ROSCOE

Robust Scientific Communities for EGI

EC Call: INFRA-2010-1.2.3: Virtual Research Communities

Project Name
Robust Scientific Communities for EGI (ROSCOE)
http://roscoe-project.eu

Project Contact
Charles LOOMIS
charles.loomis@cern.ch
CNRS/LAL (Orsay, France)

Estimated Project Start Date and Duration
The project will tentatively start 1 May 2010 and will run for 36 months.

Project Abstract

Context: Over the last decade, grid infrastructures have become indispensible scientific platforms, facilitating European scientific research from a diverse set of disciplines. The move from project-based organizations to sustainable ones will reassure researchers of their long-term viability and thereby increase their use and impact. The European Grid Infrastructure (EGI) will federate current cluster-based grids into a sustainable European grid while ensuring continuity from the existing grid infrastructures.
Aim of the Project: Sustainable funding levels will not permit EGI alone to fully support all of the grid's users and communities. The user communities themselves must fill the gap. This project aims to create robust scientific communities that are self-supporting and self-sustaining. These Specialized Support Centers (SSCs) in EGI jargon will be long-lived Virtual Research Communities that will coordinate their grid activities, safeguard the community's expertise, and provide specialized support.
ROSCOE targets seven scientific areas: High Energy Physics (HEP), Life Sciences (LS), Computational Chemistry & Material Science Technology (CCMST), Grid Observatory (GO), Complexity Science (CS), Photon Science (PS), and Humanities (H), including both large, mature communities and new, strategic scientific disciplines.
Work Program: As required for I3 projects, the work program is divided into Networking Activities, Service Activities and Joint Research Activities. The Networking Activities include the project management, SSC coordination, dissemination, and training. These activities aim to expand the user community of each SSC, to ensure efficient communication within the community, and to make users aware of grid functionality and best practices.
The Service Activities include user support, scientific gateways, and targeted application porting. The user support will augment the standard support services with domain-specific knowledge. The scientific gateways will form the hub of the community providing information pertinent to the community and in many cases providing direct access to grid services and community data sets. The porting activity will target analysis frameworks, common libraries, or high-profile applications that will have a large impact on the use and visibility of the grid for the community.
The Joint Research Activities will revolve around topics of interest to several different communities such as tools (HEP, LS, CS, PS), molecular dynamics code and studies

(CCMST, LS, PS), data mining of complex systems (CS, GO, HEP, H), large-scale data management (HEP, PS), and next generation data access and management (all areas).
Expected Impacts: The most difficult aspects of creating SSCs for EGI will be the social and political ones; the SSC coordination activities will directly tackle these problems and facilitate the transition to robust scientific communities on the grid. At the same time, the project will expand the reach of the grid by increasing the number of users, improving their grasp of grid technologies, and easing their use of the grid through its comprehensive work plan. The project will positively impact the multidisciplinary research already carried out in the seven targeted communities and similarly impact collaboration and cooperation between them.

Project Partners
The institutes involved in ROSCOE cover a broad range of different scientific disciplines. In total there are 57 institutes involved from 16 countries. The partners include two large industrial companies and one SME. All of the partners have been asked to contact their local NGI representatives; in addition all of the EGI council members have been made aware of this proposal.

High-Energy Physics

High Energy Physics (HEP) is one of the world's most cohesive large-scale international disciplines. Many researchers on today's flagship projects, such as those at the LHC at CERN, have worked on previous projects at the same or other laboratories. Many are or can be expected to continue their work on future projects and/or their design, such as the FAIR facility (an ESFRI project) or the International Linear Collider (ILC) project. HEP also has a long tradition of sharing its technology, an example is the CERN Program Library built up over several decades and installed at all major physics institutes worldwide. Furthermore, HEP pioneered the use of virtual organizations that cut across institutional and national boundaries, driven by the need to form functional collaborations around large-scale experiments. These types of organization, still a novelty in many areas, are now widely used in many forms in HEP, centered both on experiments but also on cross-cutting themes such as computing (e.g. WLCG, the worldwide LHC Computing Grid, and the HEPiX international forum for IT support staff at HEP institutes). The largest HEP communities today are gathered around the LHC experiments and the WLCG project. It is therefore natural that the primary area of work of the proposed SSC is oriented towards this activity whilst still addressing the needs of small to medium experiments worldwide as well as future projects. This will be done in close collaboration with existing structures, such as WLCG and HEPiX, in which members of the foreseen SSC are deeply involved. The LHC experimental program, and hence the need for an associated computing service, is foreseen to last well over a decade and hence sustainability is a cornerstone of a HEP SSC. This will be guaranteed by building on the experience gained over many decades in maintaining long-term international projects as well as the above-mentioned structures. Key institutes involved include the WLCG Tier0 (CERN) and Tier1 sites worldwide (ASGC/Taipei, BNL/US, CNAF/Italy, FNAL/US, FZK/Germany, IN2P3/France, NDGF/Nordic countries, NIKHEF-SARA/NL, PIC/Spain, RAL/UK, TRIUMF/Canada) as well as well over 100 Tier2 and Tier3 sites (other major national laboratories, including DESY/Germany, and many Universities and other research institutes).

The timescales of today's HEP projects, from conception and design through to construction and exploitation, are typically measured in decades, with the first proposals for a large hadron facility in the then, still-to-be-constructed LEP tunnel being made at the end of the 1970s. The first LHC computing proposals were made well over one decade ago and the LHC exploitation phase is expected to last well beyond 2020. The main tasks of the proposed HEP SSC, covering not only the LHC community but also others that plan to exploit the same basic computing technology, will need to continue throughout this period, even if significant changes in some elements of the environment must be foreseen over such a long period. The existing support effort—funded partially through European DataGrid and EGEE projects, including important long-term contributions from both Italy and the UK—has already demonstrated the potential for long-term sustainability. Beyond the foreseen funding period, it is expected that the main institutes involved would continue to provide support as part of their baseline contributions to the associated projects, with additional manpower being provided through the corresponding experiment collaborations (i.e. ALICE, ATLAS, CMS and LHCb for the LHC program).

Life Science

The SSC will leverage the work of several European projects (EGEE, EMBRACE, eLico, e-NMR, Health-e-Child, VLe etc.) to provide services for accessing the resources of the National Grid Initiatives federated in the European Grid Initiative to the research communities in molecular biology, medical imaging, drug discovery and next generation sequencing. Through the involvement of key European institutes and associations, it will foster the adoption of grids and the use of EGI resources by the Research Infrastructures that will map the field of life sciences in the coming years. The European Molecular Biology Laboratory (EMBL) is one of the EIROFORUM members and will participate through the European Bioinformatics Institute (EBI) to Life Science SSC activities.

Long-term plans for Life Science SSC build upon the involvement of key stakeholders in the definition, design and operation of services for the communities. These partners were chosen because of their roles and responsibilities in ESFRIs but also within their research communities and in the National Grid Initiatives. The sustainability of the SSC's Scientific Gateways is foreseen as follows:

- Operation of the Scientific Gateways for the specifically targeted ESFRIs (ELIXIR, INSTRUCT) will be supported by these infrastructures once they kick off.
- The Scientific Gateways for medical grid and drug discovery applications will be operated by the HealthGrid association.
- The EMBnet network of life sciences laboratories will provide the framework for supporting the scientific gateways for molecular biology in addition to the ELIXIR one.

Moreover the SSC's events will be integrated into the existing cycles of EMBnet and HealthGrid conferences and will continue after the end of ROSCOE.

Computational Chemistry and Material Science Technology

The areas of science covered by Computational Chemistry and Materials Sciences and Technologies are rather scattered. They include a large variety of research projects targeted to build molecular based simulations of natural phenomena and to design on silicon new substances and materials. This has produced along the years a vast amount of dedicated suites of programs and software tools which have made the CCMST community not only a key user of large scale computer facilities and ad hoc developed commercial software packages but also an important player for the in-house assemblage of computer clusters and development of computer programs. Moreover the produced codes have become increasingly used both by the CCMST community members as such but also by the members of other communities when trying to provide a rationale to phenomena and processes occurring in biological cells, living organisms, the earth and the universe. Most of these codes and packages handle CCMST problems starting from first principles. As a matter of fact, they often carry out *ab initio* computations of the field on which atoms and molecules move and then assemble physical observables based on the outcome of microscopic events occurring in a virtual nanometric world. Other CCMST codes and packages follow the more phenomenological approach and manage molecular (theoretical and experimental) data as an ingredient of information processing. In both cases the CCMST community is a heavy consumer of computing power and is a fast growing user of computing grids. For this reason in a few years the CCMST community has gathered together in some key virtual organizations bearing specific requirements like GAUSSIAN and TURBOMOLE (to enable commercial codes on the grid) and COMPCHEM or VOCE (to enable scientific advances to develop based on cooperation of distributed expertise and knowledge exchange driven scientific advance). As a matter of fact the CCMST community has become in a few years one of the most active players of the scientific grid-computing arena.

The creation of a CCMST SSC will boost the activities of spurring discovery through collaborative computational endeavors started within the EGEE projects and the COST international initiatives. It will also bring the European Grid Initiative closer to the market of scientific know how and technology transfer allowing it to tackle significantly more complex applications.

As to the first objective, the CCMST SSC will provide the existing virtual organizations in the field with a backbone framework allowing the common exploitation of the available hardware, software, and expertise. The framework will be based on quality of service and quality of user parameters and will develop an internal economy fostering significant scientific advances.

As to the second objective, the systematic exploitation of the potential of the grid made possible by the CCMST SSC will allow computational chemistry methods to gain in efficiency so as to develop realistic simulations of highly complex systems leading to a speed up the whole process of introducing new materials at all levels: design, synthesis, and advanced testing.

The combination of the two objectives will not only spread grid computing virtually to all chemical laboratories but it will also make sustainability at reach in a reasonable in short time so that it is that partners of proposed SSC will support the CCMST community even beyond the end of the project.

Grid Observatory

The usage of the SSC data and services is, by definition of the SSC, open to the whole scientific community. Beyond that, the GO SSC will act as a catalyst for developing synergies at the European and global levels, between scientific communities that have had, so far, limited opportunities to interact. Although we note that the Computer Science and engineering community have no international body comparable to CERN, ESA, or even the large biomedical collaborations. The only reference institution at this scale that could play the role of end-user (at least for the engineering aspect) is EGI itself. The recognition of the GO by the CS community is thus structurally bound to be more fine-grained, through actual usage by reputable institution or projects (national or EU), on one hand, and general scientific reconnaissance of the GO active participants on the other hand. A special emphasis will be put on the cross-fertilization of autonomic computing on one hand and grid research and engineering on the other. It must be noted that the Autonomic Computing community strongly centered in the US, with the NSF Centre for Autonomic Computing and a strong involvement of industry (notably IBM). In turn, the AC community might become a bridge to the TeraGrid community. The GO SSC is naturally willing to contribute to interactions with PRACE if opportunities appear.

The GO SSC is intended to be a stable entity whose primary goal is safeguarding and publishing datasets for the long term and providing stable analysis tools. Consequently, GO SSC has to evolve towards a permanent structure and define a sustainable financing model. However, the activity is much younger than all other scientific SSCs (the corresponding EGEE cluster has been created only within EGEE-III) and the SSC is still in it ramp-up phase. Thus it requires initial development funding, and has to invent a permanent structure and a governance model in the course of its existence.

The GO SSC must have support from, on one hand stakeholders involved in actual production, such as some NGIs and EGI, and on the other hand research institutes not presently involved in EGI, but prospective users of the GO data and services. The French NGI will provide the bulk of the hardware resource and participate in the operation tasks. The other NGIs are expected to contribute to the acquisition task under the general operation scheme, thus will not have to provide dedicated human resources. Interpretation from operation experts is a requirement for adding value to these data and will be bootstrapped by the constitution of a network of experts. The GO SSC requires interaction with EGI and UMD: at the operational level, in order to keep pace with the general development of the infrastructure, software monitoring resources, and operational issues, and at the conceptual level, to evolve to a sustainable set of services.

Photon Science

Currently, photon sciences are seeing a dramatic and dynamic enhancement of the experimental opportunities that demonstrates the great impact and scientific potential of the light sources for most areas of fundamental and applied sciences. The following table shows the beam lines (BL) hosted by the partners of the Photon Science SSC.

	Facilities	Services	Users / year Petabyte / yr	Communities	ESFRI Projects
	31 BL at Doris3 14 BL at Petra3 6 BL at FLASH	All IT Services, including e.g. AAA, IDM, dCache, Grid Services	~ 5000 ~ 5.0	Biology, Chemistry, Material Science, Surface Science, Arts, Drug Design, ...	EuroFEL Euro. XFEL
	25 BL 1 fermi@elettra	Computing IDM, AAA Virtual Control Room	~ 4000 ~ 0.5	Biology, Chemistry, Material Science ...	EuroFEL
	7 BL at Doris3 3 BL at Petra3	Crystallization, Computing, DM, Remote Access, Analysis Framework	~ 2000 ~ 2.0	Biology, Chemistry, Drug Design, ...	INSTRUCT
	51 BL	Computing, AAA, DM, Remote Access	~ 5000 ~ 2.0	Biology, Chemistry, Material Science, Paleontology, ...	ESRFup
	User Research Group (MX)	Crystallization, Analysis	~ 0.002	Protein Crystallography	INSTRUCT
	27 BL at Diamond 28 BL at ISIS	All IT services including IDM, Meta-Data, HPC ...	~ 5000 ~ 2.0	Biology, Chemistry, Material Science, Paleontology, ...	ESRFup PRACE Euro. XFEL
	5 BL	Full Experiment support, incl. DM, Storage/Archive, Detector Development	10-100	Biology, Chemistry, Material Science, ...	Euro. XFEL
Total	188 Beamlines		> 20.000 20-100		

Through the Photon Science (PS) SSC experiments from various scientific disciplines, tens of thousands researchers from Europe and worldwide will be supported. This exceptionally large number of user and research activities potentially benefiting from the PS SSC effort implies a tremendous impact and outreach.

The European light sources support experiments from various scientific disciplines with tens of thousands researchers worldwide. An average light source hosts typically 5000 users per year and many users visit more than facility often utilizing a series of different instruments. Only roughly 20% are actually long-term users, so that each European light source typically supports 10,000-15,000 users over 3 years, which renders support an eternally ongoing effort.

The members of the Photon Science SSC are essentially all integrated into projects of the current ESFRI roadmap: DESY and ELETTRA through participation in EuroFEL; EMBL and ITQB through the INSTRUCT project, which focuses on key biomedical questions and environment problems with be a strong involvement of European industry and SMEs; ESRF through ESRFup, STFC through participation in XFEL, ELI, HIPER and ESRFup and the European XFEL itself is one of the largest (and most challenging) ESFRI projects.
So community building in particular takes place through the usage of the long-lasting infrastructures, provided by the partners.

Complexity Science
The SSC of Complexity Science is coordinated by a group of six (6) European Universities in 6 different countries together with one of the leaders in the field in USA. During the past years the partners of this SSC have active participation in European and National projects in the field of Complexity Science. It is important to mention the participation in the Coordination Action GIACS, which was the only FP6 CA project dealing with Complex Science, and involved about 30 different Institutes in Europe covering practically all European countries. GIACS lead to the foundation of the European Complex Systems Society, in which the Coordinator of this SSC is a member of the Council. The coordination of the Complexity Science Project SSC will be performed by the CS SSC Steering Committee, which will include representatives from each country that is member of the SSC. The initial list of members will include the representatives from the countries that participate in this proposal. The participation of the CS SSC Steering Committee will expand to include also representatives from other countries, as they will be joining the CS SSC. The goal of the CS SSC is to provide services to the wider European Complexity Science research.

The goal of the CS SSC is to create a sustainable structure within the Complexity Science community through the development and deployment of services, which will lower the entry barrier for new users of the infrastructure and provide robust mechanisms to perform optimal usage of the underlying DCI resources. The CS SSC will provide services to the wider CS research community on one hand and will be the voice of the community within EGI on the other. The CS SSC Steering Committee will be in direct contact with the participating NGIs in order to ensure that the necessary human, software, and hardware resources, as they will be identified by continuous consultation with the research community, will be available at the required levels through National and European funding.

Humanities

The community of humanities researchers in Europe actively using, or with an advanced interest in, grid services and resources is currently relatively small. The Humanities SSC will therefore focus its attention on established digital humanities communities. The partners have excellent institutional bases from which to do this, and have extensive collaborative networks in their own countries and internationally. Furthermore, the SSC's membership includes DARIAH and CLARIN, which together represent the humanities in ESFRI. We will aim to ensure that the early adopters to whom these activities relate have a central point of contact and support to focus their research engagement with the EGI. The Humanities SSC will leverage its partners' existing experience with the (long established) digital humanities (DH) communities to ensure that every opportunity for EGI to reach out to the DH is identified and exploited, and conversely that any research problems or debates that occur in the DH that *could* benefit from Europe-wide e-infrastructure is identified. We will articulate these activities in case studies, reports and conference papers in the domain that are credited and trusted (and where possible, peer-reviewed) by the academic community.

Many of the sustainability issues faced by the Humanities SSC reflect those faced by digital resource creators and users in the existing Digital Humanities community. The Humanities SSC will therefore model its approach on those that have worked well in the past. Mainly, it will identify, prioritize and articulate what is useful: interactions between humanities researchers and the EGI services and community that produce high quality research that is publishable in prestigious journals are likely to sustain themselves through the normal academic publication and dissemination channels; and corresponding demand for e-Infrastructure services will be maintained. We have seen this at national level. The SSC will also add significant value to the process by creating and supporting sustainable networks of individuals, repositories and knowledge within and beyond the early adopter community referred to above. The SSC will target its technical resources towards facilitating their use of EGI. These networks will be underpinned by (SSC-produced) publications and case studies in digital humanities conferences and communities, in the humanities domains themselves (where possible and appropriate) and in the User Forum. These outputs will articulate 'primitives' of interaction between humanities research and EGI that can be documented and referred to by established activities, and by nascent/new ones.