Virt&I-Comm.5.2014.15

Does the Chemistry, Molecular & Materials Sciences and Technologies community need a Competence Centre?

Alessandro Costantini[1], Antonio Laganà[2],
1. IGI-INFN, Perugia, Italy
2. Department of Chemistry, Biology and Biotechnologies
University of Perugia

1. Introduction

The European Grid Infrastructure (EGI) has recently established a Virtual Research Community (VRC) named Chemistry, Molecular and Materials Science & Technology (CMMST), as a follow up of the activities of the EGI-InSPIRE Virtual Team (VT) project "Towards a CMMST VRC" (see ref [1]) promoted by the COMPCHEM Virtual Organization (VO) in a joint endeavour with other VOs, computational Molecular Science research projects and the participation of some Associations and spinoffs.

The CMMST community is by definition a service community because its advances and developments in molecular sciences serve as a ground for assembling multi-scale applications aimed at reproducing the observable physical properties of realistic systems in materials, energy, chemistry, biology, medicine, etc. and supporting as well technological innovation.
One of the main objectives of CMMST is to capitalize on the large quantity of molecular science programs and packages either designed or developed (or even just used by its members) in order to build Universal Computational Engines enabling accurate ab initio molecular simulations of a large variety of innovative higher level of complexity applications. In order to progress in this, CMMST relies on an efficient opportunistic use of the grid platforms made available by the recent impressive progress of distributed computing hardware, middleware and applications. As a first step this means the dressing of the already implemented molecular simulators with appropriate workflows and their offering (through easy to use specialized portals) not only to the CMMST members but also to researchers, technologists and educators of other communities whose applications are based on the use of Molecular Sciences and Technologies for their applications.

In order to do this CMMST has proposed to EGI the assemblage of a subject specific Competence Centre (CC). The central point of the proposal is the adoption of a distributed computing model encompassing much more than the simple feature of aggregating a large amount of distributed compute resources for an opportunistic access by the users. The proposed CC aims, in fact, at adopting a truly collaborative/competitive (synergistic) model in which not only the barriers between compute platforms are lowered but also the difference between users and producers is weakened, programs designed and implemented for different research areas interoperate and collaboration is encouraged and rewarded in terms of its quality parameters to the end of fostering sustainability through a higher level of competitiveness.

2. The nature of the CMMST VRC

VRCs are groups of like-minded individuals and clusters of researchers organised by discipline or computational model. European Grid Infrastructure [2] VRCs [3] typically have an established

ISSN: 2279-8773

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 several 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 which are 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.

For the above mentioned reasons, several members of the Virtual Organization (VO) COMPCHEM [4] together with other EGI VOs (GAUSSIAN [5] and CHEM.VO.IBERGRID.EU [6]) with the support of some National Grid Infrastructures (like IGI [7], PL-Grid [8] and IBERGRID [9]), projects (like MoSGrid [10] and ScalaLife [11]), resource providers (like CINECA [12], CYFRONET [13], BSC-CNS [14], CSC [15]) have established the already mentioned CMMST VRC. The CMMST VRC relies also on strong interactions with the European Chemistry Thematic Network (ECTN) Association and the EUCHEMS Division of Computational Chemistry (DCC) whose characteristics are sketched in ref. [16].

Moreover, in order to achieve sustainability, CMMST has also established links with a cluster of academic spinoffs (MASTER-UP srl, Perugia, IT [17], Exact lab srl, Trieste, IT [18], Arctur d.o.o., Nova Gorica, SI [19], Polymechanon, Thessaloniki, GR [20], Krebs Inc, Gumpoldskirken, AT [21].

Such composite nature of CMMST is at the same time the strength and the weakness of its community. On one side, in fact, the evolution of computing technologies has been increasingly more successful to cope with heterogeneous platforms and to combine their specific virtues. As an example it is becoming an added value to network remotely accessible high-performance platforms (High Performance Computing, HPC) and local (departments, research groups, institutions, etc.) clustered machines (High Throughput Computing, HPC). After all, while the latter are tailored to the user needs and requirements and, as typical of platforms adopting opportunistic models, are highly adaptable to the user requests and assign available resources to the users requests as they come (virtually with no selective filters apart from some mitigation criteria) fostering so far design and development policies, the former use a centralized grant model that by assigning 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 are more suitable for production runs once compute time is obtained.

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 contrary to the local opportunistic model that is, instead, 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 Europe in PRACE [22]) 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 complementing of both models in a synergistic one is therefore an important added value that the CMMST community wishes to promote also for other reasons. There are, in fact, weakness of the mentioned models (when they are taken individually) that are turned into advantages in the

ISSN: 2279-8773

synergistic model adopted by CMMST. This is, indeed, the case of the large variety of in-house codes developed by the community members and/or by third bodies that may turn useful for the assemblage of multi-scale applications aiming at reproducing the already mentioned observable physical properties of realistic systems in materials, energy, chemistry, biology, biotechnologies, medicine, etc. starting from first principles and integrating different expertise. Such activities represent a solid support to research, innovation and development in a broad spectrum of fields (ranging from pure science to innovative technologies) of strong economic and social impact through a service/user oriented approach by basing the cooperative endeavour on:
1. the combined expertise and efforts of a large number of European experts in the field of molecular sciences and technologies;
2. the adoption (with appropriate adaptations) of high level ICT instruments and platforms.
3. the introduction of quality parameters to rank services provided and involved users.

In order to exploit these advantages, the CMMST VRC is planning to further develop the Grid Framework GriF [23] originally designed within the activities of the Perugia group of COMPCHEM. GriF is a Service Oriented Architecture (SOA) Collaborative Framework designed to facilitate the use of the Distributed Compute Infrastructure (DCI) by non specialists with a particular focus on the optimization of the selection of computing elements (based on Quality of Service (QoS) parameters worked out from the information provided at run time by the grid 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 QoS offered by the computing site and of the use made of the resources by the users.
GriF is accompanied by the introduction of a metrics aimed at evaluating the Quality of a User (QoU) and of his/her contributions of the community according to requirements set for that purpose by the community itself. Using the other tool GCreS [24] also originally designed within the activities of the Perugia group of COMPCHEM and to be further developed by the CMMST VRC, the work done by the users on behalf of the community will be rewarded by credits' assignment.

This is bound to foster a profound innovation in molecular simulations by paving the way for high performance ab initio exact treatments of both the electronic structure and nuclei dynamics of reactive processes of increasingly larger systems that are at present dealt using force field and classical mechanics (commonly referred to as molecular dynamics) techniques which ignore completely quantum effects which are of vital importance for some scientific and technological applications.
More details will be given later when discussing the development of the synergistic model;

3. The CC and the services of the CMMST members

In order to achieve the goal of meeting user requirements for an efficient access and use of high throughput and high performance computing resources, to the end of enabling the composition of higher level complexity applications through the sharing of hardware and software, developing a new collaborative model of carrying out research grounded on a quality evaluation of the work done for the community, the proposed CC will examine both technical and non-technical aspects related to the structuring of the VRC by following the evolution of technology, structure and organization of distributed computing. In practice a selection 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. A first list of people involved and of the programs and packages that are being considered for this purpose and for storage in the EGI AppDB, are those related to the

ISSN: 2279-8773

simulator (Grid Empowered Molecular Simulator) [25] as specified in the already quoted VT document.
Most of them are large packages and often contain several separate programs developed over many years that can be grouped as:

- 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).

At present, the ab initio electronic structure programs available to the CMMST VRC of relevance for the INTERACTION module are well known (*GAMESS-US* [26], *GAUSSIAN* [27], *NWChem* [28] *Turbomole* [29], *ORCA* [30]) as apparent from the quoted references. They are, in fact, apart from the in-house developed *NB-MCTDH* [31] one that can equally well deal with Non Born Oppenheimer cases and Born Oppenheimer ones, ab initio electronic structure packages (Hartree-Fock (HF), post Hartree-Fock, Density Functional Theory (DFT)), either open source or commercial software based on the Born Oppenheimer approximation.

The molecular dynamics codes available to the CMMST VRC and relevant to the DYNAMICS module are largely well known of the quantum type (like *ABC* [32], APH3D [33], *FLUSS-MCTDH* [34, 35], *RWAVEPR* [36], *DIFFREALWAVE* [37]) and of the classical type (like *VENUS96* [38], *DL_POLY* [39], *GROMACS* [40], *NAMD* [41], *AutoDock* [42], *CADDSuite* [43], *FlexX* [44] and *DESMOND* [45].

Such a rich articulation of the available data base of programs and packages provides a fertile ground for exploiting the already mentioned QoS selection on the compute platform (from personal systems to supercomputers) and the structuring and/or adoption of innovative workflows and portals. And even more fruitful is going to be the building of a suitable metrics for rewarding the active users in return for the work done on behalf of the community through a credit acquisition/redemption system. This is expected to foster a completely new level of accuracy of ab initio molecular simulations able to ground a sustainable Grid economy (generalizable to other communities) and a high impact on the society.

4 The longer term action of the CC and related partners

Thanks to the fact that the CMMST VRC gathers together molecular scientists and technologists belonging to a broad spectrum of application areas, the CMMST CC is bound to develop, as already mentioned, workflows, portals and other tools enhancing the perspective of assembling higher level of complexity applications, of designing and implementing simple and interoperable data standards and of being involved, as well, in extending their use in science education, technology training and market activities (especially with academic spin-offs and SMEs).

A list of EGI products useful for that purpose is already given in the above mentioned “Towards a CMMST VRC” document. In the followings, we single out in particular a list of instruments also tailored to fit such purpose: *GC3PIE* [46], *GriF* [47], the *IGI Portal* [48], *InSilico Lab* [49] *MyGAMESS* [50], *G-LOREP* [51], *KEPLER* [52].

ISSN: 2279-8773

The most strategic product for achieving the planned advanced purposes, however, is WS-PGRADE [53] (MTA-SZTAKI, Budapest (HU)) and the SCI-BUS gateway framework gUse (grid and cloud user support environment) that is permanently under improvement as an open source science gateway (SG) framework developed by the EU FP7 SCI-BUS project. It enables users to conveniently and easily access 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. Such framework can be used by National Grid Initiatives (NGIs) to support small user communities that cannot afford the development of 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, also the IGI Portal of a participating Institution integrates this technology among others (like DIRAC and an ad hoc designed Data Management). Produced by the same MTA-SZTAKI laboratory is the *SHIWA SIMULATION PLATFORM* [54] that provides a multi-systems workflow execution platform and interoperable solutions. 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, while several members of the CMMST community use WS-PGRADE workflows (MoSGrid) others use Kepler. By using the SSP they can mutually reuse the workflows developed by the other community.

In this perspective a proposal has been made to EGI aimed at establishing a roadmap for the actual implementation of the CMMST CC articulated into the following tasks.

Task 1: User support and training
This task is aimed at determining the success of the CC activities by integrating up the competences of the CMMST members belonging to different research areas, having different background and possessing different levels of expertise. This task will enhance the awareness of the community and make available the expertise (also in terms of software and technologies) necessary to the whole community through training events and tutorials.

Task 2: The consolidation of a Synergistic High Throughput High Performance Performance Universal Molecular Simulator
This task consists of a path-finding mini-project aimed at equipping the CMMST community with a reference version of the Grid Empowered Molecular Simulator. Such version leverages on tools and technologies developed within EGI-Inspire to the end of facilitating and take advantage from the interoperability between HTC and HPC calculations.

Task 3: The integration of the simulator into WS-PGRADE technology
This task consists of a light-house mini-project aimed at equipping the CMMST community with specific tools best suited for selecting resources and fostering collaborative work through the use of both portals and workflows. The associated enhancement of productivity will be fostered by the introduction of quality criteria (aimed at exploiting also in-kind contributions). The outcomes of the present task will contribute to the adoption of a quality based credit economy in which credits are awarded and redeemed.

Task 4: The adoption of data standards for the various components of the simulator
This task consists of a path-finding mini-project aimed at exploiting the strong collaborative frame of the enhanced version of the synergistic model developed in Task 3. In such version common data

ISSN: 2279-8773

and knowledge standard (even if only de facto) formats will be adopted. The typical CMMST de facto standard of the various modules of the simulator (electronic structure, molecular dynamics, statistical treatments, graphical rendering, etc.) will be extended for use in training, education and knowledge production.

Task 5: Dissemination and sustainability
This task will provide at the same time the dissemination of molecular sciences and of the distributed ICT technologies competences associated with the synergistic operational model. The CMMST-CC will try to extend such competences and services to other communities and SMEs in order to both enhance the transfer of science outcomes into innovative technologies and to improve sustainability.

The planned activities will imply the participation of the Organizations listed below in Table 1

TABLE 1: LIST OF THE INVOLVED ORGANIZATIONS

Target user communities (*for science-oriented CCs only*)	Chemistry, Molecular & Materials Science and Technology (CMMST) Virtual Research Community (VRC)
List of organizations representing the user communities (*for science-oriented CCs only*).	**EGI Virtual Organizations (VO)s** COMPCHEM (https://www3.compchem.unipg.it/compchem/), GAUSSIAN (https://indico.cern.ch/event/1504/session/131/contribution/216/material/slides/0.pdf), CHEM.VO.IBERGRID.EU (http://ibergrid.lip.pt/USP) **Consortia** INSTM (Italian National Institute of Materials Science and Technology, http://www.instm.it/en/instm.aspx) **National Grid Infrastructures** IGI (http://www.italiangrid.it/), PL-Grid (http://www.plgrid.pl/), IBERGRID (http://ibergrid.lip.pt/), NGI_HU (https://www.mgkk.hu/ngi_hu), MetaCentrum (http://www.metacentrum.cz/en/), GRNET (https://www.grnet.gr/) NGI-LT (http://mif.vu.lt/ngi/)

ISSN: 2279-8773

	European projects The MoSGrid (http://www.sci-bus.eu/mosgrid), ScalaLife (http://www.scalalife.eu/) **The resource providers** CINECA (ww.cineca.it), CYFRONET (http://www.cyfronet.krakow.pl/), BSC-CNS (http://www.bsc.es/), CSC (http://www.csc.fi/english) **Supporting European Chemistry Associations** European Chemistry Thematic Network (ECTN) Association (http://ectn-assoc.cpe.fr/) Division of Computational Chemistry (DCC, http://www.euchems.eu/divisions/computational-chemistry.html) of EUCHEMS **Related Spinoffs** MASTER-UP srl, Perugia, IT (*www.master-up.it/*), Krebs Inc, Gumpoldskirken, AT (http://www.krebs.company/234/), Exact lab srl, Trieste, IT (http://www.exact-lab.it/), Arctur d.o.o., Nova Gorica, SI (https://www.arctur.si), Polymechanon, Thessaloniki, GR (http://www.polymechanon.eu).

5 The key problems to solve

The key problems to be solved for the implementation of the CC project are:

The proposed simulator, though being a derivation of GEMS, is now going to be composed of software produced and/or maintained by research groups not belonging to the same virtual organization. In addition, the used software will be made of a large fraction of codes native of supercomputer environments and of middleware differing from those popular in EGI. This will require a huge effort of harmonization and standardization. Such effort will be in charge mainly to the coordinating institution that will have to identify technologies and services suitable to support the implementation of a synergistic and collaborative model among users and providers (both resources and technology). For that purpose a comprehensive support is going to be given as

ISSN: 2279-8773

deliverables produced during the EGI-InSPIRE Virtual Team (VT) project ”Towards a CMMST VRC”. The implementation of the first prototype HPC/HTC workflow of GEMS will focus, once more, on the exact treatment of gas phase elementary reactions by making use of the WS-PGRADE technology. WS-PGRADE will be used to develop the related workflow consisting in a procedure that distributes, depending on the chosen system and parameters, either on a HTC or on a HPC platform, the generation of an appropriate set of electronic structure calculations and then launches on them a set of strongly coupled calculations to be run either on a HTC or on a HPC platform (again depending on the chosen system and parameters). The chosen prototype application will be then transformed into a HPC/HTC pre-production workflow.

WS-PGRADE has not a built-in broker. This implies that the user 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 WMS for gLite, or a particular site for this (if there is no VO or broker in the VO, such as for HPC sites of XSEDE). Within the framework of the CC this will require a significant collaborative effort between SZTAKI and the coordinating Institution. Moreover, in order to ensure access to individual grant-based usage of the resources (typically the HPC ones) or adopt an alternative community grant type further modifications will have to be introduced into WS-PGRADE in order to allow it to deal with community access and robot certificates. WS-PGRADE accounting information will have to be modified in order to deal with collected data 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. This will be also an effort mainly in charge of SZTAKI that will 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.

To the end of bridging in a synergistic fashion HPC and HTC platforms and overcome so far 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) the functionalities of the improved version of GriF will be exploited to 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 HPC 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 contribution of MoSGrid as well as the goal of the agreement reached between EGI and XSEDE (https://www.xsede.org/) aimed at allowing to interoperate from the European Grid infrastructure on the supercomputer US network.

Essentially, starting from the work performed in Task 1 to enable a quality selection of the available hardware and software resources among those considered suitable for running the application

ISSN: 2279-8773

proposed, the present task aims at equipping the CMMST community with specific tools developed by other EGI internal and external partners. The adopted tools will target the offer of appropriate portals, manage workflows and allow dynamic allocation of resources depending on the requirements set by the users of the community. In particular the work will concentrate on adopting the WS-PGRADE web portal and workflow tools in order to distribute among the HTC and HPC platforms the various jobs. This will allow to embody the simulator into appropriate workflows in order to dynamically allocate the compute and storage elements, to aggregate the most efficient pieces of software and utilities so as to optimize the efficiency and accuracy of the computation. Moreover, a similar work will be carried out on data and knowledge sharing in distributed environments in order to single out the basic elements of how quality parameters can be used to activate a grid economy in terms of assignment and redemption of credits.

In order to identify the technologies and services suitable to support the task, the available tools specialized in service quality evaluation and credit assignment will be analysed and solutions for their integration as part of the prototypical HPC/HTC workflow will be transferred for integration in WS-PGRADE portals and tools for programs and compute resources selection. A similar transfer will be adopted for data and knowledge management by focusing mainly on training and educational purposes. These community specific features integrated in the WS-PGRADE web portal will be offered, after a trial period led by the members of the VRC, to the other communities. Further effort will be paid for the definition of specific key service indicators (and related parameters) for the evaluation (through the assignment of credits) of those services provided to the community by its members. To do this, a specific user levels schema aimed at identifying and evaluate specific competences and roles have to be negotiated and adopted. The outcome of the present task is expected to bring a significant contribution in the assignment of the in-kind contributions coming from the community members through which a cost and business models can be developed.

The possibility within the synergistic distributed model of choosing the platform better fitted for retrieving/storing programs and data for his/her applications provides the CC CMMST with the advantage of providing tools for better combining different pieces of software when assembling complex realistic applications using GEMS (or any other collaborative simulator) and undertake more ambitious research projects. Such possibility of building workflows of shared programs has, for example, set stringent requirements for the choice of proper (de facto) standards of data in quantum chemistry and quantum dynamics (like the Q5Cost and D5Cost [55]). This will be also a motivation for the development of tools (out of 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). In this way the CMMST community is able to start collaborative projects with XSEDE. Such activity is expected to generate and improve a set of tools and recommendations that the CMMST community can use to integrate its CMMST workflows.

The distributed repository mini-project aims at extending the tools adopted by the CMMST-CC for chaining computational programs of the simulator (let us call it HACE-MAS) to molecular data-bases and knowledge management systems of molecular structures and processes. Related knowledge is created by CMMST members either using the HACE-MAS simulator in a massive distributed way or collecting from existing published data or data bases. For them appropriate representations obeying to at least de facto standards of format will be worked out. The target use of such data and knowledge is education and training purposes and the use of a synergistic model (of

ISSN: 2279-8773

users and providers of both resources and technology) allows to deal with molecular data and knowledge contents by following the approach adopted for selection and quality evaluation of compute and storage elements.

One of the activities of this task will be the generation and implementation of the requirements for the dealt data and knowledge. In particular the proper metadata will be identified in order to ensure both the interoperability of the data among different packages of the simulator (especially between the modules of electronic structure and dynamics) and the support to the development and retrieval of the information for comparison of the quality of sources of different quality. This will pave the way to a knowledge exploitation by targeting both education (of senior and younger generations) and publication of the materials (including their prompt re-usability) produced by the different members of the community. This will foster the activation of a virtuous circle of continuously improved versions of relevant materials and their efficient use for training and learning. This will also foster the development of editorial initiatives facilitating the re-use of published materials.

Finally, the CMMST-CC will target the promotion at the same time of molecular sciences and of distributed ICT technologies competences dealt in a synergistic fashion by the community members. In this task, the CMMST-CC will aim at disseminating the project activities and establish sustainability measures. This will focus in particular on extending competences and services to other communities and linking with SMEs for science transfer into innovative technologies.

These measures will target in a diffuse way not only all the members of the CMMST VRC but also related projects and Associations. They will also target some SMEs operating in the field of ICT based education and training technologies recently networked under the umbrella of the Virtual Education Community standing committee of the ECTN Association. The goal is the development of an operational modality aimed at ensuring sustainability to the Consortium beyond the lifetime of the present project. Particular effort will be paid for establishing links with spinoffs and SMEs operating in science transfer into innovative technologies. To this end specific channels facilitating the export of services to external communities and enterprises (with particular relevance for environmental agencies, research based education institutions, pharmaceutical companies, molecular biology and biotechnologies, etc.) will be investigated for activation. The strategy that will be elaborated will focus on providing a financial basis for establishing links with some start-ups and spin-offs in particular by 1) promoting a Consortium with chemistry and ICT based companies and institutions; 2) establishing a user panel aimed at advising the Consortium on the recent requirements and interests of the society; 3) developing a sustainability model.

Conclusions

In the present document the motivations and the guiding ideas for establishing a CMMST Competence Centre are discussed. The document depicts the present scenario of computational resources and their usage by analysing the characteristics of the existing compute time allocation model as opposed to the synergistic one to be adopted by the proposed CMMST CC. The document enumerates also the technical and non technical aspects of such CC fostering 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 the advanced usage of QoS and QoU for introducing community economy based on a credit system will produce. On this ground, the CMMST CC will be able not only to share hardware and software to run out higher level of

ISSN: 2279-8773

complexity application but will act to support users intending to provide services. This will have a twofold effect:

a) enhance collaboration to the end of developing real-like simulations using stable versions of their programs as components to be offered as a service to the other users;
b) encourage competition to the end of making the data produced the most accessible possible and the services provided the most efficient possible so as to exploit their use

in a clearly sustainable way allowing also the re-use of the produced data in science as well as in knowledge handling for education, training, dissemination and the development of big data science to which chemistry is ancillary.

REFERENCES

[1] https://wiki.egi.eu/wiki/Towards_a_CMMST_VRC
[2] EGI, http://www.egi.eu/
[3] http://www.egi.eu/community/vrcs/
[4] https://www3.compchem.unipg.it/compchem/
[5] https://indico.cern.ch/event/1504/session/131/contribution/216/material/slides/0.pdf
[6] http://ibergrid.lip.pt/USP
[7] http://www.italiangrid.it/
[8] http://www.plgrid.pl/
[9] http://ibergrid.lip.pt/
[10] http://www.sci-bus.eu/mosgrid
[11] http://www.scalalife.eu/
[12] ww.cineca.it
[13] http://www.cyfronet.krakow.pl/
[14] http://www.bsc.es/
[15] http://www.csc.fi/english
[16] http://www.euchems.eu/divisions/computational-chemistry.html
[17] www.master-up.it/
[18] http://www.exact-lab.it/
[19] https://www.arctur.si
[20] http://www.polymechanon.eu
[21] http://www.krebs.company/234/
[22] http://www.prace-ri.eu/
[23] C. Manuali, A. Lagana', GRIF: A New Collaborative Framework for a Web Service Approach to Grid Empowered Calculations, Future Generation of Computer Systems, 27(3), 315-318 (2011)
[24] C. Manuali, A. Lagana', A Grid Credit System Empowering Virtual Research Communities Sustainability, Lecture Notes Computer Science, 6784, 397-411 (2011)
[25] 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)
[26] 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)

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

[28] 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)

[29] [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).

ISSN: 2279-8773

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

[31] 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)

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

[33] A. Lagana', G.A. Parker, R.T Pack,
Li + FH Reactive cross sections from J=0 accurate quantum reactivity,
J. Chem. Phys. (ISSN:0021-9606) 99, 2269-2270 (1993).

[34] 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

[35] 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)

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

[37] 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).

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

[39] 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)

[40] 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

[41] 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).

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

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

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

[45] 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.

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

[47] C. Manuali, A. Laganà GRIF: A New Collaborative Framework for a Web Service Approach to Grid

ISSN: 2279-8773

Empowered Calculations Future Generation of Computer Systems, 27(3), 315-318 (2011)

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

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

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

[51] 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)

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

[53] 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

[54] SHIWA project website: http://www.shiwa-worklfow.eu/project. Last seen May 2013

[55] 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. Journal of Computational Chemistry 35(8): Code Interoperability 611-621 (2014).

ISSN: 2279-8773