

- 0 (March 6): Antonio and Mariusz discussion at the Catania user forum**
- 1 (.....): Mariusz web declaration of intents to set an SSC**
- 2 (March 29): Antonio list of items to be dealt in the draft of the proposal**
- 3 (April 3?): Mariusz a preliminary draft**
- 4 (April 6): Antonio draft version MMST-SSC-draft-vo**
- 5 (April 8): Oral presentation of Ludek Matyska at the Prague COST meeting**
Antonio presentation of MMST-SSC-draft-vo
- 6 (April 16): Antonio revision MMST-SSC-draft-vo-1**
- 7 (April 23): Antonio revision MMST-SSC-draft-vo-2**
- 8 (April 29): Antonio revision MMST-SSC-draft-vo-3**
- 9 (April 30): Peter Szalay contribution**
- 10 (May 6): Jan contribution slides**
- 11 (May 8): Stefano Cozzini, Fermin Huartecontribution**
- 12 (June 2): Antonio Revision MMST-SSC-draft-vo-4**
(June 3): Silvio Migliori contribution
- 13 (June 11): Elda Rossi, Antti Pursula contributions**
Antonio Revision MMST-SSC-draft-vo-5
- 14 (June 13-18): Stavros contributions**
- 15 (June 18): Antonio Revision MMST-SSC-draft-vo-6**
- 16 (June 21): Antonio new draft CMST-SSC-prop-1**

DRAFT PROPOSAL TO BE COMPLETED AND REVISED

Proposal full title: Chemical and Material Sciences and Technologies SSC
Proposal acronym: CMST-SSC
Type of funding scheme: Combination of
Work programme topics addressed: INFO-...
Name of the coordinating person:
List of participants:

Part. no.	Participant organisation name	Part. short name	Country
1	University of Perugia, Dept of Chemistry	UNIPGCHIM	Italy
2	Cyfronet	CYFRONET	Poland
3	National Center for Biomolecular Research	NCBR	Czech Rep.
4	IT Center for Science ltd	CSC	Finland
5	Ente nazionale energie alterantive	ENEA	Italy
6	Consorzio Interuniversitario CINECA	CINECA	Italy
7	Theoretical chemistry and computational grid applications	UZH	Switzerland
8	Foundation for Research and Technology Hellas, Inst. Electronic structure and lasers	FORTH	Greece
9	Democritos, ICTP, Trieste	DEMOCRIT	Italy
10	University of Barcelona	UBQP	Spain

Section 1. Scope

1.1. Concept and objectives

BACKGROUND AND PARTNERS OF THE PROPOSAL (to be completed in national projects and VOs and revised)

At present, advanced modelling and simulations based on multi-scale and multi-physics approaches starting from nanoscale level are an integral part of the process of scientific discovery in the field of molecular and material sciences which are at the hearth of any research and innovative applications in chemistry, physics and life sciences. In these fields, the large amount and variety of programs and suites of codes developed and implemented leads inevitably to the exploitation of both the collaboration on the network and of the massive use of concurrent and distributed computing. This has led, indeed, in the past to the establishing of seminal Virtual Organizations (VO)s gathering some grid users of the chemistry and material sciences and technologies (CMST) community to the end of better structuring their programs for the grid and at the same time specialize a segment of the grid to better fit to the requests of their applications.

For this reason the scientists of those VOs and of other laboratories active in promoting the use of the grid for CMST computational purposes find themselves naturally aligned with the scopes of the European Grid Initiative (EGI) project (of which they share the strategic view and are willing to support the efforts) aimed at coordinating at European level the action of national grid promoting the establishing of infrastructures allowing the collaborative utilization of hardware, software and knowledge. They are therefore ready to implement the EGI policy of making the User Communities and the related VOs key components of this action by consolidating and enhancing their grid infrastructure and activities through the establishing of some Specialized Support Centres (SSC). After all, the proposal of establishing of a CMST SSC represents only the last step of a process initiated within several European countries (among which is worth mentioning GRID.IT (Italy), RBNED1KNFP, MIUR FIRB,PL-Grid, MetaCentrum) in which the researchers operating in the field of molecular and material sciences and technologies were exposed to the exploitation of the Grid innovative features to better match their computational needs. At the same time the proposal of establishing a CMST SSC is, indeed, also the result of a pan European endeavour aimed at establishing collaborations (COST Actions D23 and D37,) and preparing a common operative ground (EGEE III) for gathering the computational scientists of molecular and material sciences around the grid infrastructure.

More specifically the proposal of establishing a CMST SSC was directly incubated by the Computational Chemistry (CC) Cluster of Excellence of EGEE III (in which the COMPCHEM, GAUSSIAN, TURBOMOLE (.....please supply details of the Lab in charge of that....) and VOCE VO were operating) and by other EGEE activities in which other VOs of the field were involved (like ENEA,). This means that the proposed CMST SSC will count from the very beginning on a solid group of research laboratories (called hereafter “support laboratories” and listed in Table 1a) which, in addition to being already operating on the European grid by contributing with hardware, software, expertise and efforts, carry out management activities and supply support to grid operations by coordinating running VOs, sustaining grid efforts through their own skilled

the necessary high level middleware tools and targeted web portals/dashboards with potential to deliver solutions customized according to the general needs of the users.

Table 1a: CMST SSC Supporters

N	ID	Cy	Name	Local Coordinators	NGI supp
1	UNIPGCHIM	IT	University of Perugia, Department of Chemistry, Perugia	Antonio Lagana', Osvaldo Gervasi, Francesco Tarantelli	IGI
2	CYFRONET	PL	CYFRONET	Mariusz Sterzel	
3	NCBR	CZ	National Center for Biomolecular Research	Jan Kmunicek	
4	DEMOCRIT	IT	Democritos, ICTP, Trieste	Stefano Cozzini	IGI
5	ENEA	IT	ENEA computational laboratory for nanotechnologies and materials	Silvio Migliori	IGI

The support laboratories

UNIPGCHIM: a cluster of research groups led by the Department of Chemistry of the University of Perugia (Italy) and belonging to the Department of Chemistry itself, to the Department of Mathematics and Computer science, to the Department of Physics, to CNISM (a consortium for material science) and to INFN (the national institute for nuclear Physics) all operating inside the Campus of the University of Perugia and having ties with the University spinoff Master-up and the cultural association ICSA

CYFRONET

NCBR

DEMOCRIT

ENEA

However, efforts will be provided also by other laboratories (listed in Table 1b and hereafter called grid applications providers) which have already matured (as part of their engagement in VO activities) high skills in restructuring computer codes for concurrent execution, enabling applications for the grid using the related middleware, developing tools for a user friendly exploitation of the grid, designing massive computational campaigns to carry out complex simulations for producing innovation. To some extent these grid applications providers already do (or are committed to do in the near future especially if the proposal is approved) contribute to the hardware platform sometimes even with specialized machines.

Table 1b: CMST SSC grid enabled software providers

N	ID	Cy	Name	Local	NGI
---	----	----	------	-------	-----

			structure and lasers		
2	UBQF	ES	University of Barcelona, Department of Physical Chemistry, Barcelona	Fermin Huarte	
3	UZH	CH	Theoretical chemistry and computational grid applications	Kim Baldrige	

These laboratories have the following characteristics:

FORTH

UBQF

UZH

Other laboratories already committed to act as provider (mainly of CMST know how though with limited competences in grid computing) are those listed in Table 1c and hereafter called high performance computing providers. These high performance computing providers despite their limited competences in grid computing are high potential contributors to the SSC because of their high level activities in parallel computing and in managing high performance computing resources or even large scale facilities.

Table 1c: CMST SSC high performance computing providers

N	ID	Cy	Name	Local Coordinators	NGI supp
1	CSC	CH	IT Center for Science ltd	Per Oster	
2	CINECA	IT	Department of Information and Knowledge Management	Elda Rossi	

The high performance computing providers laboratories are

CSC

CINECA

THE USER COMMUNITY

The functionalities of the CMST SSC guaranteed by the above mentioned supporter and provider laboratories of Tables 1a, 1b and 1c will become critical to the end of attracting new people (hereafter called users) to the SSC and to the end of allowing them to gain added value and to become able to provide themselves expertise and resources to the community. This is a key feature of the proposed CMST SSC. As a matter of fact, there is a large number of laboratories which have already manifested their interest in the institution of a CMST SSC since they have been already exposed to some extent to the already mentioned national or European grid initiatives. A (by defect) list of these laboratories (hereafter called user laboratories) is given in Table 2. User laboratories which have in particular occasion manifested their interest in concurrent computing and related implementations will be specific object of the CMST SSC action. Users will be divided into a “passive” and an “active” subset. Passive users are those who will utilize, after being registered, the user-friendly software and applications offered by the partner VOs. The status of passive user is meant to be valid for a limited period of time aimed at clarifying whether the use of the grid should be of commercial type (in this case users will be encouraged to develop start-ups utilizing the grid services of the SSC) or at allowing people to get trained to scale up to the level of active user.

To become user a laboratory must apply to any of the VOs belonging to the SSC while to become a VO a group of laboratories must apply to the SSC. To facilitate the acquisition of new users from the CMST area the SSC will pay particular efforts to the end of attracting resource providers offering also hardware, software and data particularly suited to the High Performance needs of the CMST community and encouraging its members to become increasingly involved in providing grid services better meeting the needs of the industrial, pharmaceutical, new material, life processes, environmental operators.

Table 2: CMST SSC institutional users

N	ID	Cy	Name	Local Coordinator
	UNIVIE		University of Vienna	Hans Lischka
	UNIVIE		University of Vienna	Wilfried Gansterer
	RCJ		Research Center Jülich	Thomas Müller
	HUB		Humboldt University Berlin	Vlasta Bonačić-Koutecký
	CASP		Czech Academy of Sciences, Prague	Jiri Pittner
	UNIHEID		University of Heidelberg	Hans Dieter Meyer
	UNIHEID		University of Heidelberg	Horst Köppel
	UNIFI		University of Pisa	Maurizio Persico
	ELUB		Eötvös Loránd University, Budapest	Péter G. Szalay
	ETHZ1		ETH Zürich	Hans Peter Lüthi
	UNIBERNE		University of Berne	Samuel Leutwyler
	RBIZ		Rudjer Bošković Institute, Zagreb	Zvonimir Maksić
	IMIPB		IMIP, Bari	Domenico Bruno
	UNIINSBR		University of Innsbruck	Maximilian Berger
	UNIVIE		University of Vienna	Matthias Ruckebauer
	ENSCP		ENSCP Chimie ParisTech 11, Paris	Carlo Adamo
	UNIVIT		Dept. Phys. Chem, Univ. Basq. Country, Vittoria	Ernesto Garcia
	UNIREN		Dept. of Phys., Univ. Rennes, Rennes	Jean M. Launay

	SNSC		Swiss National Supercomputing Centre, Manno "	Mario Valle
	CESGA		Centro de Supercomputaci3n de Galicia, Santiago de Compostela "	Aurelio Rodriguez
	UNIBO		University of Bologna	Gian Luigi Bendazzoli
	UNITOUL		University of Toulouse "	Stefano Evangelisti
	UNIBODA		University of Budapest	Mihaly Kállay
	UNITROMS		University of Tromsø	Kenneth Ruud
	UNIFER		University of Ferrara	Renzo Cimiraglia
	UNILILLE		University of Lille	Valerie Vallet
	UNIVAL		University of Valencia	Josè Sanchez-Marin
	ZIB		Zuse-Institut, Berlin	Thomas Steinke
	ETHZ		ETH Zürich	Martin Brändle
	FAU		FAU Erlangen-Nürnberg	Tim Clark
	UNICAMBR		University of Cambridge	Peter Murray Rust
	ICL		Imperial College London	Henry Rzepa
	UNISEVIL		Universidad de Sevilla	Antonio M. Márquez Cruz
	UNIZUR		University of Zurich	Juerg Hutter
	UNICAMBR		University of Cambridge	Michiel Sprik
	UNIGHENT		University of Ghent	Michel Waroquier
	UNIPMC		University Pierre et Marie Curie, Paris	Rodolphe Vuilleumier
	SCSC		Swiss Center for Scientific Computing	Peter Kunszt
	ETHZ		ETH Zurich	Michele Parrinello
	UNIAMST		University of Amsterdam	Evert-Jan Meijer
	SIB		SIB; Ludwig Inst. for Cancer Research	Olivier Michielin
	SISSA		SISSA, Trieste	Paulo Carloni
	UNIBA		Department of Chemistry, University of Bari	Fabrizio Esposito
	UNIZAGR		Department of Physics, University of Zagreb	Goranka Bilalbegovic
	UNIPG		Department of Physics, University of Perugia	Massimiliano Alvioli

			Greece	
	UNINA		University of Naples, Italy	Michele Pavone
	UNIPD		University of Padoa, Italy	Laura Orian, Gianpietro Sella
	UNIMI		University of Milan, Bicocca, Italy	Ugo Cosentino
	UNIFI		University of Pisa, Italy	Alberto Coduti, Alessandro Ferretti
	UNINA		University of Naples, Italy	Mauro Causà, Luisa Caracciolo
	UNIPD		University of Padoa, Italy	Andrea Vittadini
	UNIFI		University of Pisa, Italy	Susanna Monti
	UNIMO		University of Modena, Italy	Cristina Menziani
	NCBR		National Centre for Biomolecular Research	Jan Kmunicek, Jaroslav Koca
	LL		Loschmidt Laboratories	Jan Kmunicek, Jiri Damborsky
	UNIPAN	HU	Huniversity of Pannonia, Vezprem	Lajos Fodor
	CSIC	ES	CSIC	Octavio Roncero

MAIN FEATURES AND ACTIVITIES OF THE PROPOSED CMST SSC

Particular efforts will also be spent, to ensure a governance to the process, to validate middleware services, to improve interoperability, to support application development and porting by assisting users, maintaining a front desks and ensuring documentation, training and dissemination within the CMST community. In other words the proposed SSC will take care of maintaining and developing the common tools and services needed by the CMST community and VOs by establishing specific science gateways to the evolution of distributed computing. The CMST SSC will also provide “umbrella” services useful to foster synergies and collaborative projects aimed at ensuring a continuity line to related scientific areas. This will make the SSC:

- open to the joining of additional partners upon request
- have clear governance rules through representatives of member laboratories and supporting NGIS
- have a broad European basis that is coordinated at European level
- enhance the aims of EGI
- promote efficient and productive uses of the grid and actively contribute to increasing its HW and SW resources
- develop tools for its sustainability
- legitimately represent the CMST community at national and European level
- have a clear roadmap articulated in appropriate milestones
- be committed to name relevant contact points for EGI processes, environmental operators.

Accordingly, the proposed CMST SSC will concentrate its efforts on the following topics:

- e) spread the knowledge produced by the members of the CMST community
- f) foster interactions, relationships and resource sharing with other VOs and SSCs
- g) increase the number of people and entities using the CMST SSC services
- h) promote the use of computer resources and Grid technologies by the CMST community (the support to the various lines of CMST computational applications will be discussed separately)
- i) collaborate with EGI operations and middleware development groups to tailor a design and an implementation of grid tools and services best suited to the necessity of the CMST community
- j) support for and help in the design of scientific projects and funding applications through a valorisation of the services provided
- k) regulate its internal life in a way appropriate to foster cooperation.

MOTIVATIONS AND PLANS OF THE PROPOSED SSC

CMST has established itself as a high ranking community in the current utilization of EGEE III cpu resources (immediately after High Energy and Life Science grid communities) despite the marginal participation to the allocation of EU resources. As a matter of fact, while the CMST research area has little benefited from the European funding to the Grid it has been able to build mainly on the efforts of the member research laboratories an appreciable amount of initiatives. In terms of computing power, for example, published figures quote that only about 8 percent (approximately 8000) of the CPUs belonging to the EGEE infrastructure do support CMST computations while, typically, from 40 to 70% of the capacity of the large scale computing facilities used for scientific purposes outside EGEE III are accredited to be utilized for CMST computations. The CMST SSC, in fact, will bring together on the side of scientific competences in the area a truly critical mass of expertise in chemistry and molecular and material sciences and technologies while on the side of computer skill will ground all this on a high level of competence in scientific computing and management of scientific knowledge in CMST together with the ability of managing grid segments, e-infrastructures, scientific visualisation and web-based portals is concerned. All this finds its rationale on a set of motivations aimed at changing the way computational sciences are currently carried out.

MAIN MOTIVATIONS

The central motivation for instituting a CMST SSC is that of spurring discovery and transforming the way simulations are performed, data is handled, and codes are interfaced through collaborative endeavours on the Grid. For this purpose the various activities will be aimed at:

Supporting the members of the CMST community in implementing and using an environment of shared hardware and software resources on the grid

This is the basic motivation for instituting an SSC in the field of CMST. The availability of an ad hoc designed environment suited for an appropriate implementation of the large variety of computer codes and algorithms designed by or for the members of the molecular and material science community is, in fact, a primary need of the research groups operating in this field. This would not only allow to lower the technological barriers to the use of the grid associated with the rapid evolution of computer technologies (which remain hard to master for researchers of other fields) but it would also push the CMST scientists to develop even more complex (and at the same time more realistic) grid-enabled collaborative simulations. This will motivate the development of specific

and packages of support to scientific computing and the acquisition to the grid platform of high performance computing nodes with higher sizes of memory and various levels of parallelism as well as the adoption of appropriate parallelization tools.

Adopting models and standards for molecular and material knowledge allowing a reuse of quantum chemistry data and codes for an easy connection of different packages

A further motivation for establishing a CMST SSC is the need for developing models and standards for Quantum Chemistry (QC) data to support collaborative work. Typically this is the case of ab initio electronic structure outputs of several QC packages (each of which has its own format) which need to be assembled together or to be provided as an input for other programs in a row or in parallel in complex CMST simulations. There are, indeed, proposals quoted in the literature for developing such standards. A proposed standard for ab initio QC data tailored on the DALTON package and meant to be ported on the grid has been already worked out as a result of the EU collaborative initiative COST. Such product, named QC5, and related procedures were specifically designed for porting on grid environments with the motivation that they will make the assemblage of repositories, tools and workflows easier and will advance, as well, efforts in prediction, design and analysis using multiple codes, across multiple disciplines, and for an open user base. An increasing availability of QC digital data (interfaced with users and codes) will have a profound impact in both the quality and rate of discovery of the CMST community.

Implementing on the grid libraries of codes relevant to CMST as web services for use by the community

User friendliness, while necessary to motivate the establishing of an SSC, might not be enough to make grid approaches popular. It is, in fact, truly important mainly for people who are willing to develop their own procedures and simulations by introducing significant (either in terms of conceptual innovation or in terms of gathering together non negligible amounts of different pieces of software) advances in methodology. However, in the CMST community there is already a large fraction of people making massive use of well established (commercial and non commercial) packages. Traditionally these researchers would try to get a copy of the package from an official library and implement it locally by surmounting the difficulties of a new installation. A further important motivation for instituting an SSC can be, therefore, that of providing these researchers (most of the people listed in Table 2 do indeed belong to that fraction not to mention the many people who do not even appear in that list since they use validated versions of important packages as black boxes (passive users in our definition of the SSC members)) with a version of those codes implemented as web services. Typical examples of this type of packages for the CMST field are GAUSSIAN (or equivalent packages for ab initio calculations) or DL_POLY (or equivalent packages for Molecular dynamics calculations).

Designing and developing tools for evaluating quality of services (QoS) and quality of users (QoU) in CMST web services

A further motivation for establishing a CMST SSC is that it can become the reference body and the driving force of the community by setting a fair system of objective and subjective evaluation means of members and users activities. Evaluation of members will have to be implemented by measuring the quality of service (QoS) they are able to provide not only to the other members but also to the users and third parties by using objective and subjective parameters. This will have to represent a first step towards sustainability. Evaluation of users will have to be implemented by measuring their quality (QoU) in using grid resources and in providing feedbacks and indications on how to extend and improve the use of the grid.

Build a credit system as part for constructing a grid economy able to reward people contributing to the SSC sustainability.

term return. This may be used also to promote the achievements of social interest and at large of more general interest. Credits will be redeemed through a better access to services, research funding and other returns that the SSC will establish as congruent with its activities.

EXPECTED RESULTS

Expected results:

- A collection of CMST programs geographically distributed on heterogeneous platforms to be used either on an individual or on a shared basis (**resource GRID**)
- A central (web based) interface for accessing and composing some computer programs and applications as web services (**web service**)
- A workflow manager finalized to the handling of a grid empowered simulation (**simulation workflow**)
- A prototype QoS and QoU evaluator of the services and the users of the CMST SSC (**quality evaluators**)
- A system of credit award and redemption to support SSC sustainability (**credit system**)
- A repository for QC data complying with the model and standards of QC5 (**QC5 standard**)
- An agreement with the owners of some widely CMST programs for supporting shared usage and/or common formats (**collaboration**)

RELATION WITH THE CALL (to be modified)

Relation to the call:

- The proposal is part of an attempt to build *a broad European collaborative on scientific infrastructure*. Therefore, even if applications are concerned with a specific scientific sector (Chemistry and material Science) most of the problems have a common background with those of other disciplines and are clearly addressed to the ICT problem posed by grid platforms
- *... able to be easily federated with other knowledge infrastructures*. The spirit of the proposal is to federate the results obtained with those of different experiences which need for example
- *Shifting the balance away from research based only on publications*: results reproducibility and programs interoperability is one of the key features of any scientific work. A researcher must be able to reproduce and verify a published result in an independent way. Since programs vary over the time and usually only final results are kept, reproducibility is often difficult to be guaranteed. Some data is stored as supplementary material to publications but this is organized in a non-standard fashion with an undesirable dependency on the publishers policy to provide access. A publicly accessible standardized repository will help by archiving intermediate data such as the output of the procedure used to obtain the results. Moreover, the standardized repository boosts the possibility of cooperative work in several computational applications including research based teaching at higher education level [refelchem]
- *This activity will provide an integrated set of services exploiting the middleware and grid capabilities to federate data*: available middleware and computing infrastructure will be exploited, a demonstration using the two main European infrastructures will be done
- *Supporting the deployment of standardized mechanisms to run, store, archive and interpret scientific data*: standardized mechanisms to handle programs and data are central in this proposal, that aims to interoperability and to the sharing of the results produced by various

earning/redemption. The ability of providing services usable by other people and to pose requests to the system on line with a best use of available resources are taken as the foundation of the community sustainability.

1.2. Progress beyond the state-of-the-art (to be further developed)

As briefly sketched above the proposed CMST SSC is meant to push scientific computing to take a significant leap ahead in the direction of distributed computing. This proposal is, in fact, strongly based on the past experience and projects of the partners while representing a significant step forward in the exploitation of the grid by the CMST research groups. In the first part of this paragraph we discuss the present state of the art (both at organizational and infrastructure level). In the second part the paragraph examines the advances with respect to the first part.

Grid infrastructures and grid applications for Chemistry

Grid infrastructures are of course a central point in this proposal. We intend to develop in this respect interoperability of codes and libraries for the CMST area. We intend also to tackle the problem of better qualify the computing nodes of the Grid for CMST calculations by introducing a wider spectrum of machines and software available on the grid by opening windows of access for the CMST community on different networks like that of large scale facilities

- **EGEE** (Enabling Grids for E-science) - www.eu-egee.org/
UNIPGCHIM is running a VO for this infrastructure
-

There have been some initiative regarding grid applications for Chemistry that is worthwhile describing,.....

COMPCHEM activity

The activity related to the implementation on the grid of the software tools and programs exploits the advantage of using the already established EGEE production Grid segment available to the Virtual organization COMPCHEM. The COMPCHEM VO counts at present about 60 users and more than 8000 cpus. The activities of COMPCHEM are at present largely centred on dynamical calculations with the core set of programs devoted to exact QD calculations (ABC, RWAVEPR, MCTDH, APH3D) of the reactive properties of atom diatom systems. However, the same programs have been extended to the approximate calculations for more complex systems (like four atom reactions, ion flows in carbon nanotubes, etc.) when only a subset of degrees of freedom are treated exactly while the others are dealt by approximate means. COMPCHEM has implemented for run on the production grid also some classical dynamics programs like ABCtraj, VENUS and DL_POLY or semiclassical ones like SCIVR in addition to some other programs dealing with the calculation of periodic orbits or carrying out ab-initio calculations.

The most ambitious goal of COMPCHEM is, however, the design and implementation of the components of Grid enabled molecular simulator GEMS out of all the mentioned codes (and possibly also many other ones) after establishing efficient communication among them. In particular, standardized data format for communication and repository are needed to link the first module of the simulator (INTERACTION) containing ab-initio programs of the already mentioned

developments of the project is FITTING, a procedure that fits to the ab-initio values a suitable functional form giving an analytical representation of the potential energy surface of the investigated system. The module of GEMS of specific interest for the project is DYNAMICS that is designed to perform either classical mechanics or quantum mechanics dynamical calculations. In particular, the project will focus on the ABC, and RWAVEPR programs which have been built to carry out exact quantum calculations of the S matrix of reactive scattering problems. These programs can either use a functional representation of the potential energy surface or run directly using the output of the ab-initio calculations if this is represented in a proper format. The last module of GEMS is OBSERVABLES that has been modified to make it possible to carry out the proper statistical averaging either of the outcome of ab initio calculations or of any other module of GEMS to derive measurable properties.

GAUSSIAN activity

VOCE activity

QC5

Another important point of this project is the **user-friendly aspect**. HDF5 comes with a library (for Fortran and C) for managing the data structures. This was not enough user oriented, so a specialized library, **Q5Cost**, was written to manage the specific QC data format. Q5Cost is a high level library where subroutines refer to “chemical” concepts and objects, orbitals, integrals, and atoms, not to datasets, attributes and groups. Q5Cost has been conceived as a Fortran 90 library. Due to its high degree of organization, however, it is now possible to automatically build bindings for other languages (already available C, C++, Python, Fortran 77) in addition to the “native” Fortran 90 programming language. Since HDF5 is “self described” and contains metadata describing the actual content, in order to maintain this characteristic for Q5Cost, some specific tools were realised: they make it possible to inspect the content of a given file (Q5dump), to easily navigate and modify it (Q5edit), to check its consistency (Q5validate).

several programs are fully or partially integrated with Q5Cost via specific wrappers: general or commercial programs like DALTON, GamessUS, MOLCAS, COLUMBUS, OpenBabel, Molekel; in house codes like MRCC (Budapest), FullCI (Bologna), CASDI and EXCI (Toulouse). Some of these codes are also directly integrated to Q5Cost by including the Q5cost library in the program (FullCI and DALTON at present).

CECAM initiatives

The interest for computational chemistry and material related grid enabled application is also witnessed by the interest devoted to this issue by important European scientific organizations like CECAM (Centre Européen de Calcul Atomique et Moléculaire), an European organization supported by an important network of widespread national research Organization. Its main activity is focused on the organization and support of tutorials and workshop which enable the scientific communities involved in computational sciences to keep in touch and develop occasions of exchange of experiences leading to a general increase in multidisciplinary experiences.

These problems have been central to some of the workshops organised by CECAM in the last years. Most of the partners of the present projects have regularly attended in the past CECAM workshops

COST D37 project

The COST Action D37 (“GridChem”) was approved in March of 2006 and put into force in May 2006 for a four year period ending 2010. The Action consists of five Working Groups, and is lead by Hans P. Lüthi (ETH Zürich). Its mission statement reads as follows:

“The main objective of the Action is to facilitate the creation and use of distributed computing infrastructures (‘Grids’) in chemistry with the goal of bringing computer modelling and simulation in chemistry to new frontiers in complexity and to a new regime of time-to-solution.

This will stimulate innovation in the creation and handling of chemical knowledge. The areas of application cover traditional chemistry, materials science, molecular biology and environmental chemistry. The presence of chemistry on the Grid will also have an impact on the development of middleware and the design and layout of Grid infrastructures.

Enabling and accelerating the transition of researchers to the ‘infrastructure of computational science of the 21st century’ will make European computational chemistry more competitive.”

The five Working Groups of the Action consist of six to eight members each representing a total of 19 European countries. GridChem covers a broad spectrum of expertise which ranges from applied information science, Grid computing, quantum chemistry, all the way to various implementations of molecular dynamics. Also, in GridChem you find a number of application code owners or people closely related to a code development team (ADF, COLUMBUS, DALTON, DIRAC, MOLCAS, etc.), which ensures a very high degree of dissemination of the work performed within this Action. In its first two years, the Action has created a number of results, which, without the interaction of experts from different areas, would have been impossible to achieve (see also Progress Report posted at the action web site). There is more information available at www.cost.esf.org.

ECTN activity

The activities of the CMST area are of interest also to ECTN (Euroeapan Chemistry Thematic Network) and its association ECTNA which are actively operating to implement on the Grid educational activities in molecular and material sciences. ECTNA has recently launched a project of creating a grid based virtual organization and has recently decided to ove. Yet, they have recently converged on the adoption of standards to manage chemical non textual knowledge for research based teaching and learning. This activity was started in the year 2000 by the Chemical education using multimedia working group which has operated in collaboration with the working group ELCHEM of the D23 (Metachem) action of the COST Chemistry European Initiative. More recently in the year 2006 this activity has converged into those of the Innovative Approaches to Chemistry Teaching of ECTN. The key goal is to establish a continental wide Learning Management System (LMS) in Europe by designing and implementing locally some teaching and learning units to be used in a distributed way in Europe.

A key need of that experience is the possibility of accessing ab initio electronic structure data formatted in a way that they are reusable and retrievable independently of the approach followed for the implementation of the teaching and learning software. In other words any progress made in the direction of standardizing QC data would be of great benefit for ECTN activities.

Beyond the state of the art

The CMST project will exploit

Grid applications for Chemistry

accessible from the main European grid infrastructures. It will be open source, available to the whole community, easy to use, able to encourage the collaborative working among different scientific domains:

- Design tools for an efficient use (storage/retrieval) of the repository, and for the graphical representation and visualization of the result. The existing interface between Q5Cost and the MOLEKEL package will be generalized to the complete data model.
- In order to permit an efficient data-mining in the distributed repository, a centralized database containing metadata and other key information about the available distributed data

Bibliography

1.3. Methodology to achieve the objectives of the project

The objectives of the project will be achieved through a coordinated set of working packages (WPs) that are representing coherent activities under the responsibility of one of the partners (WP leader). The WPs belong to one of the three main areas requested for this type of project:

- JRAs (Joint Research Activities): three WPs focusing on three main areas of science. The aim is to describe the data quantities used in each specific area by creating a coordinated data model. For one of the area (WP1: Wavefunction-based Quantum Chemistry) a great part of the work has already be accomplished and will serve as a model to the other two sectors (WP2: Density Functional Theory and WP3: Quantum Dynamics)
- SAs: Service Activities – three WPs focusing on the set-up and operation of the service infrastructure that is the goal of the project. WP4 is about the global Data Model and takes care in a modular way of the different scientific domains; WP5 is focused on the repository itself and the WP6 is about an integrated set of tools, that will make the repository easy to be used. It is also responsibility of WP6 the day by day support of the service during the project life-time, as well as the design of a plan for assuring the service follow-up after the project.
- NAs: network activities – two WPs assuring the coordination activities with the leading experts that are not part of the project. WP7 aims to maintain strict links with European (and extra-European) programs' developers in order to guarantee a consensus on the data model, the compatibility of the programs with the DAMM-Sim platform and the maintenance of the “translation” wrappers. Also relationship with the reference e-infrastructures (DEISA, EGEE and NGS in primis) will be coordinated by this WP. WP8 will maintain the contacts with the scientific community for disseminating the data model and the DAMM-Sim platform and also to take care of its suggestions in the service set-up.

The complete list of WPs is reported in the following. In parenthesis is reported the type of activity – JRA, SA, NA - and the partner responsible for that WP:

1. **WP1 (JRA -):** WQC: Implementation and support for utility and application software on the grid
2. **WP2 (JRA -):** DFT: Development of models for reuse and connection of data to connect different packages
3. **WP3 (JRA -):** Restructure libraries and validated codes as web services
4. **WP4 (SA -):** Design Quality evaluation criteria and tools for Services and users
5. **WP5 (SA -):** Establish a field grid economy through a credit system
6. **WP6 (SA -):** Implementation of a distributed repository of CMST data.
7. **WP7 (SA -):** Tools and Workflow
Design and implementation of *tools* for the analysis of the results of the multi-scale simulations, and *workflow* for the incorporation of the various computational steps.
8. **WP8 (NA -):** Linking Establishing links with European codes' owners, production of wrappers to integrate the different programs, definition of a strategy for ensuring their maintainability.
9. **WP9 (NA -):** Dissemination: Establishing relationships with the European scientific community for supporting the data model and DAMM-Sim design. Dissemination of the outcome of the project,
10. **WP10 (.....):** Management of the project; tracking of progress, expenditure, deliverables, contact with the EC.

RESOURCES

Obviously, further investments of the CMST research laboratories into complementing the existing hardware, basic software and computational tools will be a high priority task of the proposed SSC and will be highly rewarding in terms of promoting grid computing. This could be straightforwardly achieved by making available more CPUs of the appropriate type to CMST computations. As we plan to build the CMST SSC as a vertical specialized support centre dealing with various research tasks of the domain, we expect that this will in any case occur thanks to the fact that some of its members will be going to further contribute to provide also part of a platform for mutual exchange of related knowledge/solutions and to mediate the access to cutting edge computational chemistry tools. As shown by Table 3 the most involved laboratories shall provide additional computer power to the Grid. This will constitute an additional infrastructure enabling a seamless solving of computational CMST tasks as well as testing new methods and/or approaches.

Table 3: COMMITTED HARDWARE

1	UNIPG	IT	20 CPU 4 core 9 Intel Pentium 4 core 17 Math & Inf ... Physics Dpt ... INFN-PG	Alessandro Costantini	
2	CYFRONET	PL		Mariusz Sterzel	
3	NCBR	CZ		Jan Kmunicek	
5	UBQF	ES		Fermin Huarte	
4	FORTH	GR		Stavros Farantos	
6	SISSA	IT		Stefano Cozzini	
7	ETH	CH		Kim Baldridge	
9	ENEA	IT	3000 cpu (shared)	Silvio Migliori	
10	CINECA	IT		Elda Rossi	

Another commitment is concerned with the development of specific software and high level tools (portals, workflows, databases etc.) or getting access to the use of high-level tools (like Bazaar). It will also enable to dynamically request computing and/or storage resources for the amount of time needed by the user (as when necessary for scientific projects or when needed for accessing specific resources) for CMST usage on the Grid. As a matter of fact, we believe that the establishing of a CMST SSC will be fundamental in supporting the related community at gaining access to computational resources at all levels (including the national one).

However the real central asset of the proposed SSC is the bunch of CMST applications that several laboratories have implemented on the grid for usage by other groups and laboratories. This is a typical best practice of the CMST community (in the past these were offered by the most popular scientific libraries (see for example the QCPE, the CPC, etc. ones). At present, the list of the applications which have been promised for shared use by the community is the one given in Table 4

Table 4 Committed applications for shared use¹

#	Application Name	Country	VO
1	ABC	Italy, Hungary	Compchem
2	MCTDH	Spain	Compchem
3	FLUSS	Spain	Compchem
4	VENUS96	Italy, Spain	Compchem
5	RWAVEPR	Italy	Compchem,
6	GAMESS	Italy, Czech Republic	Compchem, VOCE
7	DL_POLY	Italy	Compchem
8	NAMD	Italy/Czech Republic	Compchem/VOCE
9	GAUSSIAN	Poland	Gaussian

12	CPMD	Czech Republic	VOCE
13	GROMACS	Czech Republic, Italy	VOCE, Enea
14	NWChem	Poland	Gaussian
15	CPMD	Italy	... ENEA?
16	CP2K	Italy	... Enea
17	PCGAMESS	Italy	... Enea

Some of these programs are already implemented on the grid for individual usage and are being analysed for their implementation for shared usage. It is also important to point out that some of the above listed software are among them alternative (either because of the method or because of the implementing laboratory). This is already meant to be the strategy to be adopted by some workflows like that of the so called Grid empowered Molecular simulator in which the determination of the electronic structure of fairly small molecular systems using *ab initio* techniques, the construction of the interaction of medium and large size molecular systems using various *ab initio* or semiempirical methods bearing different levels of accuracy, the integration of quantum scattering equations for light-matter (or matter-matter) small systems or subsystems, the calculation of the dynamics and kinetics of large molecular systems, etc.) are believed to be usable in a sequence or in an alternative.

This will also lead to the assemblage of highly complex simulators based on nanolevel approaches and extremely useful for innovative applications (like the modelling of combustion, of secondary pollutant production, of innovative materials, of biological processes, of pharmaceutical activities, of structure properties relationships, etc) using the former as building blocks shown in Table 5.

Table 5 Applications for personal use²

#	Application Name	Country	VO
1	CHIMERE	Italy	Compchem
2	ABC	Italy	Compchem
3	MCTDH	Spain	Compchem
4	FLUSS	Spain	Compchem
5	VENUS96	Italy	Compchem
6	RWAVEPR	Italy	Compchem,
7	GAMESS	Italy	Compchem, VOCE
8	DL_POLY	Italy	Compchem
9	COLUMBUS	Italy	Compchem
10	NAMD	Italy, Poland, Czech Republic	Compchem/VOCE
11	SC-IVR	Italy	Compchem
12	GAUSSIAN	Poland	Gaussian
13	CPMD	Italy	...
14	NWChem	Poland	Gaussian
15	CPMD	Italy	... ENEA?
16	CP2K	Italy	... Enea
17	PCGAMESS	Italy	... Enea

16	GROMACS	Czech Republic	VOCE
17	NWChem	Poland	Gaussian
18	ABCD	Hungary	COMPChem
19	VENUS 88	Hungary	Compchem
20	ABCspli	Hungary	Compchem

The availability of these composable highly complex application will turn out to be of extreme value for the rest of the community.

STRUCTURE - The proposed organization of the SSC, including the relationships between institutes, persons, Virtual Organizations, EGI.org, and the NGIs.

The structure of CMST SSC as proposed below is based on the organization scheme suggested for SSC by EGI.

- a) Local coordinator – Each institution or research group member of the SSC should nominate a local coordinator (as well as a deputy). Local coordinators have the responsibility of keeping close contacts with the corresponding NGI, follow the evolution of the Grid in the country and spread inside the institution information coming from the NGI. Local coordinators form the SSC Management Board (MB) that is responsible for the governance of the SSC and for its evolution from the strategic point of view. The MB will be structured in a way that will keep the decision-making process as smooth as possible.
- b) SSC Chair – The MB will nominate an SSC Chair who will coordinate its activities and will represent it in all instances (including the EGI User Forum). The Chair will interact with the EGI council, the EGI.org administration middleware coordinators and all other SSC coordinators. The Chair will also have to provide information about the activities of the CMST community/SSC to external parties and disseminate at the same time information from external parties within CMST community
- c) To make interaction and cooperation among the members and VOs of the SSC as well as the decision making processes of the MB effective and as smooth as possible, the following bodies will be nominated by the MB:

Planning Board (PB) is the restricted board appointed by the MB to develop strategic initiatives and make recommendations to the MB about future developments. In CMST this is also the body taking care of fostering collaborations and managing the credit system to reward activities carried out on behalf of the community.

Front Desk (FD) is the technical unit responsible for several activities concerning consulting (for example direct interaction with application developers to get their application(s) running on the grid infrastructure), integration of CMST community resources with the grid infrastructure or assistance for application porting including integration of grid services necessarily to utilize the application in grid environment. Front Desk can also be used to spread information about the SSC among its members, to offer information about the membership to NGIs or consortia taking care of aspects relevant to the SSC.

User's Support (US) is the technical unit responsible for User support. We can distinguish

FUNDING - The required funding for the cluster along with a plan for securing that funding.

A first estimate of the funding required for the CMST SSC to operate is strictly related to the evaluations of the manpower necessary to properly operate it. Most of the manpower needs not to be centralized in a single place. The actual structure and composition of expenses, however, will have to be discussed in more detail once the SSC will be in operation.

Effort estimation – approximately 14 FTE

- The SSC Chair and its deputy(s) – 1.5-2 FTE
- Front Desk – 2-2.5 FTE
- Users Supporting teams
 - o Direct User Support – 2.5-3 FTE
 - o Documentation (specific to CMST SSC) coordination – 0.5 FTE
 - o Technical Support – 2-3 FTE
 - o GPE – database creation, maintenance and application porting – 2.5-3 FTE

The core of SSC will be formed from CESNET, CYFRONET, UNIPG, PARSC and FORTH. A more detailed splitting of funds and responsibilities concerning each institute will be discussed among the partners and presented around September this year. Also the list of institutes/partners presented above is subject to discussion and will be finalized in autumn.

EVOLUTION TO SSC - A plan for the evolution from the current cluster/team organization to the proposed organization, including time scales for the major steps in the evolution.

The Laboratories of the EGEE III CC cluster are well suited to migrate to a Specialized Support Centre EGI. To our knowledge also a significant fraction of the laboratories belonging to the COST CMST Action D37 is a suitable candidate for the migration. Other CMST research laboratories from various parts of Europe (like France, Croatia, Turkey, etc.) are moving some steps forward to join the CC cluster in making this move. The proposed timetable is as follows

Month	Step
April 2009	Collect information about possible candidates and their NGIs. Since many NGIs are still being shaped, the gathering of preliminary information is crucial in order to understand how Grid infrastructures evolve in each country. CMST laboratories should also be encouraged to take active part to the construction of the related NGIs by providing useful information about how to operate on the Grid. This can be done by providing a set of user requirements and use-cases, both to EGI and to NGIs.
May 2009	Collect information about CMST research groups (if any) involved in Grid-related activities but not members of the CC cluster.
June 2009	Make an inventory of all currently active CMST-related VOs, both registered and unregistered, and collect the relevant set of parameters. The term “CMST-related VO” indicates VOs directly managed within the CC cluster or a third-party VO supporting (in terms of shared resources and services) to research groups of the CC cluster. Parameters of interest for each VO are:

	<ul style="list-style-type: none"> ▪ Number of computing “slots” available to the VO. ▪ Amount of disk storage available to the VO. ▪ Total amount of CPU consumed by the VO. ▪ Total amount of storage used by the VO.
July 2009	Write the document describing the migration plan of the CC cluster to EGI detailing the steps to be undertaken for the construction of the CMST SSC.
September 2009	Detailed Work Plan for SSC. Final number of “core” partners.
October 2009	The agreed structure of the CMST SSC is put in place and its managerial bodies nominated, namely: <ul style="list-style-type: none"> ▪ the LC (Local Coordinators) for each Institutes/Research Group ▪ the MB (Management Board) formed by all LCs ▪ the Chair of the SSC
November 2009	Create the technical bodies of the SSC, namely the PB, the FD and the US.
December 2009	Hold the first SSC general meeting. The main focus of the meeting will be put on these topics: <ul style="list-style-type: none"> ▪ Interactions with NGIs. Verify that all CMST SSC partners contacted their respective NGIs and collected all necessary information on their structure and on regional/local Grid infrastructure. Verify that relevant information (CMST requirements and use-cases) have been provided to NGIs. ▪ CMST-related VOs. Verify that VOs whose management falls under the responsibility of the CMST SSC are fully operative. Verify that enough support comes from other VOs in order to guarantee that CMST final users can fully exploit the EGI Grid infrastructure. ▪ Grid middleware. Verify that the evolution of the EGI Grid middleware (especially in terms of deployed tools and services) takes into account all requirements coming from CMST applications. ▪ Final check on the SSC structure, namely on its bodies (MB, PB, TS, FD) and on interactions with central EGI bodies.
April 2010	Get ready for the transition from EGEE-III to EGI by taking care that corrective actions for the issues highlighted during the first general meeting have been undertaken.