domain specific services of the community,.

4.1 EGI Platforms

4.1.1 EGI high-throughput platform

EGI high-throughput platform[1] is a global high-throughput data analysis infrastructure, linking hundreds of independent research institutes, universities and organisations delivering top quality computing resources. As of January 2014, EGI offers more than 400,000 CPU cores of installed computing capacity interoperating through the gLite (http://glite.web.cern.ch/glite/) middleware and supporting about 1.8 million computing jobs per day. Within this platform resources are allocated to 'Virtual Organisations' (VOs). Each VO represents a scientific community and the resources these communities have access to from the production infrastructure. The computational chemistry community operates in EGI through 5 VOs[2] and several regional VOs (like the VO of Central Europe). Intended members of the CMMST VRC can either join the five existing computational chemistry VOs or can establish new VOs as well. All VO-related operations are available through the EGI Operations Portal[3].

4.1.2 EGI Federated Cloud platform

EGI Federated Cloud[4] is a single, standards-based, open system to federate academic clouds from multiple providers, offering scalable computing resources with increased flexibility. It provides advanced compute capabilities for research, virtualised resources to run any chosen environment, cloud storage for easier sharing of data, and a number of support services to ensure applications run as efficiently as possible. EGI Federated Cloud can be used for running cloud-based applications and services needed by CMMST VRC researchers. The EGI Federated Cloud platform currently operates in testbed mode, and is open for early adopter scientific use cases. Changes and improvements in the testbed is frequent, and these occasionally cause disruption in its services. The testbed will evolve into a production system after 2013.

4.2 Services

4.2.1 Operation services

EGI Operation services[5] are technologies, processes and people required to federate computing and data resources from resource providers into EGI. They provide a cost-efficient framework to manage operations within a federated environment, while retaining responsibility for local infrastructure. EGI Operations services can be used by the CMMST VRC to federate clusters, storage and desktop systems into EGI, and share these capacities in a secure manner with VRC members.

4.2.2 Applications database

The EGI Application Database (AppDB)[6] is a centralised service that stores and provides information about the rich variety of software that has been integrated with the EGI production infrastructure. AppDB gives visibility to and facilitates the re-use of scientific software and associated tools by providing a platform for developers to publish their software in the repository part and/or to

[1] http://www.egi.eu/solutions/htc/index.html
[2] Compchem (international), Gaussian (international), chem.vo.ibergrid.eu (Spanish), Trigrida (Turkish), Trigridb (Turkish). The 5 VOs include around 40.000 CPU cores and 3PB storage in total.
[3] http://operations-portal.egi.eu/vo
[4] http://www.egi.eu/solutions/fed-cloud/index.html
[5] http://www.egi.eu/solutions/fed-ops/index.html
[6] http://appdb.egi.eu

provide user-friendly description of their software and related access conditions. AppDB can be used to publish information about scientific applications, tools, science gateways, workflow systems and middleware components. The CMMST VRC can use existing entries from AppDB, and can publish their own software in the database, making these available not only to the CMMST members, but also to research communities around the world. AppDB has a 'web gadget'[7] that enables the integration of lists of its relevant software into any website. The list presented in such a gadget is clickable, so those who wish to get further information about items in the list are redirected to the particular entry. The CMMST VRC could for example use the gadget to list software in the VRC website from AppDB that are registered from the 'computational chemistry' domain in the database. AppDB can be used, therefore, as a central 'shop window' by the CMMST VRC to search for and browse applications, tools, science gateways, workflows and middleware software from EGI.

4.2.3 Training Marketplace

The EGI Training Marketplace[8] is an online registry of training events, training materials, training resources (e.g. computing sites for training), webinar events and university courses that are relevant to EGI user communities. The service allows users to browse and search for tailored content, and institutions, projects and VRCs to advertise their training offering. The Training Marketplace provides rating and commenting facilities for the registered items, and web gadgets for integration with third party websites. It can be used to publish events, training material, online courses that the CMMST VRC develops for its scientific community, and share this information within the VRC (through the Training Marketplace web gadget7) at CMMST website, or with the whole EGI community (through the Training Marketplace web portal).

4.2.4 Document Database

EGI Document Database[9] (DocDB) is a powerful and flexible collaborative document server. It was originally developed for use by the BTeV collaboration at Fermilab and is now used by twenty or more other experiments, Fermilab divisions, organizations, national laboratories, and companies (a partial list). DocDB is well suited for managing and sharing documents (not just official publications) among groups of up to several hundred people. It can be used to publish and manage (version control, updates) internal and public documents that the CMMST VRC produces with the read and write permissions for documents to be controlled separately.

4.2.5 Requirement Tracker

EGI Requirements Tracker[10] (RT) is used by European Grid Infrastructure collaboration to collect, store, monitor and provide solutions to requirements and recommendations communicated by scientific communities. It is powerful and flexible tool with a web frontend and the ability to accept requests submitted by email. RT can be used to submit requirements and recommendations allowing to keep track of their status, and to receive alerts about progress and solutions. Submitted requirements can be also tracked either via custom RT dashboards or custom RT web gadgets which can, for example, display only the CMMST VRC requirements. Such a dashboard and/or web gadget can be created and embedded into CMMST portals and gateways.

[7] http://go.egi.eu/gadgets
[8] http://training.egi.eu
[9] http://documents.egi.eu
[10] http://go.egi.eu/requirements

4.2.6 Webex tele and video conference system

Webex[11] is a high level communication platform which provides on-demand collaboration, online meeting facilities, web conferencing and video conferencing services. EGI.eu, the coordinator of the EGI collaboration has a subscription to Webex. The CMMST VRC can use this EGI Webex account to organise teleconferences for VRC coordination calls, and for interaction calls with EGI. The Virtual Team project that authored this document was also using this system for its teleconferences.

4.2.7 EGI SSO and GrIDP identity Federation

EGI Single Sign On (SSO) system provides the federated authentication and authorization mechanisms within EGI services. With an EGI SSO account one can access the protected features of EGI services, for example to register applications in the Applications Database, to add new documents to the Documents Database, to register new training events or materials into the Training Marketplace and so on. Anyone can obtain an EGI SSO account at http://www.egi.eu/sso. EGI does not provide identity vetting for the SSO, i.e. the real identity of the requester is not checked, the accounts are issued based on the data the requester provides.

The EGI SSO system participates as an Identity Provider in the Grid Identity Pool (GrIDP) identity federation[12]. GrIDP is jointly managed and operated by GARR and INFN Catania from Italy, supporting cross-institutional e-Infrastructure services and providing federated authentication to its members. EGI SSO account holders through GrIDP can access high-level services, such as web portals and science gateways, in an easy yet secure and confidential way, using the credentials provided by the EGI SSO system.

The CMMST VRC can use the EGI SSO system to get controlled access to the protected services of the EGI Tools, and can use the GrIDP federation to access services with EGI SSO accounts. In addition it can offer services that are accessible with accounts issued by the Identity Providers of GrIDP[13].

4.3 Community support

4.3.1 Software components and services

AppDB, the EGI Applications database (See section 4.1) stores information about software components, packages, services that are integrated with the EGI production infrastructure. These include scientific applications, science gateways, workflows, tools, middleware components. Many of the software that are registered in the database are relevant to the CMMST VRC. These are scientific code (e.g. GAMESS, GAUSSIAN, NAMD, MOPAC, etc) that are already accessible on some EGI Virtual Organisation, developer tools that members of the VRC could use to integrate their own simulation code with EGI, and middleware, workflow and other components that can extend and improve the core capabilities of the CMMST software.

4.3.2 Helpdesk

EGI.eu coordinates the distributed user and operations support activities provided by the individual NGIs and resource centres in order to ensure that operational issues are properly handled at both the Resource Centre and NGI level. This includes a centralised helpdesk ticketing system with dedicated oversight and follow-up. The EGI Helpdesk[14] can be used for reporting problems and bugs

[11] http://www.webex.com
[12] http://gridp.garr.it/
[13] http://gridp.garr.it/identity-providers.html
[14] http://ggus.org

in the existing setups. The tickets are routed to NGIs, project specific support teams, or other types of experts able to resolve the issue. The CMMST VRC can use the helpdesk to submit tickets for example about issues affecting their applications and Virtual Organisations.

4.3.3 Community representation

The EGI User Community Board (UCB)[15] includes representatives of all VRCs (as well as from other projects collaborating with EGI where necessary) in addition to the service providers within EGI. UCB develops the strategy and defines the priorities for the human and technical user community support services provided by EGI.

The User Community Board can be used to provide feedback and discuss the issues within EGI relating to the User Communities, such as high priority requirements from the CMMST VRC. After established, the CMMST VRC will delegate representative(s) to the UCB.

4.3.4 NGI user support teams

Each country member of the EGI federation of resource providers has a dedicated user support team[16] that can help existing and prospective users to access the e-Infrastructure resources and software services. The contact points of every NGI User Support Team is listed on the EGI Website. Members of CMMST VRC can ask for user support within the country they are present or where the resources they use are located.

4.3.5 Outreach via the NGI International Liaisons

Every NGI of EGI has an 'NGI International Liaison' (NILs) delegated to EGI. NILs[17] have a key role in the EGI community: They link the national activities from the 'non-technical areas' into the European landscape coordinated by EGI.eu. NILs work with teams in EGI.eu on integrating outreach, marketing, communication, training, new community engagement with the rest of the community. With the NIL acting as a spearhead, each country is able to demonstrate their added value at a European level and to share their skills with the whole community. CMMST VRC can use NILs network to outreach the achievements and communicate with a local communities, researchers to expand the Virtual Research Community of CMMST in specific countries, under targeted actions locally supervised by the NIL of the country.

4.3.6 Marketing and outreach services

EGI.eu coordinates the marketing activities[18] of EGI on behalf of the NGIs and partner projects. The aim is to publicise the work of EGI and its user communities to target audiences, including: journalists, general public, resource providers, collaborating projects and decision makers. The EGI outreach services[19] are set up to engage with new users and user communities and to continuously interact with existing ones. The goal is to ensure that the largest number of researchers is able to take advantage of the benefits that EGI offers. CMMST VRC can benefit from the EGI marketing and outreach services provided by EGI such as: publishing news at project website, EGI's publications, the EGI newsfeed, social media channels, representing the community at key events, benefit of the developed EGI Champions network, publication of materials tailored to different interests, and the organisation of two yearly flagship events as well as several topical workshops. Use the strategic

[15] https://www.egi.eu/about/policy/groups/User_Community_Board_UCB.html)
[16] https://www.egi.eu/services/support_contacts/index.html
[17] https://www.egi.eu/community/ngis/NILs.html
[18] https://www.egi.eu/services/catalogue/marketing.html
[19] https://www.egi.eu/services/catalogue/outreach.html

relationships developed by EGI with various organisations, projects, user communities and resource and technology providers through collaboration agreements.

5 TECHNOLOGIES AND SERVICES OFFERED BY THE CMMST-VRC THROUGH EGI

In this section a prospect of technical and non technical aspects related to the applications and services offered by the CMMST VRC to its users is given. Applications and services are the shared patrimony of the VRC and are also the means through which the VRC enables its members to develop higher complexity activities.

In subsections 5.1, 5.2 and 5.3 a list of computational chemistry codes already implemented (or in the process of being implemented) on the Grid by the members of the involved VOs and related Contact Points (CP) is given. Leveraging on the competences of the CMMST members and on the technological support of EGI these codes (and possibly other codes among those given in the list of chemistry and solid state physics software given in Ref. [9]) will be adapted to become interoperable components of multi-scale simulators addressing real-like problems concerning research, innovation and development in different fields ranging from chemical engineering to biochemistry, chemometrics, omic-sciences, forensic chemistry, medicinal chemistry, food chemistry, energy production and storage, new materials, space technologies, etc. The reference simulator for the CMMST VRC will be the already mentioned Grid Empowered Molecular Simulator, GEMS [3] already relying on a set of programs devoted to

- the ab initio calculation of the electronic structure of molecular systems and the assemblage of ab initio based potential energy surfaces (INTERACTION module),
- the integration of quantum and/or classical equations of motion of molecular systems and the evaluation of their basic scattering properties and measurable quantities (DYNAMICS module),
- the concerted handling of additional statistical and higher scale treatments pertaining to realistic simulations (OBSERVABLES).

A list of the applications relevant to CMMST already ported (tentatively) on the GEMS Grid environment and receiving dedicated user support is shown in Table 1. Such applications are already registered in the EGI AppDB and related CPs are provided.

5.1 Ab initio electronic structure applications

We list here (together with a short description, references and CPs) some ab initio electronic structure packages (Hartree-Fock (HF), post Hartree-Fock, Density Functional Theory (DFT)), either open source or commercial, based on the Born Oppenheimer approximation, such as those given in Ref. [9], suited for insertion into the INTERACTION module of GEMS. Most of them are large packages and often contain several separate programs developed over many years.

GAMESS-US [10] CP (K Baldridge, S. Maffioletti)

GAMESS-US is an ab initio electronic structure molecular quantum chemistry package that calculates potential energy values for moderately large molecular systems using direct and parallel techniques on appropriate hardware.

GAUSSIAN [11] CP (GAUSSIAN VO – M. Sterzel)
GAUSSIAN is a quantum chemistry package to calculate potential energy values for moderately large molecular systems performed using direct techniques, or in parallel on appropriate hardware.

NB-MCTDH [12] CP (D. Skouteris)
NB-MCTDH is a multiconfigurational time-dependent Hartree (MCTDH) program for calculating bound states of a generalized N-Body system (including Non Born Oppenheimer cases)

NWChem [13] CP (S. Herres-Pawlis)
NWChem is a package (ported on the UNICORE middleware (http://www.unicore.eu/)) providing its users with computational chemistry tools that are scalable both in their ability to treat large scientific computational chemistry problems efficiently, and in their use of available parallel computing resources from high-performance parallel supercomputers to conventional workstation clusters. NWChem software can handle: Biomolecules, nanostructures, and solid-state; quantum and classical (in all combinations); Ground and excited-states; Gaussian basis functions or plane-waves; Scaling from one to thousands of processors; Properties and relativistic effects.

Turbomole [14] CP (S. Herres-Pawlis)
TURBOMOLE is a package designed for robust and fast quantum chemical applications. It provides all standard and state of the art methods for ground state calculations (Hartree-Fock, DFT, MP2, CCSD(T)), excited state calculations at different levels (full RPA, TDDFT, CIS(D), CC2, ADC(2), ...), geometry optimizations, transition state searches, molecular dynamics calculations, various properties and spectra (IR, UV/Vis, Raman, CD), fast and reliable code, approximations like RI are used to speed-up the calculations without introducing uncontrollable or unkown errors in parallel version for almost all kind of jobs.

ORCA [15] CP (S. Herres-Pawlis)
ORCA is a flexible, efficient and easy-to-use general purpose tool for quantum chemistry with specific emphasis on spectroscopic properties of open-shell molecules. It features a wide variety of standard quantum chemical methods ranging from semiempirical methods to DFT to single- and multireference correlated ab initio methods. It can also treat environmental and relativistic effects. Due to the user-friendly style, ORCA is considered to be a helpful tool not only for computational chemists, but also for chemists, physicists and biologists that are interested in developing the full information content of their experimental data with help of calculations.

5.2 Molecular Dynamics applications

We list here (together with a short description, references and CPs) some codes designed to deal either with quantum or classical and semiclassical treatments of the molecular motion as part of the DYNAMICS module of GEMS. Most of the codes, however, include also routines for the calculation of measurable quantities.

5.2.1 Quantum dynamics

ABC [16] CP (D. Skouteris, L. Pacifici, E. Garcia, S. Rampino)
ABC is a time independent atom diatom quantum reactive scattering program using a coupled-channel hyperspherical coordinate method to solve the Schrodinger equation for the motion of the three nuclei (A, B, and C) on a single Born-Oppenheimer potential energy surface.

FLUSS-MCTDH [17,18] CP (F. Huarte)
FLUSS-MCTDH is a pair o programs carrying out a multiconfigurational time-dependent Hartree (MCTDH) calculation of thermally averaged quantum dynamics properties of multidimensional systems based on a modified Lanczos iterative diagonalization of the thermal flux operator.

RWAVEPR [19] CP (D. Skouteris, L. Pacifici)
RWAVEPR is a time dependent atom diatom quantum reactive scattering program using Jacobi coordinates to integrate rigorously the three-dimensional time-dependent Schroedinger equation by propagating wave packets.

SC-IVR [20] CP (M. Ceotto)
SC-IVR is a Semi-classical (SC) initial value representation (IVR) program based on the outcome of a classical trajectory code used to calculate the power spectrum of medium size molecules and other spectroscopical properties.

DIFFREALWAVE [21] CP (E. Garcia)
DIFFREALWAVE is a parallel real wavepacket code for the quantum mechanical calculation of reactive state-to-state differential cross sections in atom diatom collisions.

5.2.2 Classical and semiclassical dynamics

VENUS96 [22] CP (W. Hase)
VENUS96 is a program developed and maintained by W.L.Hase (QCPE-671). It calculates the trajectory for two reactants (atoms or molecules) by integrating the Hamilton equation in cartesian coordinates. VENUS96 is linked to the Semi-Classical (SC) Initial Value Representation (IVR) program SC-IVR [13] in order to perform semiclassical calculations of vibrational and vibro-electronic spectra of molecules.

DL_POLY [23] CP (M. Albertì)
DL_POLY is the most used code in the CMMST VRC for the integration of the classical equation of motion of Molecular dynamics calculations. It is a general purpose package of subroutines, programs and data designed to facilitate Molecular Dynamics simulations. DL POLY is continually developed at Daresbury Laboratory by W. Smith and I.T. Todorov under the auspices of the British EPSRC and NERC in support of CCP5. It can be used to simulate a wide variety of molecular systems including simple liquids, ionic liquids and solids, small polar and non-polar molecular systems, bio- and synthetic polymers, ionic polymers and glasses solutions, simple metals and alloys.

GROMACS [24] CP (IGI, C. Soares)
GROMACS is a versatile molecular dynamics package integrating the Newtonian equations of motion for systems with hundreds to millions of particles. It has been used in a large number of case studies and a complete workflow aimed at exploiting the interoperability between a local cluster platform (HPC capable) and a Grid platform (mainly HTC capable) has been developed. In the workflow, the coupling among the various jobs is taken care using a link (a semaphore) that defines the dependency job chain.

NAMD [25] CP (IGI)
NAMD is a parallel molecular dynamics code designed for high-performance simulations of large bio-molecular systems and it has been used to study the behaviour of a lipidic bi-layer in a water box. Ported on the Grid environment by using OpenMPI parallel libraries, a Direct Acyclic Graph (DAG) has been implemented to run the code in a semi-automatic way and facilitate the user to carry out his/her calculations.

AutoDock [26] CP (J. Krüger)
AutoDock is a suite of automated docking tools. It is designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure. Current distributions of AutoDock consist of two generations of software: AutoDock 4 and AutoDock Vina.
AutoDock 4 actually consists of two main programs: autodock performs the docking of the ligand to a set of grids describing the target protein; autogrid pre-calculates these grids. In addition to using them for docking, the atomic affinity grids can be visualised. This can help, for example, to guide organic synthetic chemists design better binders. AutoDock Vina does not require choosing atom types and pre-calculating grid maps for them. Instead, it calculates the grids internally, for the atom types that are needed, and it does this virtually instantly.

CADDSuite [27] CP (J. Krüger)
CADDSuite offers modular tools for most commonly used tasks in the field of computer-aided drug design, that all have the same interface and can easily be used to create even complex workflows. There are algorithms and tools for data storage and retrieval, data preparation, chemical checks, QSAR, Docking, Rescoring and for the analysis of results. CADDSuite has also been integrated into the workflow system Galaxy, in order to make submitting jobs (to a cluster/cloud/grid) or creating, modifying or starting workflows even easier for the user. In essence, a user can thus easily create drug-design pipelines directly from a web browser, without any need for software installations on his local computer.

FlexX [28] CP (J. Krüger)
For a protein with known three-dimensional structure and a small ligand molecule, FlexX accurately predicts the geometry of the protein-ligand complex within a few seconds. The intuitive GUI permits the set up of docking runs within a single minute and provides you with instantaneous visual feedback.
FlexX sets new records in vHTS. You can screen a library of ~1 000 000 compounds in about 8 hours on a 30-node cluster. The new Screen-module also allows you to filter out false positives on-the-fly. If you are screening compounds from a combinatorial library, you can take advantage of a novel pharmacophore-based combinatorial docking to further gain significant speed-up and enrichment.

DESMOND [29] CP (J. Krüger)

DESMOND can compute energies and forces for the standard fixed-charged force fields used in biomolecular simulations. A variety of integrators and support for various ensembles have been implemented in the code, including methods for thermostatting (Andersen, Nose-Hoover, and Langevin) and barostatting (Berendsen, Martyna-Tobias-Klein, and Langevin). Ensembles typically used in membrane simulations (constant surface area and surface tension) and semi-isotropic and fully anisotropic pressure coupling schemes are also available. Desmond supports algorithms typically used to perform fast and accurate molecular dynamics. Long-range electrostatic energy and forces are calculated using particle-mesh-based Ewald techniques. Constraints, which are enforced using a variant of the SHAKE algorithm, allow the time step to be increased. These approaches can be used in combination with time-scale splitting (RESPA-based) integration schemes. The Desmond software includes tools for minimization and energy analysis (which can be run efficiently in a parallel environment); methods for restraining atomic positions as well as molecular configurations; support for a variety of periodic cell configurations; and facilities for accurate checkpointing and restart.

5.3 Grid tools and services

As part of the activities of the community addressed also to improve the utilization of the synergistic model not only in research but also in education some Grid tools and services have been designed and are being further developed. These tools and services are listed here (together with a short description, references and CPs)

GC3PIE [30] CP (GC3 - Maffioletti)
GC3PIE is a library of Python classes for running large job campaigns (high throughput) on diverse batch-oriented execution environments (such ARC) providing facilities for implementing command-line driver scripts in the form of Python object classes whose behavior can be customized by overriding specified object methods. In several computational campaigns the adoption of the framework GC3Pie allowed the users to define event-related dependencies between different applications and execute them simultaneously on a large scale distributed computing infrastructure. Thanks to this the programmatic approach to workflows, the entire execution schema is assembled at runtime and steps can be added and removed dynamically as the program progresses, adapting to the outcome of individual computations;

GriF [31] CP (COMPCHEM VO - Manuali)
GriF is a Service Oriented Architecture (SOA) Collaborative Framework designed to facilitate the use of the DCIs by non specialists with a particular focus on the optimization of the selection of computing elements for running single and parameter study applications. It consists on a set of java modules aimed at submitting and monitoring jobs on different computing platforms and evaluate the quality of the service offered by the computing site and of the use made of the resources by the users. More details will be given later when discussing the development of the synergistic model;

IGI Portal [32] CP (IGI)
IGI Portal is a science gateway developed within the activities of the Italian Grid Initiative (IGI), which is operating the Italian Grid infrastructure. Already used by other by COMPCHEM to offer some of its applications as a service and used as well by other communities (https://wiki.italiangrid.it/twiki/bin/view/UserSupport/UserSupportDisciplines), the IGI web-based portal provides to the final user several important services such as job submission, workflow definition and data management for both Grid and Cloud environments. Moreover, it implements a

robust Authentication and Authorization (AA) mechanisms to provide a secure computational environment and, at the same time, hiding the complexity of the X.509 digital certificates on which most of the DCI resources rely;

InSilico Lab [33] CP (CYFRONET)
InSilico Lab is an application portal designed to support in silico experiments by easily running computational chemistry software on grids. Unlike manual job submission or Grid portals, InSilicoLab enables to run computations on grids without technical knowledge of how to operate it;

MyGAMESS [34] CP (GC3 - Maffioletti)
MyGAMESS is a Framework for Integration of Chemistry applications with the Swiss Grid Portal aimed at providing dynamic access to applications services. Users will profit for example by a considerable reduction of time-consuming data format conversions and/or other code restructuring, once the capabilities of different applications are integrated;

G-LOREP [35] CP (COMPCHEM – Tasso; Varella)
G-LOREP (Grid LOs Repository) is a manager of a federation of distributed repositories of research based learning objects. It relies on a central database and allows registered users to create and/or download learning contents for research based education activities. G-LOREP makes also the content automatically shareable among the federation servers and enables the creation of dependency management software attachment as well as fault tolerant submission of the simulations from any federated site.

KEPLER [36] CP (MosGrid)
KEPLER is an open source, scientific workflow application. Using Kepler graphical interfaces and components, scientists can create and execute scientific workflows which allow accessing scientific data and executing complex analyses on them. Kepler has been recently adopted in a recent work [48] where a prototypical workflow has been designed and tested for the calculation of the time independent quantum probabilities of the $H+H_2$ benchmark reaction. For this purpose a set of quantum mechanical codes, both belonging to the set of the aforementioned computational applications, have been used;

6 MANAGEMENT AND SUSTAINABILITY

The VRC ecosystem involves many different actors like VOs, University Departmental structures, Research centers, EGI and other resource infrastructure providers. These actors undertake the provision or consumption of human, infrastructure and technical services within the ecosystem that need to be supported by an appropriate business model. This is, indeed, the role of the VRC management body that will be described in the present section.

6.1 Coordination and Management

The formal acceptance of the Grid model and the establishing of VRC in EGI is the signature of a Memorandum of Understanding (MoU) between EGI and the VRC. The CMMST community has already expressed its wish to be formally part of EGI through an expression of interest formulated by the COMPCHEM VO. Yet, due to the fact that only the signature of a MOU provides a means of documenting the relationship that EGI has with the community as well as the activities and objectives of such a collaborative relationship (a MoU, is used when both parties do not want to pursue a contract that is legally binding while it still clarifies the relationships, responsibilities and communication

channels between two or more parties that may share services, clients, and resources) we give below a sketch of the coordination and management structure that in the first instance will be provided by the members of the participating VOs for the time period needed for the first consolidation of the VRC.

6.1.1 Management bodies for the VRC

The Management body of the VRC is the Board of Stakeholder (BoS) that is made by 5 representatives of the stakeholder VOs (one representative for each fraction of 20% of the total number of members of the VRC belonging to the VO), 3 representatives (one each) for the three main stakeholder resource providers and 2 representatives of the most contributing (in terms of nationality of the members) stakeholder National Grid Infrastructures.

In this spirit the VRC management body is here defined:

- VRC leader: is the person representing the VRC and is the MoU contact point and is elected by the BoS. Such position will be taken for the first three years by the COMPCHEM coordinator Antonio Laganà).

- VRC deputy leader: is the person of the BoS acting on the behalf of the VRC leader and appointed by him/her.

- VRC activity coordinators: are the persons appointed by the BoS, after the proposal of the VRC leader, representing the activity of the VRC in strong coordination with EGI and the participant VOs and acting as experts in the assigned roles.

    - User support and training Coordinator: is the person taking care of managing user support and training activities (A1).

    - Operations Coordinator: is the person taking care of managing the services and operation activities (A2).

    - Technical Coordinator: is the person taking care of managing requirements gathering activities (A3).

    - Communication and Dissemination Coordinator: is the person taking care of managing communication and dissemination activities (A4).

6.1.2 TASK A1: User support and training

VRC representatives: to define

Roles:

- provide information about estimates on the size of the VRC and possibly its composition by country;
- provide dedicated user support activities for code building, adapting and porting on multi platform computing models and storage;
- provide tutorials aimed at enhancing the use of services made available to the community;
- provide examples, training material, details of specialist applications, documentation and presentations to be made accessible to members of the scientific community.

6.1.3 TASK A2: Services and operations

VRC representatives: to define

Roles:

- provide robust, well designed, user centric services to scientific user communities;
- provide local and global operational services as needed to support international users and operations;
- drive the adoption of standards within tools and applications production;
- provide compliance with the operation interfaces needed to ensure seamless and interoperable access to networked (Grid and Cloud) resources;
- utilize a credit system to encourage CMMST users to cooperate in developing higher level of complexity applications;
- produce VRC-specific availability statistics make use of the EGI availability computation system.

6.1.4 TASK A3: Requirement gathering

VRC representatives: to define
Roles:

- define and monitor Service Level Agreement (SLA) for third-level support on incidents and requests;
- participate to EGI security police team to contribute to the development to the security police fabric of the infrastructure;
- participate in the Operations Management Board to contribute to the EGI operations agenda;
- subscribe to a mandatory set of EGI policies, procedures and Operation Level Agreements (OLA)s.
- design a roadmap for the harmonization of the participating VO and research groups structure

6.1.5 TASK A4: Communication and Dissemination

VRC representatives: to define
Roles:

- define communication strategies aimed at attracting more CMMST users into a common endeavour offering the possibility of assembling higher level of complexity applications and services;
- establishing contact points for communication channels and publications;
- disseminate results of the collaboration by planning joint sessions at EGI.eu and CMMST events and by informing EGI.eu and each VRC members of any scientific/academic publications published/issued related to the VRC activities.

6.2 Sustainability

A crucial test for the CMMST VRC is, as usual, the achievement of a sustainability regime. This implies a balance between the value of the services provided to the members and the amount of the resources invested (such balance is easier to achieve if external funding from projects and donations is

available). In such regime, services can be provided for free (to a large extent) to all members of the community (including the simple users of level 1) and may be used for dissemination activities.
Accordingly, to the end of measuring the quality of the offered services a proper metrics that allows to estimate the value of the services provided internally by members of level 2 and 3 and by pure resource providers are proposed while the value of externally provided services is set by the market.

6.2.1 Quality evaluation

The CMMST VRC members, in addition to paying significant efforts in order to attract financial resources specifically targeted to the development of the VRC, are committed to provide proposal to extend the original version of GriF to enable it to support the evaluation of the Quality of the Services (QoS) provided by its members to the community (services can be provided under the form of participating to the management of the Grid, writing (and submitting and (once approved) managing) proposals to calls, designing and developing tools and applications, opening new lines of research and innovation, activating commercial channels for procuring revenues, etc.

QoS parameters play an important role not only because, as already mentioned, provide an objective evaluation of the services and allow a better use of the resources but also because they provide an objective ground to mutual trust in collaborative work. As a matter of fact the QoS parameters adopted by the CMMST VRC:

- enable users to ask for Grid Services by specifying as keywords high-level capabilities rather than, for example, memory size, cpu/wall time and storage capacity;

- allow managers to guarantee the automatic selection of the most appropriate low-level capabilities related to the current Grid job and the adoption of different running policies (in other words, when a Grid job has to be run, GriF can make use of different system requirements according to the class level of the user owning the Grid job);

- offer a Selection of resources [37] (rather than a pure Discovery) mechanism based on a measure of the quality of the available Grid Services;

- articulate the Grid Economy Model as Costs to be paid for the Services utilized and Credits earned in return for the efforts spent on behalf of the VRC.

6.2.2 The adopted requisites

In the implementation of the QoS, reference will be made to the following requisites: Accessibility, Integrity, Performance, Reliability, Availability and Security) . Yet, a further requisite (Innovation) will also be considered to meet a central goal of the VRC that is in need of Active Software Providers (see level 2 of the above mentioned members' scheme).

Innovation can be expressed in terms of objective variables like the submitted number of jobs, the amount of the result produced or the services (availability and and support for applications and tools) offered to the community.

Therefore, other kinds of variables have been considered as. for example, those concerning the so called social (e.g. the promotion of universal human values like peace. the promotion of universal availability like open source software or the promotion of a universal welfare) and environmental

spects (e.g the amount of natural harmlessness including energy saving, ecocompatibility of materials or physical space conservation of the computing systems).

It is worth emphasizing here that in applying QoS to COMPCHEM, we did put particular emphasis in defining the parameters concerning the social and environmental aspects.

7 PERSPECTIVE RESEARCH AND DEVELOPMENT

In the spirit of implementing the proposed synergistic model through an increase of QoS, the CMMST community will seek a continuous extension of the services by fueling a regular stream of additional research, design, implementation and validation work for new solutions. As a consequence, the porting of further applications relevant to molecular simulators (see subsection 7.1) and the implementation of an improved version of GriF (see subsection 7.2) will be actively pursued. Moreover, significant collaborations will be activated with the developers of workflows and portals (see subsection 7.3) as well as with the managers of XSEDE (see subsection 7.4) in order to bridge the EGI HTC with its HPC platform and to establish de facto big data format standards (see subsection 7.5).

7.1 Further applications for CMMST

We list here, using the same format as before, some additional applications planned to be ported in the Grid environment:

CRYSTAL [38] CP (COMPCHEM, F. Filomia)
CRYSTAL is a commercial quantum chemistry ab initio program written by V.R. Saunders, R. Dovesi, C. Roetti, R. Orlando, C.M. Zicovich-Wilson, N.M. Harrison, K. Doll, B. Civalleri, I.J. Bush, Ph. DArco, and M. Lunell within a collaboration of the Theoretical Chemistry Group of the University of Torino (IT) and the Computational Materials Science Group at the Daresbury Laboratory (UK) with other researchers. CRYSTAL is recognized within the computational chemistry community as a powerful tool for carrying out solid state simulations useful for scientific and technological applications;

QUANTUM ESPRESSO [39] CP (S. Cozzini, M. Ceotto)
QUANTM ESPRESSO is an integrated suite of Open-Source computer codes for electronic-structure calculations and materials modeling at the nanoscale based on density-functional theory, plane waves, and pseudopotentials;

APH3D [40] CP (G. Parker, L. Pacifici)
APH3D is a time independent atom diatom quantum reactive scattering program using a coupled-channel hyperspherical coordinate method to solve the Schrodinger equation for the motion of the three nuclei (A, B, and C) on a single Born-Oppenheimer potential energy surface;

GFIT3C and GFIT4C [41] CP (L. Pacifici)
GFIT3C and GFIT4C are routines devoted to the fit of the Potential Energy surface of three and four body molecular systems using the formulation of M. Paniagua based on Many Body expansions and polynomials of exponentials time internuclear distances;

POMULT [42, 43] CP (S. Farantos)
POMULT is a fortran program supporting the spectroscopy studies of large molecules by locating periodic orbits and equilibrium points in Hamiltonian systems based on 2-point boundary value solvers which use multiple shooting algorithms;

FORWCONV [44] CP (N. Balucani)
FORWCONV is a routine devoted to the forward convolution of the Newton Diagrams of a distribution of Newton diagrams to simulate the composition of a gas phase molecular beam.

DALTON [45] CP (SCALALIFE)
DALTON is an electronic structure program capable of performing quantum mechanics calculations using Hatree-Fock, Multiconfigurational Self Consistent Field and Coupled Cluster wavefunctions, and Density Functional Theory for molecular systems of various sizes.

DISCRETE [46] CP (SCALALIFE)
DISCRETE is a package devised to simulate the dynamics of proteins using the Discrete Molecular Dynamics (DMD) method. The difference with respect to a standard molecular dynamics simulation is that the particles are considered to move with constant velocity until a collision occurs. As the particles move in ballistic regime, the motion equations do not need to be integrated.

ACCQD [47] CP (M. Ragni)
ACCQD - Adiabatic Curves Calculator for Quantum Dynamics is a program automatizing the calculation of the adiabatic energy values obtainable when solving on a given Potential Energy Surface the time-independent Schrodinger equation at fixed values of the hyperradius. The program is written in C++ and makes use of parallel algorithms.

CHIMERE [48] CP (Crocchianti)
CHIMERE is a package implementing a multi-scale model primarily designed to produce daily forecasts of ozone, aerosols and other pollutants and to make long-term simulations for emission control scenarios. CHIMERE runs over a range of spatial scales from the regional scale (several thousand kilometers) to the urban scale (100-200 Km) with resolutions from 1-2 Km to 100 Km.
The version of CHIMERE already implementated at the University of Perugia (IT) is the 200606A one (this was the most stable version of the code available when we started this investigation and the first validated parallel version) is made by two different modules:.

1. Module A: which prepares the meteorological data in the netCDF [49] format needed by CHIMERE, diagnoses turbulent parameters and calculates biogenic emissions (meteorologically dependent);

2. Module B: which contains a central chemistry-transport calculation unit with interfaces to several databases (emissions, meteorology, boundary conditions, land use). Each interface transforms original or processed data to CHIMERE input formatted data adapted to the time period and region to simulate. CHIMERE offers the option of including different gas phase chemical mechanisms. The original, complete scheme [50], called MELCHIOR1, describes more than 300 reactions of 80 gaseous species. In order to lower the computing time a reduced mechanism with 44 species and about 120 reactions is also implemented Efficiency parameters of MELCHIOR1 are tabulated and can be grounded using the ab initio approach of GEMS.

Similar treatments are being adopted for simulations of interest for combustion of non fossil fuels [51] and spacecraft re-entry [52].

7.2 GriF

The present version of GriF consists of two types of Java servers (an information registry of available services (YR) and a container providing Web Services (YP)) and one type of Java client (YC, used as interface with the user). The Java servers describe the services and reference self-describing interfaces which are published in platform-independent documents. The Java client, instead, takes care of running the jobs on the associated User Interface (UI) and supports the correct interfacing of the Web Services offered by GriF. The program running on the GriF client takes care of:

– The management of the basic Grid operations (e.g. checking single and multiple job status even implemented on mobile devices);

- The search of the programs on the information registry YR (e.g. searching for a YP exposing a Web Service offering the execution of the requested programs);

- The optional introduction of changes in the requested programs;

- The compilation of new executables on the selected YP by replacing the ones provided as default. Users can also utilize the results of any previous actions to start the execution of the Grid job after passing a user-specific input file (when the default one is not appropriate) and monitor the jobs status eventually retrieving the results.

In other words using GriF the members of the Grid community are able to set up and manage YPs to expose their services in an open, standard and secure way even if having little familiarity with the wrapped programs and the Grid platform. Moreover, thanks to its robust SOA framework nature GriF can support collaboration among researchers bearing complementary expertise and contributing to collaborative work when they want to articulate their computational applications as a set of multiple tasks.

GriF provides Service Orchestration features which foster the establishing of collaborative operational modalities and enhance the possibility for users and providers to cooperate in carrying out higher level of complexity computational activities leading to a gain of Credits. These Credits can then be redeemed via a preferential utilization of the resources (including the financial ones) of the community.

It is important to stress here again that in the improved version of GriF, the concept of Service Discovery (in which users searching for Grid Services are provided with an unranked list of matchings) is surpassed by that of Service Selection (in which the ranking is based on QoS). At the same time, simple user quality metrics are surpassed by the Credits award mechanisms based on the new concept of Quality of User (QoU).

In order to quantify QoS, we have identified appropriate sets of non-functional properties and parameters characterizing it [53]. It is worth mentioning, however, that we have considered here as a Grid Service any set of collaborative Web Services of GriF running on the Grid by sharing a common

distributed goal. As to QoU, instead, we refer to the collection and filtering of different implicit and explicit information provided by users as will be implied in the discussion below.

We want further comment that such development, in addition to leveraging on collaboration, also stimulates a significant extent of competition among the VRC members by introducing an incentive to produce innovative Services and improve the existing ones. Moreover, such mechanism has also the ”social” effect of encouraging the VRC members to futher step up the quality of membership and to contribute to the infrastructure development.

Moreover, for the use in the proposed CMMST VRC improvements are planned to cope with codes of higher levels of complexity that need to be dealt by adopting appropriate interfaces and portals in order to efficiently support both users having a low level of expertise and users wishing to customize their applications for advanced studies.

7.3 Workflows and portal technologies

As already anticipated, an important step of the planned work is the use of suitable portals. Significant interactions are presently being developed with the developers and users of WS-PGRADE [54] (Sztaki, Budapest (HU), INAF, Trieste (IT)) thanks to the recent advances of the product (see also http://www.sci-bus.eu/science-gateways) and the recent interaction with the managers of the XSEDE project fostered by EGI in the final phase of the VT.

Some details of the tools that will be used for this are given in the followings:

SCI-BUS gateway framework (WS-PGRADE/gUSE) [54] CP (MTA-SZTAKI)
gUse (grid and cloud user support environment) is a permanently improving open source science gateway (SG) framework developed by the EU FP7 SCI-BUS project. It enables users the convenient and easy access to grid and cloud infrastructures. It has been developed to support a large variety of user communities. It provides a generic purpose, workflow-oriented graphical user interface to create and run workflows on various Distributed Computing Infrastructures (DCIs) including clusters, grids, desktop grids and clouds. The framework can be used by National Grid Initiatives (NGIs) to support small user communities who cannot afford to develop their own customized science gateway. Currently more than 30 science gateways are established based on this technology in Europe including the MoSGrid Science Gateway. Moreover, the already mentioned IGI Portal integrates this technology with other technologies (like DIRAC and an ad hoc designed Data Management). Further details are given in: http://sourceforge.net/projects/guse/, https://www.sci-bus.eu/, http://guse.hu/.

SHIWA SIMULATION PLATFORM [55] CP (MTA-SZTAKI)
SHIWA is a project providing a multi-systems workflow execution platform and an interoperability solution. The platform currently supports seven workflow systems: Askalon, Kepler, LONI pipeline, MOTEUR, Taverna, Triana and WS-PGRADE. The SHIWA Simulation Platform (SSP) enables scientific communities working in the same field but using different workflow systems to collaborate, share and reuse their workflows even written in different workflow languages. For example, several members of the CMMST community use WS-PGRADE workflows (MoSGrid) other use Kepler. By using the SSP they can mutually reuse the workflows developed by the other community.

In this respect the options to consider are:
1. Connect an existing GUI (in this case GriF GUI) to the WS-PGRADE gateway through the 'remote API'. WS-PGRADE would act as a high level middleware that submits and manages jobs and data (files) on various DCIs (Particularly on gLite-EGI; Unicore-XSEDE)
2. Setup a complete WS-PGRADE stack with a GUI and customise/extend it with portlets for CMMST. WS-PGRADE would act both as the job-data manager, and as a graphical environment for users. The GUI can incorporate community-specific components, such as QoS, QoU services, from GriF (programming language should be checked).

Further problems that will be tackled are:
* due to the fact that WS-PGRADE has no a built-in broker, one has to decide a priori which type of resource to use for the different tasks of the workflow, to whom should be left the responsibility of the workflow development for allocating the workflow tasks to VOs (for this one has to consider whether the VO has a broker, such as a the gLite VO has WMS, or a particular site for this (if there is no VO or broker in the VO, such as for HPC sites of XSEDE). Within two months there will be a new WS-PGRADE release that will contain a simplified built-in broker to solve the above mentioned problem. Within the framework of the planned CMMST VRC project SZTAKI is ready to substantially improve this built-in broker.
* how to ensure access to individual grant-based usage of the resources (typically the HPC ones) or adopt an alternative community grant type. Due to the fact that WS-PGRADE can at present deal only with grant holders (though it could be easily modified for community access and WS-PGRADE can already use robot certificates in EGI VOs), find out what is needed on the user side to systematically work with robot instead of personal certificates)
* how to extend and integrate WS-PGRADE accounting information collected from the grid sensors about running time performances of the jobs and of the users in order to satisfy GriF requirements. In particular in order to formulate QoS and QoU evaluations it is necessary that the framework (grid layer) automatically records in a data base new information about users and related jobs that can be then regularly collected (say via command line) to feed the adopted metrics. Then, following an exhaustive and correct recording of such data, it is also necessary to implement a server (or a service) to evaluate the related formulae and produce the resource ranking (typically at regular intervals) for the users themselves and the management. Within the framework of the planned CMMST VRC project SZTAKI is ready to extend WS-PGRADE with the resource metrics and ranking mechanisms required by the CMMST community. The above mentioned built-in broker will be able to use these information for improving its resource selection algorithms.

* Negotiations already started with XSEDE resource providers to test the integration of WS-PGRADE with XSEDE resources.

7.4 Bridging HTC and HPC

To the end of bridging in a synergistic fashion HPC and HTC platforms to the end of overcoming the highly unsatisfactory situation in which neither the use of HPC nor that of HTC alone is optimal to meet the requests of a variety of complex CMMST applications (at present use is made of an experimental SSH procedure). In the work planned for establishing the CMMST VRC the functionalities of the improved version of GriF will be exploited the end of interoperating large computational applications. This in fact allows an optimization of the usage of both HTC and HPC computing resources because it is not infrequent the case in which a user utilizes HPC platforms not as such but as a bunch of loosely coupled processors underutilizing their fast dedicated network. At the same time HTC users may utilize massively distributed HTC platforms to solve tightly coupled

computational tasks ending up by wasting a large amount of time in transferring data on the net. A coordination of the two types of platforms to interoperate via a single workflow (or workflow of workflows) and properly manage the various components on the most appropriate hardware, would instead allow a clever composition of complex applications optimizing the use of the various computing resources and providing the users with the best level of performance. This is, indeed, a key contribution of MoSGrid to the VRC as well as the goal of the agreement reached between EGI and XSEDE (https://www.xsede.org/) aimed at allowing COMPCHEM to interoperate from the European Grid infrastructure on the supercomputer US network.

7.5 Data Format standards

An already mentioned advantage of the synergistic distributed model consists in the possibility for the user to choose the platform better fitted for retrieving/storing programs and data for his/her applications. This provides the CMMST members with the possibility of better combining different pieces of software when assembling complex realistic applications using GEMS [4] (or any other collaborative simulator) and undertake more ambitious research projects. Such possibility of building workflows of shared programs has, for example, stimulated the setting of proper (de facto) standards of data in quantum chemistry and quantum dynamics (like the Q5Cost and D5Cost [56]). This has also prompted, as already mentioned, the development of tools (like the framework GriF) enabling the redirecting of computer applications to run on the best suited sites (including HPC machines) and enhancing cooperative compute capabilities by opening the perspective of combining different complementary know how into single (higher level of complexity) realistic applications and applying as a community for more ambitious research grants. A fundamental necessary implication of all that is the further development of standards for the representation of data (sometimes huge sets) of different types for CMMST applications (e.g. data standards developed by the ScalaLife project (http://scalalife.eu/content/data). Moreover the CMMST community wishes to collaborate in some other EGI activities as, for example, the EUDAT and PRACE pilot ones [57] aimed at bringing together infrastructures and user communities by focusing their interest in the data sharing facilities (see Appendix A).
Such activity is expected to provide a set of tools and recommendations that the CMMST community can use to integrate its CMMST workflows.

8 CONCLUSIONS

In the present document the advantages that a VRC status would offer to the members of the CMMST community in terms of access and use of the computing resources federated in EGI are listed. The document depicts the present scenario of computational resources and their usage by analysing the characteristics of the grant and the opportunistic models as opposed to the synergistic one proposed for the CMMST community. The document enumerates also the technical and non technical aspects of such cooperative model based on the selection of the resources (from personal systems to supercomputers) and services (from number crunching to massive data handling on heterogeneous platforms) as well as on an advanced usage of QoS and QoU and on an introduction of a community economy based on a credit system. On this ground, the CMMST VRC (as well as other communities which use distributed computing resources) will be able not only to share hardware and software that

(thanks to the expertise offered the various groups) can be integrated to address complex problems and computing simulations of high social impact.

Table 1

Packages and frameworks in-use by the CMMST Community over the EGI Grid

Application	Description	**License**
ABC	Solve the Schrodinger equation for triatomic systems using the time independent quantum method	Academic
FLUSS-MCTDH	MultiConfigurational Time Dependent Hartree method Lanczos iterative diagonalization	Academic
NB-MCTDH	N-Body MultiConfigurational Time Dependent Hartree method	Academic
VENUS96	Quasi-classical dynamics of reactive collisions	
DL_POLY	Classical Molecular Dynamics	
NAMD	Classical Molecular Dynamics	Academic
GAMESS-US	General Atomic and Molecular Electronic Structure Package	Academic
RWawePR	Time Dependent Method to Solve the quantum reactive Scattering equations for triatomic systems	Academic
GROMACS	GROningen MAchine for Chemical Simulations	Academic
SCIVR	Semiclassical initial value representation method	Academic
Framework	Description	
GriF	Grid Framework enabling efficient and user-friendly scientific massive calculations	Free
Gcres	Quality of Users (QoU), Quality of Services (QoS) evaluation Framework	Free
MyGAMESS	Front-end script for submitting multiple GAMESS-US jobs	Free
InSilicoLab	Application portal designed to support in silico experiments	Free
IGI Portal	science gateway providing a workflow environment and a Cloud access	Free
GC3PIE	Framework providing building block to build dynamic workflows	Free

APPENDIX A – EGI/EUDAT/PRACE ACTIVITY

The goal of this EGI activity is to bring together European infrastructures (EGI, EUDAT and PRACE) and user communities to discuss the requirements in sharing and using their data between different environments, which can be general infrastructures or domain specific facilities.

The main objectives of this activity are here highlighted

- identify common data access and transfer tools and protocols
- demonstrate real benefit for the involved communities
- identify technology and/or organisational gaps and suggest measures for improvements

On behalf of such objectives, pilot activities were set up as a result of the EGI/EUDAT/PRACE workshop on data management [48] and address the user communities' requirements by defining and setting up test environments using existing technologies like GridFTP, UNICORE, Globus Online, iRODS, EMI data management services, etc.

Several pilot activities were defined to address specific but generic use cases of usage of different resource infrastructures:

- Pilot 1. & 2: Data sharing and uniform data access across e-infrastructures and community centres
- Pilot 3. ScalaLife and PRACE
- Pilot 4. Monitoring and resource discovery

APPENDIX B – THE CMMST VT MEMBERSHIP

(list all those involved in the team and doing the work)

Country Code:	Name:	e-mail:	Role:
IT	Antonio Lagana University of Perugia	lagana05@gmail.com	VT Leader Application List QoU-QoS DCC liaison VRC structure (coord)
IT	Daniele Cesini IGI, INFN	daniele.cesini@cnaf.infn.it	Attract and Train users
IT	Alessandro Costantini IGI, INFN	alex.costantini@gmail.com	Application List QoU-QoS Attract and Train users
IT	Carlo Manuali University of Perugia	carlo.manuali@gmail.com	QoU-QoS (coord)
IT	Michele Ceotto University of Milano	michele.ceotto@unimi.it	Application list
GR	Stavros C. Farantos, IESL/FORTH and Chemistry Univ. of Crete	farantos@iesl.forth.gr	Application List
GR	Stamatis Stamatiadis, IESL/FORTH and Materials Univ. of Crete	stamatis@materials.uoc.gr,	Application list
GR	Manos Giatromanolakis, IESL/FORTH	gmanos@iesl.forth.gr	Application list
GR	Giannis Remediakis, Materials Univ. of Crete	remed@materials.uoc.gr	Application list
GR	Evangelia Varella University Thessaloniki	varella@chem.auth.gr	ECTNA liaison
HU	George Lendvay (MTA TTK Research Center for Natural Sciences)	lendvay@chemres.hu	Application list
HU	Anna Vikar (MTA TTK Research Center for Natural Sciences)	vikaranna@hotmail.com	Application list
HU	Peter Szabo (MTA TTK Research Center for Natural Sciences)	xysma@msn.com	Application list
HU	Tibor Nagy (MTA TTK Research Center for Natural Sciences)	nt2001@freemail.hu	Application list

HU	Akos Bencsura (MTA TTK Research Center for Natural Sciences)	bencsura@gmail.com	Application list
NL	Remco Havenith University of Groningen	r.w.a.havenith@rug.nl	Application list
ES	Fermin Huarte Larrañaga University of Barcelona	fermin.huarte@gmail.com	Application list
ES	Amaia Saracibar (University of the Basque Country)	amaia.saracibar@ehu.es	Application list
ES	Ernesto Garcia (University of the Basque Country)	e.garcia@ehu.es	IBERGRID liaison Application list
FR	Marco Verdicchio (CNRS)	m.verdicchio@gmail.com	Application list
CR	Petr Hanousek (CESNET)	petr.hanousek@cesnet.cz	Application list
UK	Peter Oliver (STFC)		Application list
USA	Alan Sill (CERN)	alan.sill@ttu.edu	HTC-HPC (coord)
USA	Bill Hase (TTU)	Bill.Hase@ttu.edu	Application list
CH	Sergio Maffioletti (UZH)	sergio.maffioletti@gc3.uzh.ch	CC3 liaison
PL	Mariusz Sterzel (CYFRONET)	m.sterzel@cyf-kr.edu.pl	GAUSSIAN liaison
EGI	Elena Tamuliene (VU)	jelena.tamuliene@tfai.vu.lt	Application List (coord)
EGI	Gergely Sipos (UCST)	gergely.sipos@egi.eu	EGI technol. (coord) HTC-HPC
EGI	Richard McLennan (UCST)	richard.mclennan@egi.eu	VRC Sustain. (coord) EGI technologies
EGI	Karolis Eigelis (UCST)	karolis.eigelis@egi.eu	Application list
SE	Rossen Apostolov (KTH Royal Institute of Technology)	rossen@kth.se	ScalaLife liason
DE	Sonja Herres Pawlis (University of Munich)	Sonja.herres-pawlis@cup.uni-muenchen.de	MoSGrid coordinator

ANNEX

Here a brief description of the partner VOs of the project (to which belong the researchers involved in the VT, see appendix B) is given.

COMPCHEM

COMPCHEM is the proposer VO that is made of a group of European molecular and materials science research laboratories (see the list of the members of the VT team) committed to implement their computer codes and computational applications on the production Grid infrastructure made available by EGI. COMPCHEM has participated in the flagship European grid projects of the EGEE series and more recently in the EGI-Inspire one as member of the IGI JRU. A. Laganà is the coordinator of the VO with the assistance of O. Gervasi (Department of Mathematics and Informatics, University of Perugia, IT) and A. Costantini (IGI, INFN).

Prof. Antonio Laganà is Chairman of the European Computational Chemistry division of EUCHEMS, member of the international committee of the Erasmus Mundus Euromaster in Theoretical Chemistry and Computational modelling, Coordinator of the Virtual Education Standing Committee of the European Chemistry Thematic Network Association, Director of the Chemistry Department of the University of Perugia.

Website: https://www3.compchem.unipg.it/compchem/

GAUSSIAN

GAUSSIAN is the VO taking care of the Chemistry and Molecular Science applications of the Akademickie Centrum Komputerowe CYFRONET in Krakow one of the largest Polish supercomputer and networking centres. It constitutes the separate organizational entity of the AGH University of Science and Technology in Krakow. The man mission of CYFRONET is to offer access to its computational facilities and network services to universities and research institutes and perform consulting expertise, training and educational activities. CYFRONET coordinates implementation of Polish NGI in PL-Grid Project. It took part in a series of FP5 – FP7 projects including the various stages of EGEE and EGI-Inspire in which it coordinated the CE ROC and the Computational Chemistry Cluster of Excellence. The main area of expertise of the GAUSSIAN VO is to enable the Gird access to the Gaussian suite of codes.

The coordinator of the GAUSSIAN VO is Mariusz Sterzel, Ph.D in Chemistry, expert in computational Chemistry (application of first principle and molecular dynamics methods to vibronic spectroscopy and NMR studies), EGEE responsible for grid ports of commercial chemical software packages Gaussian, Turbomole, Dalton and GAMESS. During EGEE-II responsible for computational Chemistry, member of NA4 and UF Steering Committees and during EGEE-III leader of Computational Chemistry Cluster of Excellence.

Website: http://www.lsr.nectec.or.th/index.php/How_to_register_on_Gaussian_VO

Chem.vo.ibergrid.eu

Chem.vo.ibergrid.eu is the joint Spanish–Portuguese VO connecting universities and research institutions to the end of providing them with the support shared resources, operational support, and

area specific applications. Typical packages offered on the grid are: Gromacs (to perform molecular dynamics for systems with up to millions of particles), RPMDrate (to carry out ring polymer molecular dynamics (RPMD) simulations calculations to evaluate the bimolecular reaction rate coefficients for thermally activated processes in the gas phase). The VO relies on accesses to compute resources like those of the *ARASGRID, BIFI, BIFI-IBERGRID, CETA-GRID, IAA-CSIC, IFCA, IFIC, IFISC, NCG-INGRID-PT, CESGA, UPV-GRYCAP,NCG-INGRID-PT, CIEMAT-TIC, ESA-ESAC, SGAI-CSIC* which reserve a minimum number of slots plus other sites with no reservation.
The liaison member of the VRC to Chem.vo.ibergrid.eu is Claudio M. Soares (Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, claudio@itqb.unl.pt).

MoSGrid
MoSGrid (Molecular Simulation Grid) is a German project (German VO) aiming at providing easy access and use of molecular simulations in computational chemistry in a grid environment. Important application domains are molecular dynamics, quantum chemistry, and docking. MoSGrid offers a science gateway that allows an easy access to complex molecular simulations. The included web-based graphical user interface allows a simulation code independent setup of simulation workflows that are submitted through the UNICORE grid middleware to the underlying clusters. All users can apply commonly used metadata enriched workflows which are available in recipe repositories. The metadata description allows an efficient search for the required workflows by a description of the underlying dataflow. This lowers the hurdle for applying computational chemistry methods even for novice users.
The coordinator of MoSGrid is S. Herres Pawlis (Sonja.herres-pawlis@cup.uni-muenchen.de) Professor of Coordination Chemistry at the department of Chemistry of the University of Munich. Partners in MoSGrid are the Departments of Chemistry of the Universities of Cologne and of Paderborn, the faculty of the University of Dortmund, the computer centres of the University of Cologne, the Technical University of Dresden, of the University of Paderborn, of the University of Tubingen as well as of the Centre for information technology of Berlin, the MTA SZTAKI and the Julich Supercomputing Centre (See Ref. [51-53]).
Website: https://mosgrid.de/portal

ScalaLife Competence Center (www.scalalife.eu)
ScalaLife Competence Center provides wide range of expert support options to users and developers of applications for biomolecular simulations (Gromacs, Dalton, Discrete etc.), and to resource providers such as HPC centers. The center works directly with the main developers of the supported applications and computer science experts. Support include guidance with setup, deployment and efficient usage of the supported codes; code analysis/profiling and algorithm implementation techniques; benchmarking studies and help with applications for resource allocations. The center provides training activities and extensive documentation knowledgebase.
ScalaLife Competence Center partners are KTH Royal Institute of Technology (Sweden), Barcelona Supercomputing Center, LRZ - Leibniz-Rechenzentrum, Germany (http://www.lrz.de/) and OeRC Oxford eResearch Center (UK).
The liaison member of the VRC to ScalaLife is Rossen Apostolov, researcher at the KTH Royal Institute of Technology (Sweden) and ScalaLife Competence Center coordinator.

REFERENCES

[1] https://wiki.egi.eu/w/images/5/5f/VT_CMMST_Proposal_v1.pdf

[2] A. Laganà, A. Costantini, A virtual research community for synergistic computing in Chemistry, Molecular and Materials Science and Technologies, Virt&l-comm 3 (2013)

[3] A. Costantini, O. Gervasi, C. Manuali, N. Faginas Lago, S. Rampino, A. Laganà, COMPCHEM: progress towards GEMS a Grid Empowered Molecular Simulator and beyond, Journal of Grid Computing 8(4), 571-586 (2010)

[4] A. Laganà, E. Garcia, A. Paladini, P. Casavecchia, N. Balucani, The last mile of molecular reaction dynamics virtual experiments: the case of the OH (N=1-10) + CO (j=0-3) → H + CO_2 reaction, Faraday Discussion of Chem. Soc. 157, 415 - 436 (2012)

[5] A. Laganà, A. Riganelli, O. Gervasi, On the structuring of the computational chemistry virtual organization COMPCHEM, Lecture Notes in Computer Science (ISSN:0302-9743) 3980, 665-674 (2006)

[6] A. Laganà, S. Crocchianti, N. Faginas Lago, A. Riganelli, C. Manuali, S. Schanze, From Computer Assisted to Grid Empowered Teaching and Learning Activities in Higher Chemistry Education in Innovative Methods in Teaching and Learning Chemistry in Higher Education, I. Eilks and B. Byers Eds, RSC Publishing (2009) p. 153-190 ; ISBN 978-1-84755-958-6

[7] C. Manuali, S. Rampino, A. Laganà, GRIF: A Grid Framework for a Web Service Approach to Reactive Scattering, Comp. Phys. Comm. 181, 1179-1185 (2010); C. Manuali, A. Laganà, GRIF: A New Collaborative Framework for a Web Service Approach to Grid Empowered Calculations, Future Generation of Computer Systems, 27(3), 315-318 (2011)

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

[9] Quantum Chemistry packages list:

http://en.wikipedia.org/wiki/List_of_quantum_chemistry_and_solid_state_physics_software

[10] GAMESS-US see http://www.msg.ameslab.gov/gamess/ M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S. Su, T.L. Windus, M. Dupuis, J.A. Montgomery, J. Comput. Chem. 14, 1347 (1993)

[11] http://www.gaussian.com/

[12] D. Skouteris, O. Gervasi, A. Laganà, Non-Born-Oppenheimer MCTDH calculations on the confined H_2^+ molecular ion, Chem. Phys. Letters 500 (1-3), 144-148 (2010)

[13] M. Valiev, E.J. Bylaska, N. Govind, K. Kowalski, T.P. Straatsma, H.J.J. van Dam, D. Wang, J. Nieplocha, E. Apra, T.L. Windus, W.A. de Jong, "NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations" Comput. Phys. Commun. 181, 1477 (2010)

[14] [Turbo] R. Ahlrichs; M. Bär; M. Häser; H. Horn; C. Kölmel, Electronic structure calculations on workstation computers: The program system Turbomole, *Chem. Phys. Lett.*, **162**(3), 165-169, (1989).

[15] [Orca] F. Neese, The ORCA program system, WIREs Comput. Mol. Sci. 2012, 2: 73–78

[16] D. Skouteris, J.F. Castillo, D.E. Manolopulos, ABC: a quantum reactive scattering program, Comp. Phys. Comm. 133, 128–135 (2000)

[17] U. Manthe, Direct calculations of reaction rates in Reaction and Molecular Dynamics, A. Laganà and A. Riganelli Eds, Springer Verlag BERLIN (ISBN 3-540-41202-6) (2000) p. 130

[18] M. Beck, A. Jakle, G. Worth, H.D. Meyer, The multiconfiguration timedependent Hartree (MCTDH) method: a highly efficient algorithm for propagating wavepackets, Phys. Rep. 324, 1–5 (2000)

[19] D. Skouteris, L. Pacifici, A. Laganà, Time dependent wavepacket calculations for the N(^{4}S) + N_2($^1\Sigma^+_g$) system on a LEPS surface: inelastic and reactive probabilities, Mol. Phys. (ISSN:0026-8976) 102(21-22), 2237-2248 (2004).

[20] M. Ceotto, S. Atahan, S. Shin, First principles semiclassical initial value representation molecular dynamics, Phys. Chem. Chem. Phys. 11, 3861-3867 (2009)

[21] M. Hankel, M., Smith, C. Sean, S. K Gray, G. G. Balint-Kurti, G. G. DIFFREALWAVE: A parallel real wavepacket code for the quantum mechanical calculation of reactive state-to-state differential cross sections in atom plus diatom collisions, Computer Physics Communications, 179 8, 569-578 (2008).

[22] venus96 Chemical Dynamics Software and Simulation System (CDSSIM System) https://cdssim.chem.ttu.edu/nav/htmlpages/licensemenu.jsp. Last seen May 2013

[23] W. Smith, T.R. Forester, DL POLY2: a general purpose parallel molecular dynamics simulation package, Journal of Molecular Graphics, Vol. 14 (3), 136-141 (1996)

[24] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. Shirts, J. Smith, P. Kasson, D. van der Spoel, B. Hess, E. Lindahl, GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit Bioinformatics (2013) 29 (7): 845-854, 2013 doi:10.1093/bioinformatics/btt055

[25] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten, Scalable molecular dynamics with NAMD, Journal of Computational Chemistry 26,1781-1802 (2005).

[26] O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, Journal of Computational Chemistry 31 (2010) 455-461

[27] CADDSuite: http://www.ballview.org/caddsuite

[28] Flexx: http://www.biosolveit.de/FlexX/

[29] Kevin J. Bowers, Edmond Chow, Huafeng Xu, Ron O. Dror, Michael P. Eastwood, Brent A. Gregersen, John L. Klepeis, István Kolossváry, Mark A. Moraes, Federico D. Sacerdoti, John K. Salmon, Yibing Shan, and David E. Shaw, "Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters," *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*, New York, NY: IEEE, 2006.

[30] GC3Pie website: http://code.google.com/p/gc3pie/. Last seen May 2013

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

[32] IGI portal website: https://portal.italiangrid.it/. Last seen May 2013

[33] In Silico Lab website: http://insilicolab.grid.cyfronet.pl/. Last seen May 2013

[34] MyGAMESS project: http://www.switch.ch/aaa/apps/projects/htdocs/detail/UZH.7

[35] S. Tasso, S. Pallottelli, M. Ferroni, R. Bastianini, A. Laganà, Taxonomy management in a Federation of Distributed Repositories: a chemistry use case, Lecture Notes Computer Science 7333, 358-370 (2012)

[36] Kepler website: https://kepler-project.org/. Last seen May 2013

[37] S. Rampino, A. Monari, S. Evangelisti, E. Rossi, A. Lagana', A priori modeling of chemical reactions on computational grid platforms: workflows and data models, Chemical Physics 398, 192-198 (2012)

[38] R. Dovesi, R. Orlando, B. Civalleri, C. Roetti, V. R. Saunders, and C. M. Zicovich-Wilson, Z. Kristallogr. 220, 571 (2005)

[39] Quantum Espresso webpage: http://www.quantum-espresso.org/

[40] A. Laganà, S. Crocchianti, G. Ochoa de Aspuru, R. Gargano, G.A. Parker, Parallel time

independent quantum calculations of atom diatom reactivity, Lecture Notes Ccomputer Science 1041, 361-370 (1995)
[41] A. Aguado, C. Tablero, and M. Paniagua, Comput. Phys. Comm. 134, 97 (2001)
[42] S. C. Farantos, POMULT: A Program for Computing Periodic Orbits in Hamiltonian Systems Based on Multiple Shooting Algorithms, Comp. Phys. Comm. 108, 240-258 (1998)
[43] S. C. Farantos, Periodic Orbits in Biological Molecules: Phase Space Structures and Selectivity in Alanine Dipeptide, J. Chem. Phys. 126(17),175101-175107 (2007).
[44] N. Balucani, P. Casavecchia, L. Banares, F.J. Aoiz, T. Gonzalez-Lezana, P. Honvault and J.M. Launay, J. Phys. Chem. A, 110, 817 (2006)
[45] http://www.daltonprogram.org
[46] http://www.scalalife.eu/discrete
[47] Ragni (private communication)….
[48] The CHIMERE chemistry-transport model. A multi-scale simulation. Institute Pierre-Simon Laplace, INERIS, LISA, C.N.R.S. (2004) http://euler.lmd.polytechnique.fr/chimere
[49] http://www.unidata.ucar.edu/software/netcdf
[50] M. Lattuati: (1997). Contribution a l'etude du bilan de l'ozone tropospherique a l'interface de l'Europe et de l'Atlantique Nord: modelisation lagrangienne et mesures en altitude. These de sciences, Universitè Paris 6, France (1997).
[51] E. Ranzi, A. Frassoldati, S. Granata, T. Faravelli. (2005) A wide range kinetic modeling study of pyrolysis, oxidation and combustion of heavy n-alkanes Ind. Eng. Chem. Res., vol. 44, pp. 5170-5183 ISSN: 0888-5885
[52] M. Capitelli Ed. Molecular Physics and Hypersonic Flows NATO ASI Series C. Dordrecht: Kluwer Academic Publishers 1996; 482: pp. 1-20.
[53] K. Karta, An investigation on personalized for web service, Honours Programme of the School of Computer Science and Software Engineering, University of Western Australia (2005)
[54] P. Kacsuk, Z. Farkas, M. Kozlovszky, G. Hermann, Á. Balasko, K. Karóczkai, and I. Márton, "WS-PGRADE/gUSE Generic DCI Gateway Framework for a Large Variety of User Communities", Journal of Grid Computing, vol. 10, no. 4, pp. 601 - 630, 2012
[55] SHIWA project website: http://www.shiwa-worklfow.eu/project. Last seen May 2013
[56] E. Rossi, S. Evangelisti, A. Laganà, A. Monari, S. Rampino, M. Verdicchio, K. Baldridge, G.L. Bendazzoli, S. Borini, R. Cimiraglia, C. Angeli, P. Kallay, H.P. Lüthi, K. Ruud, J. Sanchez-Marin, A. Scemama, P. Szalay, A. Tajti, Code Interoperability and Standard Data Formats in Quantum Chemistry and Quantum Dynamics: the Q5/D5cost Data Model, J. Comp. Chem. (in the press)
[57] EGI_EUDAT_PRACE pilot project: https://wiki.egi.eu/wiki/EGI_EUDAT_PRACE_collaboration