ROSCOE

Robust Scientific Communities for EGI

Draft V30
10 November 2009

Part B

Type of funding scheme:

Combination of Collaborative Project and Coordination and Support Action: Integrated Infrastructure Initiative (I3)

Work programme topic addressed:

INFRA-2010-1.2.3—Virtual Research Communities

Name of the coordinating person:

Charles LOOMIS

List of Participants:
NOTE: Please use consistently the partner short names given in this table. These come from the submitted Forms A2 and are the ones recognized by EPSS. For CNRS and CNR that represent multiple laboratories, please use the specific laboratory name in all cases.

Part. no.	Participant organisation name	Part. short name	Country
1 (coordinator)	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS)	CNRS	FR
		CNRS/CREATIS	
		CNRS/IRSAMC	
		CNRS/LAL	
		CNRS/LRI	
		CNRS/LPC	
		CNRS/HG	
2	BT INFRASTRUCTURES CRITIQUES SAS	BT	FR
3	INSTALLATION EUROPEENNE DE RAYONNEMENT SYNCHROTRON	ESRF	FR
4	UNIVERSITE DE PICARDIE JULES VERNES	UPJV	FR
5	EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH	CERN	CH
6	HAUTE ECOLE SPECIALISEE DE SUISSE OCCIDENTALE	HES-SO	CH
7	SWISS INSTITUTE OF BIOINFORMATICS	SIB	CH
8	UNIVERSITAET ZUERICH	UZH	CH

9	CESNET, ZAJMOVE SDRUZENI PRAV-NICKYCH OSOB	CESNET	CZ
10	STIFTUNG DEUTSCHES ELEKTRONEN-SYNCHROTRON DESY	DESY	DE
11	EUROPEAN MOLECULAR BIOLOGY LABORATORY	EMBL	DE
12	GESELLSCHAFT FUER SCHWERIONEN-FORSCHUNG Gmbh	GSI	DE
13	JUSTUS-LIEBIG-UNIVERSITAET GIES-SEN	JLU	DE
14	MAX PLANCK GESELLSCHAFT ZUR FO-ERDERUNG DER WISSENSCHAFTEN E.V.	MPI	DE
15	GEORG-AUGUST-UNIVERSITAET GOET-TINGEN STIFTUNG OEFFENTLICHEN RECHTS	UGOE	DE
16	UNIVERSITAT DE BARCELONA	UB	ES
17	UNIVERSIDAD POLITECNICA DE VA-LENCIA	UPV	ES
18	CSC-TIETEEN TIETOTEKNIIKAN KE-SKUS OY	CSC	FI
19	ARISTOTELIO PANEPISTIMIO THESSA-LONIKIS	AUTH	GR
20	ETHNIKO IDRYMA EREVNON	EIE NHRF	GR
21	GREEK RESEARCH AND TECHNOLOGY NETWORK S.A.	GRNET	GR
22	FOUNDATION FOR RESEARCH AND TECHNOLOGY HELLAS	FORTH	GR
23	BAR ILAN UNIVERSITY	BIU	IL
24	CONSOREION INTERUNIVERSITARIO RISONANZE MAGNETICHE DI METAL-LOPROTEINE PARAMAGNETICHE	CERM	IT
25	CONSORZIO MULTIENTE PER LA PRO-MOZIONE E ADOZIONE DI TECNOLOGIE DI CALCOLO AVANZATO	COMETA	IT
26	CONSIGLIO NAZIONALE DELLE RICER-CHE	CNR	IT
		CNR/ITB	
		CNR/DEMOCRITOS	
27	SINCROTRONE TRIESTE SCPA	ELETTRA	IT
28	ENTE PER LE NUOVE TECNOLOGIE, L'ENERGIA E L'AMBIENTE	ENEA	IT
29	ISTITUTO NAZIONALE DI FISICA NU-CLEARE	INFN	IT
30	UNIVERSITA DEGLI STUDI DI PALERMO	UNIPA	IT
31	UNIVERSITA DI PERUGIA	UNIPG	IT
32	UNIVERSITA DEGLI STUDI DEL PIEMONTE ORIENTALE AMEDEO AVOGADRO	UNIPMN	IT
33	Academisch Medisch Centrum bij de Univer-siteit van Amsterdam	AMC	NL
34	LOGICA	LOGICA	NL
35	UNIVERSITETET I OSLO	UIO	NO
36	ACK CYFRONET AGH	CYFRONET	PL

37	INSTITUTO DE TECNOLOGIA QUIMICA E BIOLOGICA - UNIVERSIDADE NOVA DE LISBOA	ITQB	PT
38	UNIVERSIDADE DE AVEIRO	UAVR	PT
39	SVERIGES LANTBRUKSUNIVERSITET	SLU	SE
40	STOCKHOLMS UNIVERSITET	US	SE
41	THE UNIVERSITY OF BIRMINGHAM	BHAM	UK
42	CARDIFF UNIVERSITY	CU	UK
43	IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE	IC	UK
44	SCIENCE AND TECHNOLOGY FACILITIES COUNCIL	STFC	UK
45	THE UNIVERSITY OF MANCHESTER	UNIMAN	UK
46	KING'S COLLEGE LONDON	KCL	UK

Table of Contents

List of Tables

Table of Figures

0. Abstract

This section will be deleted in the final version of Part B. A text-only abstract of 1992 or characters or less must be provided in Part A of the proposal.

Grid infrastructures have become indispensible scientific platforms, facilitating European collaborative research in a diverse set of scientific disciplines. The current move from project-based infrastructures to a unified, sustainable European Grid Infrastructure (EGI), based on resources from National Grid Initiatives (NGI), will reassure researchers the long-term viability of grid technologies, increasing its use and impact.
The regionalization associated with the transition to EGI could fragment and isolate existing grid user communities and inhibit new international communities, reducing the scope and pace of scientific advancement in Europe.
To avoid these risks, the ROSCOE (Robust Scientific Communities for EGI) project focuses on building robust, self-supporting, and self-reliant scientific communities. The project will help the communities with this transition, enable the communities to achieve their scientific goals, nurture nascent communities through interactions with mature ones, and transfer management expertise to EGI.eu, the coordinating body of EGI.
Mature Virtual Research Communities (VRC) within ROSCOE are High-Energy Physics, Life Sciences, Computational Chemistry & Materials Science Technology, and Computer Science & Engineering. ROSCOE targets three nascent VRCs: Photon Science concerning users of synchrotron radiation (light) sources, Complexity Science concerning the analysis of complex, interconnected systems, and Humanities. The 36-month work plan includes activities to foster collaboration within and between VRCs, to lower barriers to grid adoption by the members of the communities, and to push the evolution of the grid to maximize its utility.
ROSCOE safeguards current international scientific collaboration, expands that collaboration to new scientific disciplines, and prepares existing and nascent communities to be reliable, stable partners within EGI ecosystem, thereby maximizing the scope and impact of their scientific work.

1. Scientific and technical quality

(Indicative maximum length for the whole of Section 1 – forty pages. This limit does not include the Gantt chart, Pert diagram and tables 1.3a-f))

1.1. Concepts and objectives

Explain the concept of your project. What are the main ideas that led you to propose this work? Describe in detail the S&T objectives. Show how they relate to the topics addressed by the call, which you should explicitly identify. The objectives should be those achievable within the project, not through subsequent development. They should be stated in a measurable and verifiable form, including through the milestones that will be indicated under section 1.4, 1.5 and 1.6 below.

Over the last decade, grid infrastructures have become indispensible scientific platforms, facilitating European collaborative research in Europe from a diverse range of scientific disciplines. This success is the direct result of a series of short-term projects that have guided the advance of grid technology in Europe from R&D through its current, heavy production use. The European DataGrid (FP5) and Enabling Grids for E-sciencE (EGEE, EGEE-II, EGEE-III in FP6 and FP7) series of projects have been the vanguard of this advance.

Despite the clear advantages of using grid technologies, many large, international scientific communities are reluctant to commit to using the EGEE grid infrastructure because the timescale of the associated projects (2-3 years) is incompatible with the timescale of the research programmes of those communities (often 10 years or more). Because of this, the current grid infrastructures are evolving into the European Grid Infrastructure (EGI)—a pan-European, sustainable grid infrastructure based on a federation of National Grid Initiatives (NGI). Each NGI will be financially supported by the country it serves; EGI will be largely funded by membership fees from the constituent NGIs.

The EGEE infrastructure, unlike other specialized infrastructures, has targeted full scientific communities in order to affect a sea change concerning how members of those communities collaborate via shared intellectual and computing resources. The project simultaneously fosters collaboration between communities and coordinates activities within communities. Lively, international communities in high-energy physics, life sciences, computational chemistry, and others testify to the success of this method.

The regionalization associated with the transition to EGI carries a huge risk for the existing and nascent grid user communities—fragmentation and isolation of those communities within NGI boundaries. Such fragmentation and isolation would reduce the scope and pace of scientific advancement in Europe.

To avoid the fragmentation and isolation of scientific communities, EGI.eu must acquire the existing EGEE experience in fostering collaboration between communities and the communities themselves must take on their own coordination. At the same time, current scientific activities must continue and new communities must be supported.

Consequently, the ROSCOE project focuses on building robust, self-supporting, and self-reliant scientific communities. Concretely the goals are:

1. Promote the organizational changes needed within communities for them to become representative of their communities and self-governing.
2. Support community activities and services that help them achieve their scientific goals.
3. Expand the number of active communities by transferring experience from existing mature communities to a select number of nascent communities.
4. Transfer expertise in fostering cross-community collaboration from current projects to EGI.eu.

The mature Virtual Research Communities (VRC) involved in ROSCOE are High-Energy Physics (HEP), Life Sciences (LS), Computational Chemistry & Materials Science Technology (CCMST), and Computer Science & Engineering (CSE). ROSCOE targets three nascent communities: Photon Science (PS) concerning users of synchrotron radiation (light) sources, Complexity Science (CS) concerning analysis of complex, interconnected systems, and Humanities (HUM). All are described in more detail below.

Supporting those communities implies that ROSCOE must also support the goals of those Virtual Research Communities:

A. Increase the number of active users within the community
B. Promote the use of grid technology for collaboration within the community
C. Encourage cooperation through sharing of intellectual and computational resources
D. Safeguard the community's grid knowledge and expertise
E. Develop scientific collaboration between members of the community and members of others

ROSCOE's work programme has been carefully designed to meet its own goals as well as to support the VRCs in meeting theirs.

Add a description of sustainability strategy for the VRCs.

Note: If you want to reference a project goal use the form (Project Goal 1). The context still must be clear even if the reference to the goal wasn't there. Similarly for the VRC goals, use the form (VRC Goal A).

1.2. Progress beyond state-of-the-art

Describe the state-of-the-art in the area concerned, and the advance that the proposed project would bring about. If applicable, refer to the results of any patent search you might have carried out.

This section must absolutely include a description of why this project must be funded. The argument should be that funding the coordination of communities (and not individual users) would lead to a sea change within the communities. This in turn would improve the scientific efficient within those communities and open them up to further collaboration. (With significantly lowered barriers because of the grid technology.)

1.2.1. State of the Art

This section should include statistics on the use of the EGEE computing infrastructure and the range of applications. If possible, this should also include concrete examples of important work that was done on the EGEE infrastructure that would not have been possible otherwise.

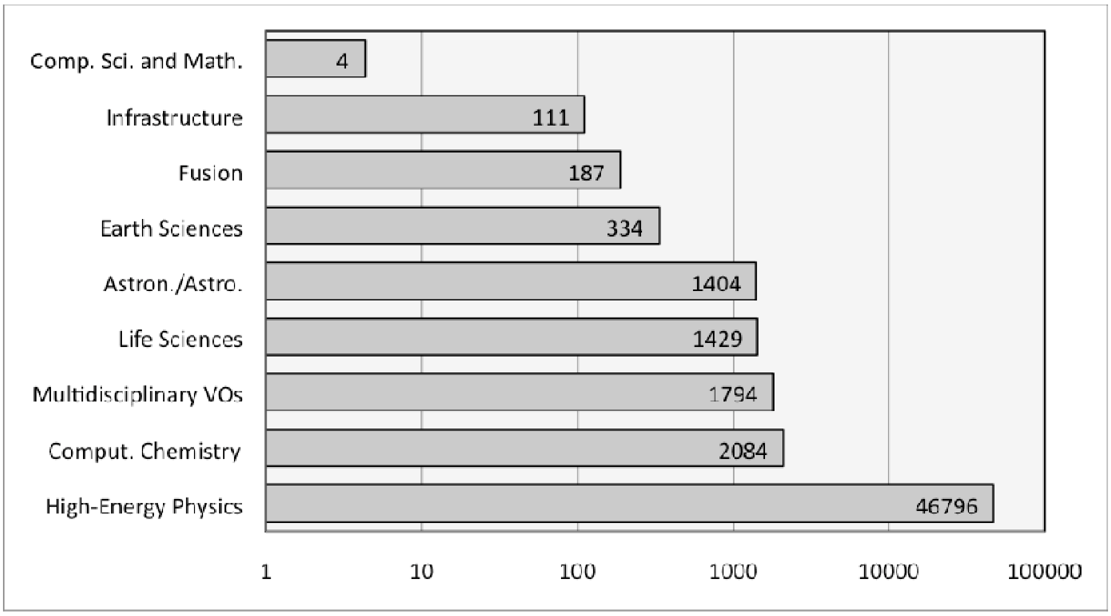


Figure 1: Approx. number of CPUs used on EGEE (9/2008-8/2009)

Provide a description of where the usage data comes from and how to interpret it. Especially point out that the three largest communities are members of ROSCOE. Associate with this is a statement that the resources contributed to the infrastructure are roughly proportional to the usage.

1.2.1.1. State of the Art for Each VRC

Provide a very brief (1 paragraph maximum) of the current state of the art for your community.

High-Energy Physics
HEP is the leading user community by all common metrics – number of active users; disk and custodial storage, CPU requirements and usage – and the major contributor of resources) to the EGEE grid. All the preparation work (simulation and optimisation of the LHC detectors' behaviour, setting up of reconstruction and analysis algorithms) has been done on the grid using its CPU (of the order of 10^5 jobs running at any time) and storage (order of 70 PB of disk deployed across all the infrastructures). The result is that HEP is the first science to have consistently moved to using grid technologies for its core scientific activities. Although the main focus of HEP in EGEE was on the LHC experiments the HEP community at large is using the Grid infrastructure and notably EGEE to complete existing experimental programmes and to plan for new ones.
The commissioning of the detectors required massive computing resources (2B events simulated, reconstructed and analysed for CMS only just in 2009): for all these activities (and for the processing of the real data) LHC experiments depend entirely on the grid.

Life Science
Following previous EU efforts, the Life Science community has gained such an expertise that applications are now able to produce scientific results based on the exploitation of grids. However, compared to other scientific disciplines, the Life-Science community is characterized by its broad diversity. This diversity embodies at various levels including VRC production monitoring, software developments and requirements of user and applications. Currently, LS service providers often repeat development efforts with no real sharing of expertise. Similarly, LS users could share practical expertise about how to use the grid for their particular problem. As a side effect of this diversity, LS activity monitoring is hardly feasible, which

makes production control and incident prevention a difficult task in the VRC. Coordination and federation of scattered efforts in a single LS gateway is now needed to further enlarge the community of LS users based on existing technologies. Besides, application porting remains an art mastered only by a few although portals do exist to provide a friendly interface to deployed applications. Tools are still missing to facilitate the deployment of existing programs on grids.

Computational Chemistry & Materials Science Technology
The large amount and variety of programs and suites of codes developed and implemented by the researchers of the Computational Chemistry and Materials Science Technology to carry out advanced modelling and simulations based on multi-scale and multi-physics approaches starting from the nanoscale level are an integral part of the process of scientific discovery and have traditionally led to a heavy engagement of the best computing facilities ever available. For the same reason, in the recent past, the CCMST computational scientist have started building an international community out of several National Grid Initiatives (GRID.IT – now IGI), RBNED1KNFP, MIUR FIRB, PL-Grid, MetaCentrum, M-Grid, and NDGF). This pan European endeavour is also based on collaborations fostered by two COST Chemistry Actions (D23 and D37) and the propaedeutic effort of the European EGEE Initiative. To consolidate this effort the CCMST scientists have gathered in seminal Virtual Organizations like COMPCHEM, VOCE, TURBOMOLE, GAUSSIAN and have started, as well to structure their computational applications for the grid and, at the same time, try to specialize accordingly a segment of the grid. These VOs find therefore natural to 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 the CCMST Virtual Research Community.

Computer Science and Engineering
The CSE VRC's use of grid and e-infrastructure is focused on advancing grid research and engineering through the exploitation of the European grid itself as its acquisition and experimental apparatus. The CSE VRC thus holds the singular position with respect to the other VRCs that the grid is not the infrastructure, but part of the scientific domain. The community strongly expresses need for real-world, real-size experimental data (for a recent example, see [Ref. 21]). So far, the existing repositories (such as the Grid Workload Archive) are limited to 1) extremely sparse sampling, and 2) specific research targets (e.g. scheduling). With extensive monitoring facilities already in place, EGEE and the future EGI grid offer the best opportunity to observe, and gain understanding of, new computing practices of e-Science and their long-term evolutions. With the Grid Observatory cluster of EGEE-III, first steps have been made in the uncharted territory of digital curation not ***by***, but ***for*** e-infrastructures: preservation, validation, indexing, and publication of the digital assets constituted by the grid behavioural data.
Obviously, the community could also exploit the infrastructure as users. However, its heavy use of high performance computing facilities, whether grids or supercomputers, occurs typically when embedded into other disciplines (e.g. bio-informatics or applied numerical algorithmics), which is taken into account in the corresponding VRCs. Broader use of the grid infrastructure by the CSE community is indeed desirable, and national plans have included this community in their expansion plans (see e.g. [Ref. 22]).

Photon Science
Photon Science user communities are coming from a wide range of scientific disciplines. Currently, Grid usage is rather punctual depending on the particular computational tasks, like e.g.

bio-informatics and statistics. The available PS relevant VOs (e.g. biomed, xfel, xray.eu) are already supported by PS VRC members, but hardly adopted in the communities. The complexity of the Grid environment and lack of PS specific gateways, portals, tools and support seemingly hinders wider utilization of the Grid infrastructure in the communities. Associated with the German NGI, the WissGrid (http://www.wisgrid.de) aims to provide generic support for new communities, among them the German PS represented by XFEL, EMBL and HASYLAB. WissGrid provides blueprints and support for example in direction of sustainable long term archives of scientific data, which is of vivid interest to the PS communities. Currently, data management, archival and curation is to a large extend within the responsibility of the individual research groups, which prevents sustainability as well as full exploitation of the experimental data by the scientific communities and across disciplines.

Complexity Science

Complexity Science is a broad field of scientific endeavour encompassing several fields of knowledge, such as natural sciences, engineering, humanities, health, economics, linguistics, social and political sciences. Complex systems such as human economies, social structures, climate systems, nervous systems and energy or telecommunication infrastructures have emerged from studies on these fields of science and the focus of Complexity Science is placed on their common properties, their structure and their underlying dynamics. In general, the complex systems under study are characterized by their intrinsic structural topology that is very sensitive to initial conditions or to small perturbations. Such fluctuations of state are not normal but may exhibit rare, catastrophic events and lead to system wide collapse. The techniques and methodologies used in the theoretical study of complex systems originate from Statistical Physics and non-linear dynamics, but are highly interdisciplinary.

Humanities

Currently, use of the Grid in the humanities is concentrated on particular kinds and instances of distributed digital information, which in turn usually depends on the engagement and interest of individual scholars and research groups. Put simply, where there is such activity, it is where a humanities scholar has (or has access to) existing data, and has identified a research activity that could benefit from grid-type services or infrastructure. Several initiatives at national level are seeking to build on such pockets of activity (most VRC partners represent such initiatives). These are sometimes, but not always, linked to NGI activity. This is however also a time of great flux for this community: the inception of large-scale digitisation programmes, coupled with the increasing availability and affordability of digitization equipment, means that the full potential of digital information for the humanities can no longer be realized with scattered points of innovation: a coordinated and concerted effort is needed within and across Europe.

1.2.1.2. Examples of Scientific Work

Provide examples of scientific work with exceptional visibility or socio-economic impact carried out on the current grid infrastructures that would not have been possible without the grid. (Limit the description to a single sentence adding a few sentences highlighting the importance of the result.)

High-Energy Physics

High Energy Physics' current flagship project is the Large Hadron Collider (LHC) at CERN. Some 10,000 physicists from around the world collaborate on this project – a discovery machine that is expected to further increase our understanding of the universe. Scientists predict that the LHC's very-high-energy proton collisions will yield extraordinary discoveries about the nature of the physical universe. Beyond revealing a new world of unknown particles, the LHC experiments could explain why those particles exist and behave as they do. The LHC experiments could reveal answers to many of the most profound questions of the physical world. The results of the LHC will be complemented with measurements at a future electron-positron linear collider. Such a machine will provide a unique scientific opportunity at the precision and energy frontiers. A candidate for such a machine is the International Linear Collider (ILC). Continuing the work carried on by the heavy-ion programme at the LHC and the ALICE experiment in particular, the Facility for Antiproton and Ion Research (FAIR) will soon become an international research centre in Darmstadt (Germany). FAIR will provide high-energy primary and secondary beams of ions of highest intensity and quality, including an "antimatter beam" of antiprotons allowing forefront research in five different disciplines of physics.

HEP drove the move of grid infrastructures (notably EGEE) to deliver production-quality services. In the past most of the computing for an HEP experiment was performed at the host laboratory (i.e. accelerator centre). The adoption of the grid in the HEP community made possible the deployment of the necessary computing power and storage in a distributed way across the participating institutes / countries. HEP requirements were driving the creation of a global grid computing infrastructure for LHC and other experiments. The creation of a global infrastructure of centres of competence was needed to provide quality services to HEP. In turn these centres attract other scientific disciplines starting from the local communities in each country.

Life Science

The exploitation of the grid yielded several important scientific results in various fields of the life sciences:

- Researchers from the Cardiogenics consortium, an EU-project coordinated in Lübeck, Germany which aims to discover genetic variations leading to coronary artery disease, used the Enabling Grids for E-sciencE (EGEE) infrastructure to identify possible genetic candidates for the causes of a disease which kills over two million people a year in Europe alone. Their results were published in the March 2009 issue of Nature Genetics (Ref. 19).
- The WISDOM (World-wide In Silico Docking On Malaria) initiative used about one millennium of CPU time from several grid infrastructures to screen biological targets of interest to fight malaria and avian influenza. The best compounds selected in silico were tested in vitro, yielding new drug-like compounds that were patented for further process towards drug discovery.
- Concerning medical imaging, the grid has been exploited by radiologists and neurologists to conduct parameter sweeps in functional imaging of the brain. The Dutch Grid has been exploited to optimise parameters used in the detection of emotional processing in the amygdalae [Ref 20].

Computational Chemistry & Materials Science Technology

(input from Antonio, Fermin and Stavros)

Complexity Science

Complexity science is trying to reveal the function and behavior of complex systems in the real world. Such systems include information-technology networks, power grids, transport networks, financial markets, social systems, nervous systems etc. By using an interdisciplinary approach we try to understand the variety of different problems.

As an example consider, the spread of information and rumors across human interaction networks. Such rumors enhanced by mobile communications can escalate out of control. Knowledge of this phenomenon will prevent further incidents like the one in Hungary in June 2003, which caused nationwide panic about a nuclear explosion, which turned out to be a hoax. Understanding the relay of information through a crowd in panic will help develop more efficient evacuation methods, for instance from football crowds, earthquakes or terrorist attacks. Understanding the way epidemic spreads will yield a more effective approach into immunizing it.

Humanities

The Medieval Warfare on the Grid project (http://www.cs.bham.ac.uk/research/projects/mwgrid/) is a pioneering project that creates agent-based hypotheses to data about which there is little historical evidence. This requires both integrating fuzzy data, and large quantities of processing power. The e-Dance project (http://projects.kmi.open.ac.uk/e-dance/) is a UK e-science project which enables distributed performances across networks of high-end videoconferencing facilities. It also deploys motion capture technologies to record choreographic movements in such environments. Both projects have received recent media attention (see http://www.timeshighereducation.co.uk/story.asp?storycode=407025).

1.2.2. European Grid Infrastructure Ecosystem

Add some text saying that the point of this description is to show how ROSCOE and the VRCs fit in to the ecosystem.

As seen above, 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 grids based on gLite, ARC, and UNICORE middleware distributions into a sustainable European grid.

EGI is a collaboration of National Grid Initiatives (NGIs) working together with a coordinating body (EGI.eu) to deliver a grid service to the European research community. It will provide:

- A governance structure that represents the major stakeholders involved in delivering a federated European e-infrastructure built from national resources and the research communities using it.
- A scalable user support model that allows international research communities within Europe and their worldwide collaborators to use the resources within the e-infrastructure to which they are allowed access.
- Interoperation and interoperability with other e-infrastructures within Europe and with those globally required by the European research community.

The three primary stakeholders for EGI are the resource providers, middleware developers, and the user communities. Figure 2 shows the management structure of EGI. As can be seen from the diagram, the resource providers represented as NGIs own and control the grid infrastructure. The other two major stakeholders—middleware developers and user communities—will interact with EGI through defined committees and particular service units.

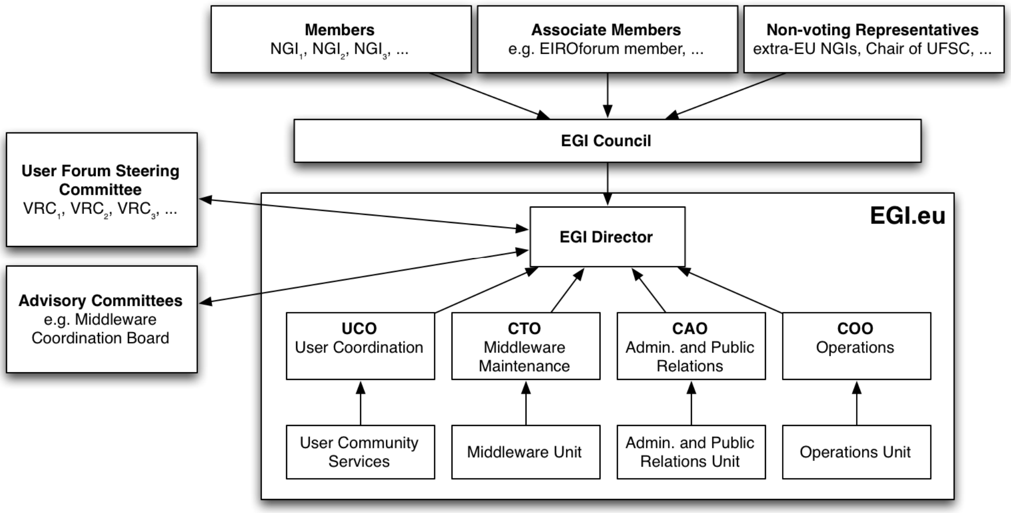


Figure 2: EGI Management Structure

1.2.2.1. Resource Providers

The resource providers directly control EGI through representation on the EGI Council. These are principally NGI representatives from the European states, although EIROforum members may choose to join. The resource providers support EGI by providing access to their computing resources and by paying membership fees to support EGI.eu.

1.2.2.2. Middleware Developers

Unlike in previous grid infrastructure projects, the relationship between EGI and the middleware developers is a client-producer one. EGI will define functional, quality, and delivery requirements for the middleware recommended for EGI.

1.2.2.3. Virtual Research Communities

The user community will interact with EGI primarily through Virtual Research Communities (also called "Scientific Specialized Support Centres (SSCs)" or "Vertical SSCs"). The Virtual Research Communities act as long-lived hubs that allow those communities to safeguard their grid expertise between projects, promote sharing of resources within the community, and provide specialized services to the community. The community itself contains many stakeholders: researchers, research institutes, ESFRI projects, European projects, and national projects.

VRCs will play an active role within the EGI ecosystem, acting as advocates for their communities, coordinating feedback on EGI services, and working with EGI service providers to optimize the support for the community. To fulfil these roles, a VRC must be a legitimate representative of its community and be highly visible within the community.

Figure 3 shows the primary interaction points between a VRC and EGI. The management interactions will take place through the User Forum Steering Committee (UFSC) where each VRC recognized by EGI will be represented. The chair of the UFSC will be a non-voting member of the EGI Council and will frequently discuss with the EGI Director.

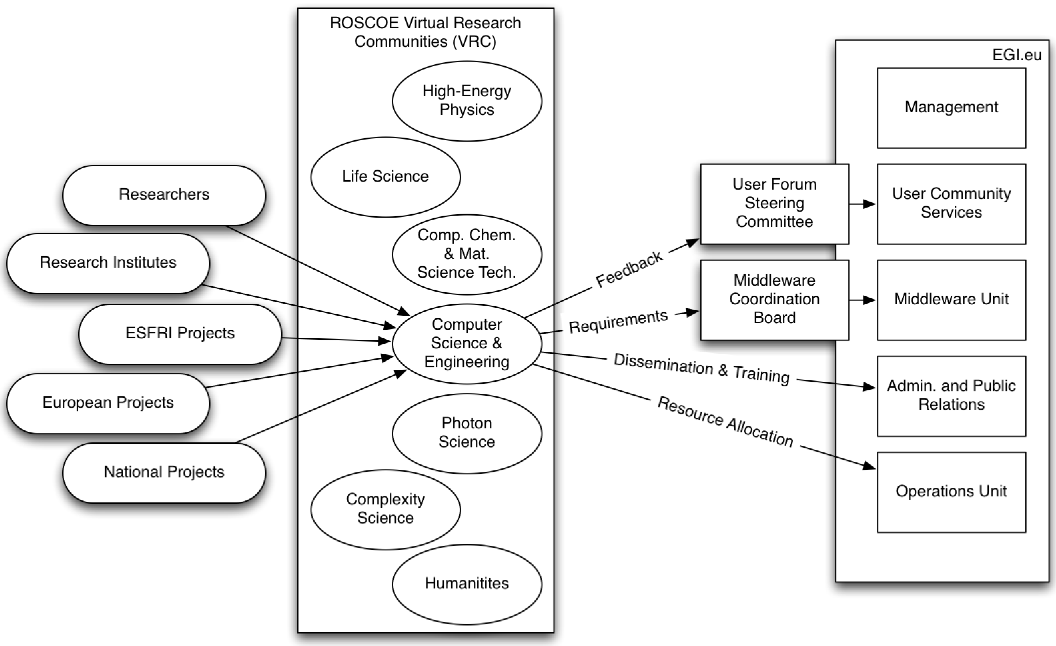


Figure 3: EGI and VRC Interactions

Similarly the VRCs will be represented on the Middleware Coordination Board that will set the middleware development priorities. This group will also act as the liaison between the VRCs and the middleware providers, although direct technical interactions between users and middleware developers are expected when new or improved features are being developed.

The User Community Services (UCS) group within EGI will coordinate and provide generic services to the user community including user support, application porting support, and training. The VRCs will augment those services by providing specialized services for their communities, for example by providing information about access to the VRC's data sets or helping port frameworks or libraries common within the community.
Finally, the VRCs will use the EGI public relations infrastructure to reach the larger European scientific community and the general public. EGI will take advantage of the VRCs to provide a dissemination channel into each VRC's scientific community.

1.2.3. Advances Beyond the State of the Art

Add to this section the expected organizational advances (reason why that is important) and specific scientific advances showing how this project supports those advances. Restate the principal risk of "regionalization" of the communities and how funding this project is mandatory to avoid that. The advances must be tied to specific milestones within the project's work plan.

1.2.3.1. HEP

The expected results of this work are a marked increase in the number of Grid users, as HEP usage expands from the data processing activities that have dominated until now into the realm of data analysis, scientific discovery and publication. This will be accompanied by wider inter-disciplinary collaboration, both through science (i.e. related disciplines) and technology (e.g. grid tools). This can only be achieved by a significant simplification of user interaction with the Grid, through further adoption of existing tools such as those described in detail below, and by a flexible and scalable end-user support model. A key element of this work will be the establishment of community support, whereby the communities are encouraged and enabled to be largely self-supporting, with expert guidance to establish and optimize the support structures and associated tools. This is essential not only to deal with the large expansion in terms of number of users but also for long-term sustainability. As a result we foresee a positive and measurable impact on the quality and effectiveness of the e-Infrastructure that in turn will lead to corresponding benefits to the research communities that use it. These activities will help to ensure Europe's leadership role in the areas of grid design, deployment and efficient exploitation.

1.2.3.2. Life Science

Life-Science grid users and developers are currently fragmented into several communities across Europe, each of them having its own requirements, skills and tools. ROSCOE offers a unique opportunity for those groups to share their expertise and coordinate efforts to avoid redundancy and foster collaboration. Our ambition is to federate regional efforts in a common gateway offering different but harmonized mechanisms to access grid resources. For the LS service providers, this gateway will simplify the development of new tools and enable to build on the experienced capitalized by other groups. For the LS users, ROSCOE will provide a single interface from which they could assemble, present and execute components to design and run experiments. To improve autonomous design of experiments by scientists, a particular emphasis will be put to porting tools enabling the wrapping of scientific codes in grid applications. In addition, this federated gateway will improve LS activity monitoring within the VRC so that incidents will be better detected, logged and handled, resulting in an improved user support.

1.2.3.3. CCMST

The key advantage for structuring the CCMST VOs as a single VRC is not just a further strengthening of the computing platform but also the consequent establishing of a higher level of collaborative grid computing that will lead to the spurring of further discovery. This will be obtained by exploiting common scientific gateways facilitating the access of the user to a large body of computational applications and associated tools (like workflows, mathematical libraries, graphical interfaces, parallelization instruments, etc.). As a result, this will enhance the possibility for the users of structuring and running more complex applications and will foster as well the adoption of models and standards for molecular knowledge and data representations allowing a reuse of quantum chemistry results and codes for an easy connection of different packages and an increase of the grid interoperability. Of particular interest for the CCMST community will be the possibility of utilizing the sensors of the monitoring tools to carry out an evaluation of the quality of the services offered on the grid and of the quality of the users whose judgments will be employed as a means to build a credit system thanks to which the contributions given to the community will be rewarded and an economy of the virtual research community developed. This will imply a structuring of the community activity as grid services and a strong development of the networking system that will produce better support to the users in dealing with the grid technicalities as well as by side strengthening effects for the community when negotiating conditions for the use of libraries and packages.

1.2.3.4. Computer Science & Engineering

The primary goal of the CSE VRC is to develop the Grid Observatory (GO). The GO will contribute to an *experimental theory* of large grid systems by: (i) setting up a *data repository*, (ii) enabling the development of *behavioural grid models* facilitating the deployment of *software experimental facilities*, and (iii) contributing to overall *grid-governance solutions*. The CSE VRC is the only entity in the EGI ecosystem ensuring *digital curation* of the internal production of the EGI grid. The CSE VRC addresses the four components of digital curation. It scales up the process, started with the GO cluster in EGEE-III, of establishing and developing long-term repositories of digital assets for current and future references. It provides digital asset search and retrieval facilities to scientific communities through its gateway. It tackles the good data creation and management issue through formal ontology building. Finally, the CSE VRC adds value to data by generating new sources of information and knowledge through both semantic and Machine Learning based inference. From these available datasets, various segments of the Computer Science and Engineering community, ranging from Machine Learning to Distributed systems design, will be able to build reference datasets, accepted by the sub-communities as representative of the real world (benchmarks), and adapted to their specific scientific interests.. Finally, the CSE VRC will contribute to better grid attractiveness by demonstrating the potential of grid-enabled data sharing and computing through the tools accessible at the gateway.

1.2.3.5. Photon Science

The Photon Science communities foresee a drastically increasing demand for computational and storage resource. Several new facilities became operational very recently or will enter the field in the near future. Particularly the European XFEL will create unprecedented new experimental opportunities and challenges, which has been proven by the exceptionally successful operation of the VUV-Laser FLASH. To fully exploit the opportunities a tight integration of instruments, data management and distributed (high-performance) computing becomes increasingly important. For example, the time structure of the Free Electron Laser beams allows studying chemical or physical transitions at a femto-second scale. To be able to analyse the experimental data, molecular dynamics simulations of comparably large systems at an atto-second scale. The PS VRC will allow exploring the EGI e-infrastructures for the photon science communities to meet these computational challenges as well as the requirements in many aspects of the lifecycle of the experimental data. Since VRCs in ROSCOE like for example CCMST face precisely the same computational challenges, synergies from these VRCs will not only improve efficiency but also combine expertise from currently rather distinct scientific communities, which hitherto can accelerate scientific progress substantially.

1.2.3.6. Complexity Science

Computer studies of complex systems have attained much attention during the last two decades, but as the science of complex systems has evolved into encompassing real world networks of agents and elements such studies have been hindered by computational and storage capacity limits. Adaption to new programming models and techniques suited for the optimal exploitation of Grid infrastructures will boost the production of scientific results and ease the collaboration of research groups in the field of Complexity Science. In addition, the storage capabilities provided by the existing Grid infrastructures (along with the emerging Data infrastructures) will assist researchers of the field in building up collaborative and distributed databases of large scale real world networks and systems.

1.2.3.7. Humanities

The burgeoning production and availability of dispersed digital resources for humanities researchers has created many new opportunities. However, approaching those resources in ways which usefully address humanities research questions (or create new ones) is a far from trivial task. Hence current applications of humanities on the grid which do exploit the potential of e-infrastructure to conduct advanced research are bespoke activities with high overheads and extremely specialized collaborations between domain experts and computer scientists and/or e-infrastructure professionals. The H-VRC will seek to support such specialized collaborations wherever they arise, but also to move the humanities research community to the point where it can engaged with certain (probably limited) parts of the e-Infrastructure without them.

1.2.4. Relationship with PRACE

Provide a general description of the relationship with PRACE. Use examples from the VRC comments below to show that HPC resources are of interest to the ROSCOE communities. Project focuses on DCI platforms (and comes from them). However, many of the researchers within the communities will have need of HPC resources and the VRC should develop a common strategy for using them as transparently as possible. There should be dialog with PRACE on how to bring them into the fold, even if the infrastructures remain separate.

1.2.4.1. HEP

The main computing resources used by the High Energy Physics (HEP) community are not supercomputers; however supercomputers and multi-core architectures are required in some specific HEP research areas, such as Lattice QCD simulations for the QCD thermodynamics studies in the context of heavy-ion collisions experiments (LHC, RHIC). One such study performed in 2008 required a pre-thermalisation phase to be run on a supercomputer – a 1.2 TFlops NEC SX8 in High-Performance Computing Centre in Stuttgart – while the bulk of the simulations was subsequently run on the WLCG/EGEE Grid. To further increase the precision of the Lattice QCD studies future Grid or cloud infrastructures should be able to support large number of locally parallel jobs exploiting multicore architectures. As multicore architectures are now predominant the ability to efficiently exploit them is now a general problem and hence collaboration with communities with experience in these areas will become increasingly important.

1.2.4.1. Life Sciences

The Life Sciences community has needs for resources on cluster-based grids and on supercomputers. This was clearly shown in a recent survey conducted in France for the French National Grid Initiative to which more than 400 scientists contributed from about 30 laboratories across the country. Needs were expressed for access to grid platforms for on demand computing but also to HPC resources. The Life Sciences VRC involves several partners (CSC, CNRS) that are also involved in PRACE. They will participate to the dialogue with PRACE.

1.2.4.2. CCMST

Computational chemistry is well known for it very high demands concerning CPU power as well as disk and memory requirements. The source of this relies in two main groups of computational methods utilized by the community: first principles (ab initio and Density Functional Theory) and molecular dynamics respectively. The first group of methods is widely used to describe molecular and matter properties on fixed geometries. The key factors influencing demands are number of basis set functions and the computational model used. For highly correlated methods memory requirements can scale with fourth power of number of basis functions. While such problem may seem not to be very demanding for small molecules (say up to 10 atoms) it becomes computationally prohibited for molecules containing 50-100 atoms (especially with heavy elements) being of interest recently among chemical community. The latter method – molecular dynamics is used to study changes in molecular systems in time. Those methods are widely used not only in chemistry and materials science but also in other science disciplines. The challenges MD has to deal with are the number of atoms of the system studied (hundred thousands to millions of atoms) and a time of simulations. The common remedy for both methods is usually parallelization of the software codes. While this can be done quite easily for molecular dynamics it becomes very difficult for first principle methods. Nevertheless both groups of methods utilize roughly 60-70% of computational resources of HPC centres. Despite the highly increasing interest in grid computing among CCMST community certain applications will still require HPC resources. Especially modelling of materials of industrial and/or biological importance requires several thousands of processors for single run. Such computations utilize highly parallelized software packages like CPMD or NAMD. It is a responsibility of the VRC's experts to guide users which computational platform as well as software is better suited for a given computational problem. At this level close cooperation with PRACE project is necessarily.

1.2.4.3. Computer Science & Engineering

Missing information.

1.2.4.4. Photon Science

The Photon Science communities, both at the synchrotron light sources as well as FEL's see a strong need for resources on cluster-based grids and on supercomputers, as outlined in a recent CECAM workshop on Computational Challenges emerging from next generation Light Sources. Demands are coming from various fields like automated macromolecular structure determination, molecular dynamics simulations at an attosecond scale to study rapid structural transitions or single-particle imaging. One of the PS VRC partners, namely STFC is participating in PRACE and also represents light source facilities as well as user communities and can therefore substantially contribute to the cooperation and dialogue with PRACE.

1.2.4.5. Complexity Science

The development and adoption of large-scale parallel models and applications in the context of modern application programming frameworks adapted for High Performance Computing will allow for the numerical computation of the dynamics of several real world complex systems. Thus, the combined usage of HPC and Grid resources in the field of Complexity Science will boost the production of scientific research by bringing scientists together in a transparent way. The Complex Science VRC will act as a bridge between the infrastructure and the Complex Science community.

1.2.4.6. Humanities

The potential user community for an VRC facilitating access to, and services for using EGI facilities for the humanities is diverse, distributed and complex. Some of the problems in developing e-Infrastructures for the Humanities will be very different for those of other VRC communities in the experimental sciences. Generally, processing capability is less needed than access to large-scale and complex data volumes. Recent years have seen huge efforts to digitise existing humanities resources and provide them online: the efficacy of these resources for research is greatly reduced if domain experts cannot access and use them as intuitively as they use other kinds of data and tools they work with. There is therefore now a clear need for a Grid Infrastructure to provide access to research data and novel tools to extract information from it. Several Humanities ESFRI projects have been set up to serve the needs of this diverse domain: CLARIN targets the linguistic community, while DARIAH will look after a wide range of user communities in what is known to be 'Digital Humanities'. Digital Humanities is a broad term which has come to mean research conducted in the humanities disciplines – history, languages, textual studies etc – which is substantially supported or enabled by digital resources, applications or tools; and which produces new knowledge that could not be produced without such resources, applications or tools.

1.3. Methodology

Describe the methodology to achieve the objectives of the project, especially the way integrated services will be provided.

ROSCOE aims to build robust, autonomous user communities while encouraging cross-disciplinary cooperation. These two interests dictate different organizational structures for the project. The first requires a project organization based on the Virtual Research Communities (VRCs); the second, an organization based on common tasks. To address these competing interests, ROSCOE adopts a hybrid structure where the work packages are based on common tasks but each consists of subtasks (and effort) tied directly to a VRC. The principal management committee, the Technical Management Board, consists of the VRC coordinators to reinforce the VRC as an autonomous structure, preparing them for the future where they must act as independent bodies. Figure 4 illustrates the project organization.

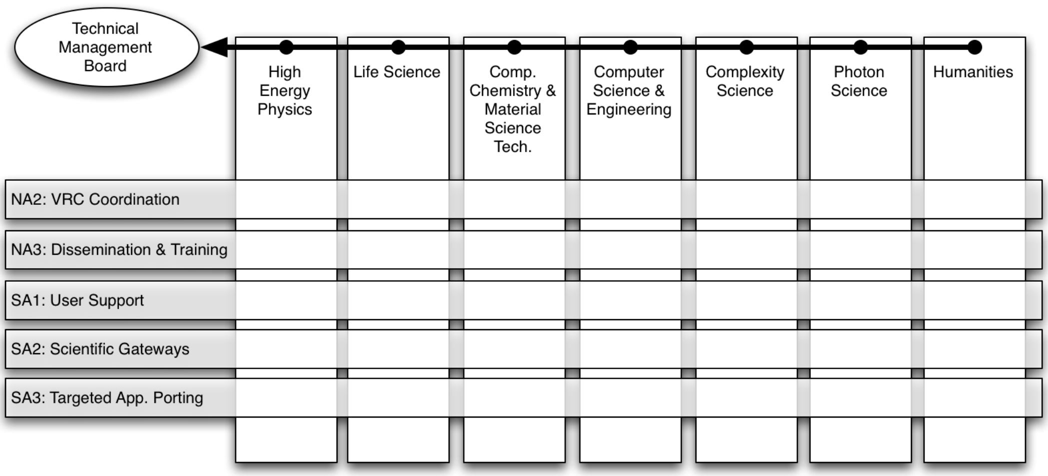


Figure 4: VRC and Work Package Organization

One consequence of this organization is that there will not be a "natural" lead partner for each of the defined networking and service activities. Instead the VRC coordinator will be responsible for ensuring that his personnel within each activity contribute effectively. In addition, the people involved in the Project Coordination (NA1) activity will be responsible for common activities within each work package. Some examples of these common activities are:

- Preparing for materials for a common ROSCOE booth and organization of common training events for NA3.
- Contributing to a common user support team in SA1 to handle high-level triage of support tickets and resolution of those tickets where possible.
- Gathering of requirements for a common scientific gateway implementation.
- Identifying common libraries or frameworks that can be useful for more than one scientific discipline.

Having personnel in each VRC participate in these cross-discipline activities will improve the awareness of what is happening in other VRCs and foster closer collaboration.
The increased awareness of activities in other areas is also expected to improve multi-disciplinary scientific research. Table 1 shows scientific and technological areas of common interest between communities involved in ROSCOE. The VRC coordinators will foster collaboration in these areas and help to identify others.
Please fill in the following table with concrete examples of collaboration between the ROSCOE communities.

Table 1: Scientific and Technical Interactions Between VRCs

HEP↔LS	Development and use of AMGA, Dashboard, DIANE, Ganga. Common use of the GEANT4 toolkit. Collaboration through the PARTNER project.
HEP↔CCMST	Common use of AMGA and Ganga tools.
HEP↔CSE	Interactions based on HEP community tools that provide monitoring information.
HEP↔PS	Common data management and archival problem. Common data (dCache) and possibly metadata (AMGA).
HEP↔CS	Collaboration and use of AMGA for metadata management.
HEP↔H	
LS↔CCMST	Joint deployment of molecular dynamics and first principles (*ab initio* a nd DFT) software codes.
LS↔CSE	Interactions based on LS community tools that provide monitoring information. Report to GO the VRC activity monitored by LS gateway. GO tools helps production control by detecting abnormalities to prevent incidents.
LS↔PS	Common user communities regarding bioinformatics and structural biology and common experimental methods (instruments). Joint effort for collaboration with ELIXIR and INSTRUCT ESFRI projects.
LS↔CS	Emerging disease monitoring and modelling. Scientific approaches to study protein interactions in the context of physiological studies.
LS↔H	
CCMST↔CSE	Interactions based on CCMST community tools that provide monitoring information.
CCMST↔PS	Common user communities around molecular dynamics, ab initio, and computational chemistry.
CCMST↔CS	
CCMST↔H	
CSE↔PS	
CSE↔CS	Complex Networks software. Sharing of expertise in the field of complex systems to better understand grid usage trends.
CSE↔H	Both communities address Digital Curation
PS↔CS	
PS↔H	
CS↔H	Sharing and exchange of socio-economic databases.

The project's Joint Research Activities (JRAs) are not featured in the above diagram because only a subset of the VRCs participate in them, even though the results may be of interest to all of them. The project's two JRAs focus on activities to improve the functionality of the grid infrastructure: Data Management Evolution and Contributions to Sustainable Operations.

The current tools for managing data on the grid infrastructure are too limited and too basic to allow scientific communities to manage effectively large, diverse data sets. Some communities (e.g. high-energy physics) currently do manage large data sets on the grid, but this is accomplished by layering heavy, complex frameworks on top of the grid middleware. The ROSCOE activities will compile requirements from the VRCs to determine what high-level services are needed and what data storage technologies must be interfaced to the grid. Using these requirements, promising new technologies will be tested against real use cases in conjunction with EGI. These experiences will be fed back through EGI's Middleware Coordination Board to drive the adoption of new data management tools.
The Contributions to Sustainable Operations contains two tasks: feeding back results from the analysis of Grid Observatory data to improve operations and expanding the types of accounting data available to a VRC to allow economic models for sharing. These will improve the cooperation between the EGI operations teams and the VRCs.
Need to add a general statement about collaborating with other projects and other communities. This could include, for example, a description of how important MPI is and what will be done to engage with others to ensure that this is solved on the infrastructure.

1.4. European Grid Infrastructure (EGI) Interactions

Because we wish to promote our VRCs as independent, stable, and sustainable entities, the VRCs will in most cases interact directly with structures within EGI. The VRCs are expected to directly participate in the User Forum Steering Committee, Middleware Coordination Board, and User Community Services. Similarly, common service activities like dissemination, training, and user support coordinators will directly interact with the appropriate personnel in EGI.eu. The Project Coordinator will interact directly with the EGI director. Figure 3 illustrates the interactions between the VRCs and EGI.

1.5. Collaborations with ESFRI Projects

Add a table of ESFRI projects and how this project will interact with them. This table should include a column about whether the ESFRI has provided a Letter of Support for the project. The most current update to the ESFRI roadmap is available here:
ftp://ftp.cordis.europa.eu/pub/esfri/docs/esfri_roadmap_2008_update_20090123.pdf
A number of ESFRI projects participated in the most recent EGEE conference (EGEE'09). The agenda listing those projects is available here:
http://indico.cern.ch/sessionDisplay.py?sessionId=12&slotId=0&confId=55893#2009-09-21

ESFRI Project Name	LoS?	Description of Collaboration
S-LHC		The Large Hadron Collider (LHC) at CERN, now starting, will be the energy frontier machine for the foreseeable future and it has the highest priority to fully exploit its physics potential. Depending on the nature of the discoveries made at the LHC, higher-statistics studies of these phenomena would naturally call for an increase in luminosity. This upgrade – referred to as Super-LHC – should increase the luminosity by a factor ten. Super-LHC features in the European Strategy Forum on Research Infrastructures (ESFRI) roadmap document [ref]. The HEP VRC will work with the team planning the Super-LHC to ensure their simulation models can be deployed on the EGI production grid infrastructure.
ILC		The results of the LHC will be complemented with measurements at a future electron-positron linear collider. Such a linear collider will provide a unique scientific opportunity at the precision and energy frontiers. This programme can be carried out by the International

		Linear Collider (ILC) or, if multi-TeV energies are needed, by a novel design called the Compact Linear Collider (CLIC) which has the potential to deliver such energies. For essentially every new physics scenario involving particles in the linear collider energy range, detailed and very promising research programmes have been formulated which will require on detailed computer simulations of the machine and detectors. The HEP VRC will work with the ILC and CLIC teams to ensure their simulation software can be deployed on the EGI production grid infrastructure.
FAIR		The HEP VRC team will also have similar interactions with the Facility for Antiproton and Ion Research (FAIR), which will soon become an international research centre in Darmstadt (Germany). Construction has already started at GSI in Darmstadt. FAIR will provide high-energy primary and secondary beams of ions of highest intensity and quality, including an "antimatter beam" of antiprotons allowing forefront research in five different disciplines of physics. CERN will interact with the FAIR collaboration to exchange of knowledge and skills of the accelerator, detector and grid technology.
DARIAH	Y	TBD
CLARIN	Y	TBD
CESSDA		Investigation of sharing of socio-economic databases and assistance in data management planning
ESS		Investigation of possibility to contribute research data
EPOS		Investigation of sharing of modeling tools for natural hazards prediction and recovery tactics
PRACE		Investigation of using HPC infrastructure for large temporal and spatial scales of real world data simulations
EuroFEL		ELETTRA & DESY are EuroFEL partners. Collaboration on security management (AAA) and data management envisaged.
INSTRUCT		EMBL is INSTRUCT member. ITQB applied for partnership. Projects like iSpyB integration are suitable for collaborations.
Eur.XFEL		DESY, STFC and the European XFEL are participating members. Development of data infrastructure in close collaboration.
HIPER		STFC is member of HIPER. Science program of HIPER is in line with FLASH, Eur.XFEL and FERMI@ELETTRA. Expect collaboration on application porting and simulation tools.
ESRFup		Upgrade project of ESRF. Collaboration of Grid studies and user management with DESY and PSI.

1.6. Interactions with Other Projects

Additionally, we know of a large number of EGI-related projects that will be providing user support, application porting support, and middleware services. There will also be other projects in the Virtual Research Communities call that will have common interests and goals and may result in fruitful collaboration. Preliminary discussions have already been held with the following proposals:

- CUE: Creating Users of E-infrastructures
- TAPAS: Team to Assist Porting Applications to e-Science infrastructures
- SAFE: VRCs for Astronomy, Astrophysics, Earth Science, and Fusion
- EMI: European Middleware Initiative
- EGI-SGI: EGI Science Gateway Initiative

- StratusLab: Combining grid and cloud technologies
- GridTalk-II: Disseminating information from communities to wider audience.

Discussions on concrete collaboration between the projects will be held once it is known which are approved.

1.7. Networking Activities and associated work plan

Describe the extent to which the proposed co-ordination mechanisms will foster a culture of cooperation between the participants, and enhance the services to the users.
A detailed work plan should be presented, broken down into work packages (WPs) which should follow the logical phases of the implementation of the project's Networking Activities, and include consortium management and assessment of progress and results. (Please note that your overall approach to management will be described later, in section 2).

1.7.1. Overall strategy of the work plan

(Maximum length — one page)

The ROSCOE Networking Activities (NAs) will enhance cooperation between the collaboration participants, between members of the Virtual Research Communities (VRCs), and between different VRCs.
The Project Coordination (NA1) work package will ensure that all of the project participants contribute effectively to the project and that there is sufficient communication between the different VRCs involved in the project. In addition, the project management will be responsible for managing the cross-discipline activities of the other two networking package. The project management will also work to transfer its expertise to EGI.eu as the project ends.
The VRC Coordination (NA2) work package will focus on community building activities in each of the scientific communities involved in ROSCOE: High-Energy Physics (HEP), Life Sciences (LS), Computational Chemistry & Materials Science Technology (CCMST), Computer Science & Engineering (CSE), Complexity Science (CS), Photon Science (PS), and Humanities (HUM). The coordinator of each VRC will work to bring about the social and political changes necessary to be as self-reliant and self-supporting as possible. The subtasks of the activity (as well as the other networking and service activities) are organized by VRC to ensure that each VRC takes responsibility for its part of the work plan, reinforcing the VRCs as a central hub for its community.
The Dissemination & Training (NA3) work package aims to maximize the number of people within a VRC that are aware of and that use the grid infrastructure as well as to ensure that they use the infrastructure optimally. Each VRC will actively work to disseminate the activities of the VRC within its scientific community, advertising success stories at relevant conferences and events. Dissemination activities aimed at the wider European scientific community and the general public will be done in conjunction with EGI.eu. The VRCs will provide material that appeals to a broad audience to EGI.eu and rely on it to disseminate that information. In turn, the EGI.eu dissemination coordinator can rely on VRCs to pass information to their scientific communities.
Add a table that identifies what project and VRC goals these activities deal with?

1.7.2. Timing of the different WPs and their components

(Gantt chart or similar)

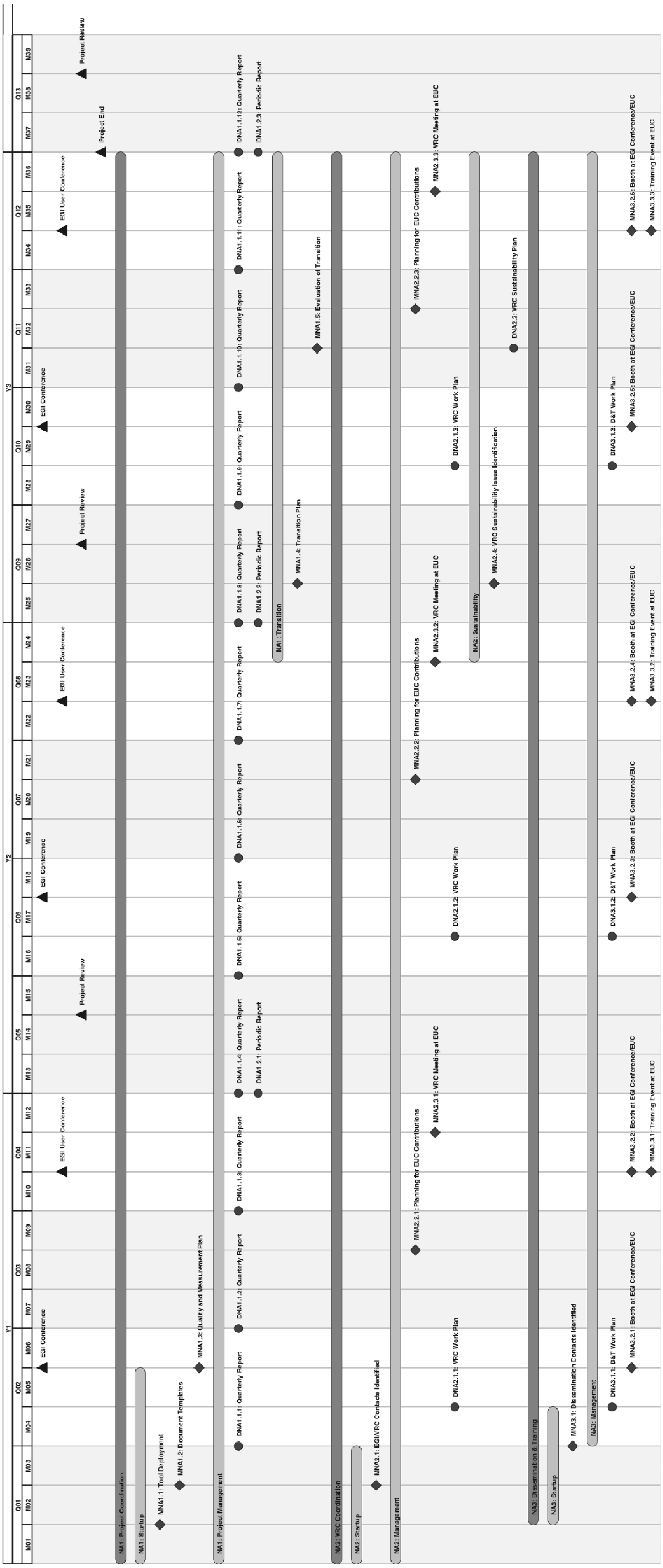


Figure 5: Gantt Diagram for Networking Activities

1.7.3. Detailed work description broken down into work packages

Table 2: Work package list

Work package No[1]	Work package title	Type of activity[2]	Lead Part. No[3]	Lead Part. short name	Person / months[4]	Start month[5]	End month
NA1	Project Coordination	MGT		CNRS		M01	M36
NA2	VRC Coordination	COORD		CNRS		M01	M36
NA3	Dissemination & Training	COORD		CNRS		M03	M36
	TOTAL						

Table 3: Deliverables list

Del. no.[6]	Deliverable name	WP no.	Nature[7]	Dissemination level[8]	Delivery date[9] (proj. month)

[1] Work package number: WP 1 – WP n.
[2] Please indicate one activity per work package:
RTD = Research and technological development; COORD = Co-ordination; MGT = Management of the consortium; SVC = Service activities
[3] Number of the participant leading the work in this work package.
[4] The total number of person-months allocated to each work package.
[5] Measured in months from the project start date (month 1).
[6] Deliverable numbers in order of delivery dates. Please use the numbering convention <WP number>.<number of deliverable within that WP>. For example, deliverable 4.2 would be the second deliverable from work package 4.
[7] Please indicate the nature of the deliverable using one of the following codes:
R = Report, **P** = Prototype, **D** = Demonstrator, **O** = Other
[8] Please indicate the dissemination level using one of the following codes:
PU = Public
PP = Restricted to other programme participants (including the Commission Services).
RE = Restricted to a group specified by the consortium (including the Commission Services).
CO = Confidential, only for members of the consortium (including the Commission Services).
[9] Measured in months from the project start date (month 1).

ROSCOE

Table 4: Project Management (NA1)

Work package number	NA1	**Start date or starting event:**			M01	
Work package title	Project Coordination					
Activity Type[10]	MGT					
Participant number						
Participant short name						
Person-months per participant:						

Objectives

- Manage and monitor progress towards stated goals.
- Coordinate interactions with the European Commission.
- Ensure effective communication between project participants and between ROSCOE and related projects.
- Provide administrative support to ensure timely, high-quality technical and financial reporting.
- Transfer community management experience to EGI.

Description of work (possibly broken down into tasks) and role of participants

The ROSCOE consortium consists of a large number of partners that are geographically distributed throughout Europe and represent a diverse range of scientific disciplines. Managing such a consortium requires a well-developed organizational structure (see Section 2) as well as significant, dedicated effort to support the consortium. This activity provides that effort.
This work package consists of the Project Coordinator and the Project Office. The organization of tasks within the work package follows the primary objectives of the work package.
Add a statement that says that we rely on the VRC coordinators to interact with partners within their area to reduce the complexity of managing so many partners. Additionally, this will help to lead to more formal structures between those partners (potentially JRU's between them, for example).

Phases

Phases: setup, coordination, transfer.

Managing and Monitoring Progress Towards Goals

ROSCOE is centred on having our Virtual Research Communities play a strong, visible role as VRCs within the European Grid Infrastructure. This task will monitor each community's progress towards becoming or effectiveness as an accepted EGI VRC. This task requires periodic review of each defined Virtual Research Community within ROSCOE.
Regular meetings of the Technical Management Board with reports to the Project Management Board will ensure that any issues are quickly identified and corrected before they can have a negative impact on the project's programme of work.

Coordinating Actions with the European Commission

All projects must interact frequently with the European Commission to report on the project progress, to raise issues encountered during the course of the project, to arrange formal re-

[10] Please indicate one activity per work package:
RTD: Research and technological development; COORD: Co-ordination; MGT: Management of the consortium;
SVC: Service activities

views of the project, and to provide input on European programmes affecting our user communities.
Interactions with the European Commission will take place through the Project Coordinator with administrative support from the Project Office. As necessary the Project Coordinator will act as a high-level liaison between the ROSCOE user communities and the European Commission.

Ensuring Effective Internal Communication
Reaching the project goals requires effective communication between all of the project participants. Having those participants widely distributed around Europe complicates effective internal communication; however, this can be controlled through clear communication channels and effective tools.
The Project Office will maintain a directory of all project participants, containing their roles within the project and their contact information. Lists based on this directory will be maintained to ensure effective electronic communication between participants and identified groups of participants.
Additional tools to manage the project's real-time communications, meeting agendas, etc. will be deployed as necessary to facilitate communication between the project's participants.

Ensuring Effective External Communication
Equally important is effective communication with the large number of EGI-related projects that will provide a range of services to the user communities. The project office will liaise with those other projects to ensure that the necessary technical communication takes place between the projects and to develop memoranda of understanding between projects to clearly delineate boundaries between projects and points of common interest.

Providing Administrative Support
All European projects require a significant level of technical reporting. The Project Office will coordinate the work on periodic reports to the European Commission. It will also define and manage the progress for tracking the project's deliverables and milestones, as well as arranging for quality reviews of them.
In parallel with the technical reporting, the project must also provide financial reporting. CNRS, the coordinating partner, has experienced and effective financial units who will be in charge of the financial justification and reporting. However, the Project Office will coordinate the gathering of the necessary information and following each partner's contributions.

Intellectual Property Management
Intellectual Property Management rules will be reflected in the Consortium Agreement to appropriately recognize the background and the expertise brought into the project by its participants. IP issues will be discussed in close collaboration with each partner's IP departments.

Partner Contributions
CNRS will participate in all of the tasks associated with the project management. Both the Project Coordinator and the Administrative Coordinator will be provided by CNRS.
BT will provide quality assurance expertise for the project's results by participating in the "Providing Administrative Support" task. Related to this, it will help monitor the progress of the project towards its goals by defining appropriate metrics ("Managing and Monitoring Progress Towards Goals" task).
GRNET will participate in all of the tasks associated with the project management, concentrating on management of common tasks such as dissemination and training and relationships with external projects.

Deliverables (brief description and month of delivery)

Provide quarterly reports and yearly periodic reports. Perhaps split these quarterly reports into administrative ones (for the project) and ones based on the scientific program. Provide minutes of the PMB and Collaboration Board meetings as annexes to these reports. There should also be a Quality and Measurement Plan for the project.

Table 5: VRC Coordination (NA2)

Work package number	NA2	**Start date or starting event:**		M01	
Work package title	VRC Coordination				
Activity Type[11]	COORD				
Participant number					
Participant short name					
Person-months per participant:					

Objectives

- Coordinate VRC activities
- Liaise with other EGI activities and EGI-related projects
- Liaise with NGIs about deployment of services and provision of resources
- Facilitate sharing of resources
- Develop and implement VRC policy and procedures
- Develop sustainability and exploitation plans
- Evaluate technical and scientific impact

Description of work (possibly broken down into tasks) and role of participants

The EGI user community is organized in to Virtual Research Communities (Specialized Support Centres (SSCs) in EGI jargon). The Virtual Research Communities (VRCs) represent a given community within EGI, acting as a liaison between the community and the various activities within EGI. Additionally, the VRCs are expected to act as a hub within the community, safeguarding the community's grid expertise and knowledge as well as coordinating grid activities within the discipline.
To fulfil these goals, the VRC needs to be a long-lived structure with appropriate plans for sustaining the community's use of the grid, exploiting the acquired knowledge, and liaising between the various community stakeholders.
A typical VRC will cover one or more Virtual Organizations. As one of the primary reasons for using the grid is the sharing of computing resources, the VRCs will encourage sharing of resources within the covered VOs and between those VOs.
To understand its effectiveness, each VRC will have to evaluate its technical and scientific impact. This analysis will be done periodically, in collaboration with EGI and other VRCs.

Phases

Phases: setup (plans), community building & coordination (sustainability plan), preparation for sustainability (status report/analysis).

Common Tasks

TMB meetings should ensure communication between the VRCs and allow them to identify common areas of interest. This may involve the setup of specific working groups.

Each VRC coordinator must review their description below and rewrite it keeping in mind the stated objectives above as well as the general description. The length of each contribution should be 0.50-0.75 pages. The description should start with a brief description of how the

[11] Please indicate one activity per work package:
RTD: Research and technological development; COORD: Co-ordination; MGT: Management of the consortium;
SVC: Service activities

VRC will be coordinated.

High Energy Physics

High Energy Physics (HEP) is a well-established global VRC with a number of mature coordination bodies and regular conferences and meetings of users and/or service providers. This activity will work through such existing structures wherever possible, adding direction where required but retaining overhead to the strict minimum. As such, a quarterly partner review is foreseen through which the overall activities of the VRC and its relations with other communities and bodies will be reviewed. HEP covers not only the four main LHC experiments but also other communities corresponding to collaborations at CERN or other facilities. The sharing of knowledge and techniques across such virtual organisations, as well as with other VOs in different disciplines, will be given special emphasis, as will the development and implementation of a long-term sustainability plan – essential in a community where it is common for VOs to have lifetimes measured in decades.

Effort: CERN (liaison to WLCG and other HEP coordination bodies): 0.5 FTE, un-funded; OSLO (liaison to EMI): 0.5 FTE, co-funded.

Life Sciences

Coordination is a major concern for the LS community given its coverage of a wide spectrum of applications and user communities. Previous experience of the LS community showed clear needs for coordinating (i) user activities, (ii) development efforts and (iii) relations with resource providers (NGIs). The specific tasks identified for the LS VRC address the main goals of NA2 as detailed below.

NA2 objective 1: "coordinate VRC activities"

- Develop expertise sharing between VRC users to facilitate grid access and application porting
- Develop expertise sharing between VRC service providers to avoid redundant development efforts

Those tasks could be facilitated by the integration of appropriate coordination tools within the LS gateway (e.g. forum, wiki).

NA2 objective 2: "liaise with other EGI activities and EGI-related projects"

- Coordination with EGI dissemination SSC, NGI and Regional grid infrastructure dissemination structures (IBRB, INFN)

NA2 objective 3: "liaise with NGIs about deployment of services and provision of resources"

- Integrate the LSVRC activities with NGI activities: provide information about resource needs and service requirements and give feedback on the infrastructure.

NA2 objective 4: "facilitate sharing of resources"

- Develop synergies between NGIs
- Integration of resources operated by ARC in the biomed VRC

NA2 objective 5: "develop and implement VRC policy and procedures"

- Identify "best practices" in grid exploitation to improve production control
- Maintain list of VRC members
- Monitor VRC activity and resources

NA2 objective 6: "Develop sustainability and exploitation plans"

NA2 objective 7: "liaise with ESFRI projects"

- Collect computing requirements from selected ESFRI projects

Explore the integration of grid resources in selected ESFRI computing infrastructures

LS partners involved in this work package are CNRS (17PM), HealthGrid (8PM), HES-SO

(4PM), CSC (3PM), CERM (7PM), INFN (4PM), EMBnet (6PM), AMC (7PM) and EBI (22PM)

Computational Chemistry & Materials Science Technology
The main management challenges are likely to be related to the dynamic nature of the VRC. Ongoing deployment and integration work, associated with the need to integrate and support our scientific partners in exploiting the VRC infrastructure, requires substantial agility of decision procedures.
The management of VRC is structured in a way to allow a good communication and flow of information within VRC as well as between VRC and EGI or NGIs. The entities composing management structure will involve the VRC chair elected by VRC Management Board, representing CCMST VRC in EGI User Forum. In addition we will nominate also the local coordinators (one from each institute or research group) to ensure a good contact points with scientific laboratories and research groups. The Management Board (formed by supporting partners) role is to ensure long sustainability of VRS. For this it is crucial to verify VRC partners' achievements periodically and evaluate impact for community to better exploit their technical and scientific potential.
Development of policies and procedures is essential to target common mechanisms for resource sharing between different parties among VRC community as well as between CCMST and other communities and national and regional grids and EGI. Especially cooperation between VRC and NGIs needs to be covered as community in first place relies on the resources and services provided by them. Such services and procedures have to cover not only implementation of new tools developed or required by the community by NGI on their infrastructure but also new computing resources coming from research laboratories attached to e-infrastructure via NGIs.

Computer Science & Engineering
Each partner will choose one coordinator. Partner coordinators form the VRC Management Board (MB) that is responsible for the effective and timely achievement of the VRC goals as described in the WPs. The MB will meet on a monthly basis. The MB will nominate a VRC Chair who will coordinate its activities and will represent the VRC in all instances where the CSE VRC should be represented (including the EGI User Forum). The major tasks are of the MB are as follows.

Resource Management
The usage of the Grid Observatory data and services is open to the whole scientific community, and not limited to the CSE partners. In the course of the project, the usage of grid resource is not expected to raise major issues, as the adequate resources have already been provisioned. The evolution of the requests for other kind of resources (notably gateway software and hardware infrastructure) will be closely monitored, in order to contribute to investigate and start implementing sustainable solutions.

Expanding the users community
The CSE VRC will act as a catalyst for developing synergies at the European level and beyond, between scientific communities that have had so far limited opportunities to interact. The evolution of the MB by involvement of participants from other bodies representing the community (Autonomic Computing Centre, COST networks and NoEs in Computer Science and Engineering, national projects) is required in the perspective of sustainability. It is expected that a representation of these communities will be included in the course of the project. Addressing interoperability, in relation with both the OGF WGs and the specific projects such as GWA is part of this task.
The strategic and organisational aspects of this integration are part of this task, while the selection of scientific themes is part of the CSE VRC tasks *Engaging the Autonomic Computing community* and *Engaging the Distributed systems community* of NA3.

Liaising with Operations and Middleware development
The CSE VRC requires interaction with EGI and EMI, both at the operational level, in order to keep pace with the general development of the infrastructure, software monitoring resources, operational issues, and to evolve to a sustainable set of services. The Chair will interact with the EGI administration coordinators, the EMI management, and all other VRC coordinators.

Transition towards a sustainable model
Building a credible sustainability plan for the Grid Observatory is a major task for the CSE VRC, as the research community must be convinced of the perennity of the data source in order to invest the effort required to exploit them. The challenging issue is that the CSE community has no international operational body comparable to CERN or ESA, or flagship projects such as those recognized in the ESFRI roadmap. The closest approximation would be EGI itself, but it covers only the engineering segment. The directions in which the sustainability plan will be investigated are a combination of the following.

- NGI support, concerning resources with the French NGI TGIR (Très Grands Instruments de Recherche) roadmap [Ref. 24]; support from other NGIs concerning the creation of a network of experts.
- EGI as the primary data provider.

Setting the Grid Observatory as a sufficiently reputable product to attract sponsoring and be supported by the national research programmes.

Complexity Science
Project coordination and liaison with other VRCs and EGI is an essential task for the VRC. The coordination of the Complexity Science VRC will be performed by the CS VRC Steering Committee that will include one representative from each European partner of the VRC. The members of the CS VRC Steering Committee will also be responsible of the liaison of the CS VRC and the corresponding NGIs.
The Steering Committee will select a User Forum Representative, who will actively participate in the User Forum Steering Committee of EGI, and a Grid Planning Officer whose responsibility will be to provide a more long term technical planning and who will represent the VRC in the Middleware Coordination Board.
To ensure that the required progress is being made at all times during the course of the project the VRC will hold regular monthly meeting conference calls in which the key personnel from each work package will participate. In addition, the organization of regular face-to-face meetings is also needed to ensure that Project aims are met.
AUTH will be responsible for the overall coordination of the Complex Science VRC and for organizing a monthly conference call focused on the progress of the CS VRC activities BIU, UNIPA, JLUG, UA and SU will in rotation organize and host a CS VRC face-to-face meeting.

Photon Science

Coordination of user relevant services for light sources; Communication with informal PS community organizations.
The typical users of light sources are using more than one facility. Therefore it's of great interest to harmonize access on all levels. Covering all the relevant issues is going too far for this project but having common and coordinated access methods to data and resources is very beneficial.

Communication with middleware providers, e.g. EMI etc; Communication with EGI
As each community, the Photon Science community has in particular with respect to the upcoming ESFRI projects different or special requirements for the middleware, for tools and for

infrastructure in Europe like the networking. These requirements have to be communicated in a professional way to the relevant bodies. Photon science will actively participate and contribute to the EGI meetings and boards, if required.

Humanities

The Humanities VRC's main focus will be on community building and engagement, with the aim of becoming a hub for the nascent humanities EGI community. Its management and coordination activities, and links with other VRCs, will be organized accordingly. A management group will be formed comprising of representatives of all the partners. One member of this will represent the VRC on the user forum's programme committee. Meanwhile, strong links with the existing Digital Humanities (DH) community (see www.digitalhumanities.org) will be established, building on the partners' existing relationships. The VRC's main focus will be to present the opportunities arising from a European infrastructure towards the DH community. H-VRC partners include the ESFRI projects DARIAH and CLARIN, and it will provide a point of focus for these organizations' EGI and other grid activities. The VRC will also ensure close liaison with relevant professional bodies such as the Alliance of Digital Humanities Organizations (ADHO), which has a very strong historic Europe-wide component, but also excellent links to comparable bodies outside the EU, especially the USA. The VRC will represent EGI at DH events, such as the annual international Digital Humanities conference (http://www.cch.kcl.ac.uk/dh2010/). The VRC will develop and maintain a list of organizations and groups in the humanities with *potential* links to EGI, and ensure that appropriate cross-representation is provided.

In summary, all Humanities VRC partner institutions have conducted an active policy of community engagement from the very beginning, and the VRC's coordination strategy will be to build on these. It will develop exemplar materials that will engage other communities of users - including those who are enthusiastic but lack the resources and/or the knowledge to engage fully with e-Infrastructure tools and methods.

Deliverables (brief description and month of delivery)

Probably quarterly or semi-annual reports including the technical and scientific impact of the VRCs is sufficient. These reports should also include standard metrics if these are not provided automatically elsewhere.

Table 6: Dissemination & Training (NA3)

Work package number	NA3	**Start date or starting event:**		M03	
Work package title	Dissemination & Training				
Activity Type[12]	COORD				
Participant number					
Participant short name					
Person-months per participant:					

Objectives

- Coordinate dissemination activities
- Disseminate VRC activities within the targeted VRC communities
- Coordinate training activities
- Coordinate participation in general training events
- Provide focused training events for VRC communities
- Coordinate VRC participation in EGI User Conferences
- Organize ROSCOE participation in strategic non-EGI conferences

Description of work (possibly broken down into tasks) and role of participants

Dissemination and training are two key areas for attracting new users to EGI and ensuring that they can use the grid infrastructure effectively.
The dissemination activities will target each VRC's user community and the general public, respectively "internal" and "external" dissemination. EGI.eu, the management structure of EGI, will coordinate dissemination activities for EGI as a whole. To avoid duplicating effort, ROSCOE will piggyback its "external" dissemination activities on those of EGI and use EGI.eu personnel to coordinate the "internal" dissemination activities. The dissemination activities within EGI are expected, in turn, to take advantage of EGI-related projects and the NGIs.
The internal dissemination activities must necessarily be specialized for each VRC, targeting important conferences within the community and publicizing the scientific accomplishments taking place on the grid. Consequently, the effort for this is embedded within each VRC with cross-VRC coordination done by EGI.eu, the lead partner.
The situation with training is similar with EGI.eu acting as the overall coordinator and acting as the liaison with EGI-related projects and NGIs for generic training materials and effort. The effort embedded within the VRCs will augment the generic training courses with specialized information for the targeted community, for example, giving examples of common use cases within the discipline or showing how domain-specific data repositories can be accessed.
Visibility of the ROSCOE VRCs at the annual EGI User Conference and at strategic non-EGI conferences identified by the VRCs is critical for attracting new users to the grid infrastructure and making people aware of the scientific advances facilitated by the grid infrastructure. The lead partner will coordinate participation in these events, handling the logistical details and ensuring that VRC dissemination contacts provide relevant materials.
A description of the training plans must be given here. The general idea will be to plan one event a year where all of the VRCs will participate. It is expected that EGI (or a delegate) will provide the general training and the VRCs will provide domain-specific training. The

[12] Please indicate one activity per work package:
RTD: Research and technological development; COORD: Co-ordination; MGT: Management of the consortium;
SVC: Service activities

structure of these meeting still need to be determined. Because the EGI User Conferences are likely to be hectic, these will be scheduled at a time other than those conferences.

A description of the dissemination and training "pot of money" needs to be provided. This fund will pay for general and VRC-specific dissemination materials. It will also provide (limited) travel money for inviting speakers and participants to the EGI User Conference and VRC meetings. The travel money will only be used for people OUTSIDE of the project. This activity will also need a travel budget for attending conferences and for renting of booths and conferences.

Add a statement that scientific gateways are critical for dissemination. This is so important that it is treated in a separate activity.

Phases

Phases: setup, dissemination/training.

Common Tasks

Organization of EGI User Conference participation; common training events.

Please rewrite each of the VRC sections keeping in mind the goals of the activity and the general strategy. For dissemination, you should identify important groups within the community to reach, strategic projects, and important conferences. For training, provide examples of how the generic training would be specialized for your domain. Also provide important training events geared to your community to which you will contribute. The length of each VRC section should be 0.50-0.75 pages.

High Energy Physics

HEP and partners have a long experience of dissemination of their grid activities via conferences and schools. The principal events are the twice yearly HEPiX (HEP Unix users) meetings, the Computing in High Energy Physics conferences held each 18 months, the annual Advanced Computing and Analysis Techniques in Physics Research conference and the successors to the annual EGEE and EGEE-User Forum meetings.

In addition to these "internal" events there will be periodical external events targeted at a wider public. Recent examples of these were a half-day LHC 'Grid-fest' event to publicise the use of the LHC Computing Grid to which scientific journalists were invited and demonstrations of the grid in action at Telecom 2009. CERN also provides regular input into e-zines such as iSGTW.

CERN is also involved in numerous training tasks including sessions (for both users and site administrators) at the CERN Computing School, User Forums and EGEE conferences and invited training sessions in regional events and international conferences. The Grid support team regularly provides external training events for the setup of generic Grid services. Underlying such training CERN maintains the gLite 3 User Guide.

Much of this work will be coordinated with that of other VRCs with contributions from other partners on specific aspects (e.g. Distributed Analysis Training).

CERN will coordinate the HEP contribution to this activity (0.5 FTE, unfunded).

Life Sciences

One of the major goals of the LSVRC is to enlarge the community of grid users in life sciences. Therefore dissemination and training are crucial tasks. As shown below, the Life-Science VRC dissemination and training activities aims at answering the general goals of NA3 in the field of LS with a special concern for addressing the digital divide in order to foster the use of grids for medical developments.

NA3 objective 1: "Coordinate dissemination activities"

- Reports of LSVRC events and technical notes will be published in EMBnet.news [Ref. 1]
- Dissemination activities will be advertised on the LS gateway operated by SA2.

NA3 objective 2: "Disseminate VRC activities within the targeted VRC communities"

- Dissemination to the research communities involved in ESFRIs (BBMRI, IN-STRUCT, ELIXIR, LIFEWATCH)

NA3 objective 3: "Coordinate training activities"

- Coordination with EGI training VRC
- Two virtual lecture rooms and an e-learning facility [Ref. 3] [Ref. 4] will be provided by EMBnet to support training activities.
- LS training activities will be advertised on the LS gateway operated by SA2.

NA3 objective 4: "Coordinate participation in general training events"

- Organization of yearly LSVRC events in conjunction with other events
- The training infrastructure will be provided by CSC (providing a PC classroom with video streaming facilities already used by many training events [http://www.csc.fi/English/csc/courses/archive] and INFN (providing its GILDA t-Infrastructure [Ref. 5] used so far with success in many grid training events, some of them specific for the life science community in projects like BioinfoGRID [Ref. 6] and e-NMR [Ref. 7], or for the regional projects like EELA [Ref. 8] and EELA-2 [Ref. 9], EUAsiaGrid [Ref. 10], EUChinaGRID [Ref. 11], EU-IndiaGrid [Ref. 12] and EUMEDGRID [Ref. 13]). INFN can also provide as unfunded contribution its virtual training facility [Ref. 14] [Ref. 15] that has already been successfully used in the International Winter School of Grid Computing 2009 (IWSGC'09) [Ref. 16].

NA3 objective 5: "Provide focused training events for VRC communities"

- Organization of yearly biomed Grid schools
- Coordination with regional grid infrastructures for organizing biomed grid schools, including in developing countries
- Integration of grid tutorials into EMBnet training events
- Production of EMBnet Quick Guides [Ref. 2] on components of the LS gateway.

NA3 objective 6: "Coordinate VRC participation in EGI User Conferences and organise ROSCOE participation in strategic non-EGI conferences"

- Contribution to EGI User Forums
- Joint demos and booths with regional grid infrastructures

LS partners involved in this work package are CNRS (6PM), HealthGrid (6PM), HES-SO (5PM), UPV (4PM), CSC (1PM), CERM (2PM), CNR (12PM), INFN (6PM), EMBnet (30PM), AMC (4PM), EBI (2PM) and UMAN (5PM).

Computational Chemistry & Materials Science Technology

Dissemination is meant to focus on scientific communities promoting the VRC achievements as possibility to better present the result of research. Trainings is meant to be delivered in a form suitable for active characters that would prefer webinars, instruction movies, on-line training materials on predefined data, etc. than standard trainings.

Training will be given in a different format with respect to standard EGI training events. Users involved will be stimulated to work closely to tutors and solve their own problem on the GRID during the tutorial itself.

For this reasons the various partners and in particular the Democritos training team will make available during the tutorials the software tools for e-science development and grid enabling procedure in the area of its competence including:

- Simulation software (quantum-espresso and related)
- Modern software developing tools (like gforge portals: qe-forge.it and

gforge.escience-lab.org).

- el3 and elbas benchmarking suite to asses performance on hardware platforms dedicated to material science simulations

Special attention will be paid to prepare specific training materials that should be ideally formed and provided by satisfied GRID users ready to share their experiences with their peers.

Dissemination plays a crucial role in every Research and Technological Development project. CCMST VRC focus will be both on promoting project results in general and on bringing these results to community members. Therefore it is crucial for community members to present results of both scientific and technical achievement not only during EGI User Conferences but also during strategic workshops and thematic conferences covering scientific areas our community. A key prerequisite is also a project portal (see more details in SA2 activity description) containing public sections with training and dissemination materials. Also to reach vast majority of community members actions will be taken to promote VRC via widely utilized free social networking services like Facebook or Twitter. This seems to be one of the most effective ways to promote grid technology among young researchers and students.

Computer Science & Engineering

Dissemination will focus on two aspects.

Promoting the VRC services and achievements in the wider scientific community

The usual instruments will be used: scientific publications, participation in conferences, demos/booths and contributions to. Major target events are the EGI user forum, and the SC (supercomputing), CCGRID and ICAC conferences. The organization of the CSE gateway will be designed so that it can also be exploited as a central instrument for internal and external dissemination (e.g. database of publications acknowledging ROSCOE, announcement of events, etc.).

Engaging the Community

The first action will target primarily the NSF Center for Autonomic Computing, and the corresponding European networks. The focus is the cross-fertilization between AC and production grids. The goal is to demonstrate the effectiveness of autonomic concepts for end users who utilize the grid infrastructure. The major issue is to define a process that could be an equivalent of undertaking the “paper-in-Nature” test, whereby a scientist could claim that through the use of autonomic computing techniques they were able to reach a particular scientific insight that they could not have obtained otherwise. The process must focus on effective dissemination of such benefit to get community commitment. Cardiff University will lead this subtask. The distributed systems community will be engaged primarily through interactions with the CoreGrid/ERCIM WG (follow-up of the CoreGrid NoE). The goal is to create a process for gathering requirements from the Distributed Systems community toward the GO, both concerning the nature and presentation of the traces.

CNRS/LRI and CU will contribute to the Dissemination Plan of the CSE VRC.

Complexity Science

Dissemination of success stories and best practices is a critical task in the course of the Project, as it is needed to enlarge the scientific community and strengthen it. Complexity Science VRC partners will constantly seek to publish the status and progress of the Project results through their participation in EGI User Conferences (expected to be organized in the context of EGI) and Complexity Science conferences.

AUTH, BIU, UNIPA, JLUG, UA, and SU will contribute to the Dissemination Plan of the Complexity Science VRC

BU will also contribute to the liaison of the CS VRC with the US Complexity Science community

Three training events specifically targeted to the CS community will be organized and delivered so that new user communities are identified and incubated, thus ensuring that the sustainability criteria set out in the start of the project are met. During these training events focus will be given on the Complexity Science VRC services and the tools developed in its context. Thus a deep insight of the services provided (see SA related tasks) and the related tools developed (see JRA related tasks) will be given to attendees. The trainings will in addition be recorded in the form of webinars and this web content will be put online through the Knowledge Base (see SA1 for further information).
UNIPA, JLUG and SU will organize and deliver the CS VRC Training Events.

Photon Science

Dissemination and Training for PS communities
It is assumed, that there is a generic facility for dissemination and training available in the EGI context. In particular basic training courses in grid usage, potentially on site are expected as a service provided by EGI or the relevant NGI's. In this work package only photon specific aspects of dissemination and training are covered, this means in detail an additional course bridging the gap between the scientist and the grid-experts. This can be either a stand-alone course or be part of a general course.
This task will be accompanied by dissemination into the user communities, training, workshops and tutorials, and documentation services. Part of the task will be the participation and contribution to workshops and conferences of the targeted user communities. The user meetings held by light sources (e.g. the annual HASYLAB user meeting) or PS schools like HERCULES at ESRF or EMBL-schools provide existing platforms to implement a significant part of this task.
To advertise opportunities, support entry points and to supply support documents for PS communities a PS VRC web-portal will be created.

Liaison with pan-European initiatives and projects
Dissemination to other PS facilities, beyond the partners of the VRC, will be reached through cross-facility participation in several ESFRI projects (EuroFEL, ESRFup, INSTRUCT, XFEL) and the cooperation of all major light and neutron sources within the pan-European PanData project. Liaison with pan-European initiatives and ESFRI-projects like those mentioned will be integral part of the dissemination activities.

Humanities

Objectives

- Disseminate the Humanities activities within the EGI community
- Communicate effectively the implications and benefits of engaging with EGI to the digital humanities community
- Maintain dissemination web site
- Promote interactive community resource

The broad objective of this work is to raise awareness of Humanities e-Infrastructure activities and promote dialogue between (potential) stakeholders. This work package aims to get directly in touch with the individual researchers, engage them through exemplars and network scholars with existing e-Infrastructure activities in the humanities.

Selecting which elements of the humanities community to target with dissemination, and how, is a non-trivial task. It is too simple to assume that humanities as a whole is an e-Infrastructure 'early adopter community'. Many parts of the humanities are already actively engaged with e-Infrastructures if not directly with its applications in e-Science. Four categories of humanities scholars will be explicitly targeted with appropriate levels of effort, which are not mutually exclusive, but can be used to understand, at a high level, support needs and activities:

1. Established Humanities Communities
These are researchers already active and confident in the use of e-Infrastructure tools and methodologies. Examples of such projects can be found http://ahessc.ac.uk/escience-projects Most A+H researchers will not use e-Infrastructures directly, but the established research communities will be involved in larger multidisciplinary projects. Their needs are therefore for a combination of technical and domain-specific support. We will produce a set of case studies and briefing papers pitched at a high level, with a strong emphasis on making this community aware of relative activities and services within EGI, combined with targeted technical support and liaison with EGI when necessary and appropriate – which is likely to be when individual projects have become established and have fully scoped their data requirements (the VRC will not of course have the expertise or capacity to do this itself). This activity will therefore feed into ROSCOE's broader knowledge-gathering programme.

2. Emerging Humanities communities
Most of the Humanities researchers not centrally involved in the e-Infrastructure projects as PI's belong to this group. They are beginning to engage with e-Infrastructures, but support to foster the establishment of tools and environments will accelerate their uptake. These researchers are often best supported by offering basic advice and working with them on further proposals towards funding their particular interest. We see particular potential for further engagement of the established Digital Humanities community with the e-Science and e-Research agenda.

3. Interested but not yet engaged
These groups have not hitherto been involved in e-Infrastructure work. In the Humanities, these are either those which have had some exposure to ICT methodologies to support their research, but have not used research infrastructures, or those which have had no exposure at all to ICT. The first group includes, for example, those projects from the national Digitization schemes that have digitized an existing (analogue) resource. The second group includes those researchers, which do not use any kind of online research data to conduct their research.

4. Unengaged
This, the largest section of the academic community, comprises of researchers who have had no contact with ICT, whether in the Humanities or elsewhere. Our materials are widely available for anyone who is potentially interested, and we seek to ensure that messages of potentially broad appeal are broadcast as widely as possible; and that the VRC's knowledge base is authoritative, trusted and reliable.

Given this, any user support for the Humainists VRC proposal will need to be carefully targeted. The VRC will responsively map the 'engaged' communities needs by identifying key 'themes' of interest to the humanities that e-Infrastructure has significant potential to support via Virtual Research Communities, and focus on these. These include, but are not limited to, Geographic Information Systems and geo-temporal computing; mining and information retrieval across huge text corpora, simulation and predictive (or postdictive) modelling of past societies, and the management and storage of very large collections of image and multimedia objects. The results of this work will be formatted and structured in such a way as to feed in to the appropriate sections of ROSCOE's user support knowledge base. Dissemination and publication activities will be developed around these themes. Where appropriate, these themes will be further mapped onto disciplines beyond the humanities (e.g. predictive modelling is a methodology originating in the Social Sciences: obviously we would wish to have close liaison with this community in this particular area). Again we will contribute to, and rely on, the ROSCOE knowledge base to develop and extend this body of information.
Text is still widely pervasive as a research data type in the humanities, although other forms

of digital material are becoming increasingly important; and the H-VRC's user support activities will reflect this. Texts are used and produced in ever growing amounts and are increasingly available in a digital format. No researcher alone will be able to cope with the plethora of published material. Furthermore, text analysis in the humanities can be a tedious and time-consuming task. But advanced computer-enabled methods make the process easier for digital or digitised works. Researchers can search large texts rapidly, conduct complex searches and have the results presented in context. The ease brought to the analysis process allows the researcher to engage with texts more thoroughly and can then lead to the development of insightful, well-crafted interpretations. We will look into mapping the landscape of text mining services and how they might fit into a European infrastructure not only for Humanities. Next to text various other mining technologies such as geo-mining have proven their value for Humanities.

Deliverables (brief description and month of delivery)

Need to have deliverables related to collecting to public relation contacts, dissemination plan, strategic conferences, etc. Probably need quarterly or semi-annual reports on dissemination and training events as well as metrics.

Table 7: Summary of staff effort

Participant no./short name	NA1	NA2	NA3	person months
Total				

Table 8: List of milestones

Milestone number	Milestone name	Work package(s) involved	Expected date[13]	Means of verification[14]

1.7.4. Graphical presentation of component interdependencies

Provide a graphical presentation of the components showing their interdependencies with a Pert diagram or similar.

1.7.5. Significant risks and associated contingency plans

Table 9: Risks for Project Management (NA1)

Risk NA1	Impact	Occurrence Probability	Mitigation

Table 10: Risks for VRC Coordination (NA2)

Risk NA2	Impact	Occurrence Probability	Mitigation

[13] Measured in months from the project start date (month 1).
[14] Show how you will confirm that the milestone has been attained. Refer to indicators if appropriate. For example: a laboratory prototype completed and running flawlessly; software released and validated by a user group; field survey complete and data quality validated.

Table 11: Risks for Dissemination & Training (NA3)

Risk NA3	Impact	Occurrence Probability	Mitigation
Low participation in the CS VRC Training Events	Enlargement of CS community will be hindered by low interest	Medium	In order to make CS VRC Training Events attractive to the wider Complexity Science community we will have to focus and disseminate upon success stories.

1.8. Service Activities and associated work plan

Describe the extent to which the activities will offer access to state-of-the-art infrastructures, high quality services, and will enable users to conduct high quality research.
A detailed work plan should be presented, broken down into work packages (WPs), which should follow the logical phases of the implementation and provision of the project's Service Activities, and include assessment of progress and results.

1.8.1. Overall strategy of the work plan

(Maximum length — one page)

Grid technologies and grid infrastructures have the reputation of being difficult to use; this reputation is well deserved. All of the ROSCOE service activities are designed to lower the barriers to using the grid infrastructure for our Virtual Research Communities (VRCs). The three service activities each address one facet in the program to make grid technologies and the grid infrastructure more accessible.

EGI will provide user support activities that deal with issues related to operational problems—unresponsive services, lost data, etc. The User Support (SA1) work package of ROSCOE will augment this to provide support for questions related to how to use grid technology or to use resources specific to a given community, for example how to access a particular public data store. This work package will be structured to provide a common team for initial triage of tickets and resolution of simple ones. Individual teams for each VRC will respond to discipline-specific issues.

All of the scientific communities have identified a Scientific Gateway as critical for building an identity for a VRC and for communicating with individual researchers within their communities. In addition, many want to use the Scientific Gateways as a simplified mechanism for accessing grid resources, effectively lowering the barrier to adoption by hiding the complexity of the grid infrastructure. These gateways also offer the opportunity to federate community resources outside of the grid with those inside. The Scientific Gateways (SA2) work package will maintain those gateways, determine common requirements, and transition to a common implementation as it becomes available.

The scientific analyses within a given community tend to use a common set of tools and techniques. Porting those common tools or applications using common techniques can therefore have a large, immediate impact on the grid utilization of a community. The Targeted Application Porting (SA3) work package will identify flagship applications, common frameworks, and common libraries in use within the community and port them to the grid. Multiple communities use many of the tools and libraries; this offers an opportunity for additional collaboration between disciplines.

All of these activities also work to cultivate the grid expertise within the community and to safeguard that knowledge. This complements the dissemination and training activities of NA2 ensuring that both novice and experienced members of the community are supported by the VRC.

1.8.2. Timing of the different WPs and their components

(Gantt chart or similar)

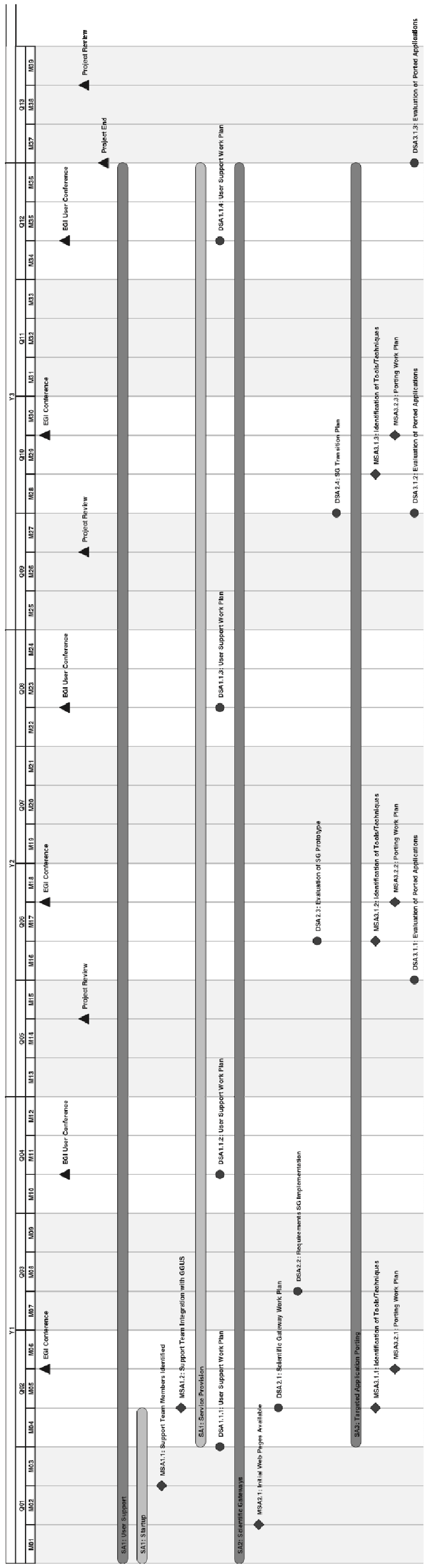


Figure 6: Gantt Diagram for Service Activities

1.8.3. Detailed work description broken down into work packages

Table 12: Work package list

Work package No[15]	Work package title	Type of activity[16]	Lead Part. No[17]	Lead Part. short name	Person / months[18]	Start month[19]	End month
SA1	User Support	SVC		CNRS		M01	M36
SA2	Scientific Gateways	SVC		CNRS		M01	M36
SA3	Targeted Application Porting	SVC		CNRS		M01	M36
	TOTAL						

Table 13: Deliverables list

Del. no.[20]	Deliverable name	WP no.	Nature[21]	Dissemination level[22]	Delivery date[23] (proj. month)

[15] Work package number: WP 1 – WP n.
[16] Please indicate one activity per work package:
RTD = Research and technological development; COORD = Co-ordination; MGT = Management of the consortium; SVC = Service activities
[17] Number of the participant leading the work in this work package.
[18] The total number of person-months allocated to each work package.
[19] Measured in months from the project start date (month 1).
[20] Deliverable numbers in order of delivery dates. Please use the numbering convention <WP number>.<number of deliverable within that WP>. For example, deliverable 4.2 would be the second deliverable from work package 4.
[21] Please indicate the nature of the deliverable using one of the following codes:
R = Report, **P** = Prototype, **D** = Demonstrator, **O** = Other
[22] Please indicate the dissemination level using one of the following codes:
PU = Public
PP = Restricted to other programme participants (including the Commission Services).
RE = Restricted to a group specified by the consortium (including the Commission Services).
CO = Confidential, only for members of the consortium (including the Commission Services).
[23] Measured in months from the project start date (month 1).

Table 14: User Support (SA1)

Work package number	SA1	**Start date or starting event:**		M01	
Work package title	User Support				
Activity Type[24]	SVC				
Participant number					
Participant short name					
Person-months per participant:					

Objectives

- Create and maintain targeted documentation
- Provide support concerning use of the grid infrastructure
- Provide user support for domain-specific services and applications
- Provide intensive debugging support for operational problems
- Contribute to the treatment of user support tickets
- Investigate novel mechanisms for providing user support
- Support for the evaluation and testing of pre-production services

Description of work (possibly broken down into tasks) and role of participants

The grid remains a complex, distributed system and its effective use requires dedicated user support at many levels.
Where appropriate the VRC will maintain documentation targeted to its user community, concentrating on domain-specific applications, techniques, and data repositories. Often documentation by itself is insufficient so the user support teams will provide help with using the grid services to accomplish scientific analyses.
Operational problems on the grid can be difficult to trace especially for those scientific disciplines that have extensive analysis frameworks built over the grid middleware. In this case, the user support teams will help with the detailed debugging of operational problems to determine where the problem lies and to follow up with site managers or middleware providers to ensure a fix. The intensive debugging also builds expertise within the community to help it become for self-sufficient.
The VRCs will use the standard EGI ticketing system to track problems and the user support teams will appear as support teams within that system. As contributors to that system the user support teams will solve tickets when possible or route tickets to other appropriate support teams.
Often ticketing systems and email are too limiting to provide effective, quick user support. The user support teams within ROSCOE will collectively investigate providing user support through novel interfaces such as chat, VoIP, or videoconference. Similarly alternate types of documentation such as podcasts, webcast, video will be tried to see if they can improve the user experience.
The description of testing of pre-production services is missing.

Phases

Phases: setup, provision of support.

---

[24] Please indicate one activity per work package:
RTD: Research and technological development; COORD: Co-ordination; MGT: Management of the consortium;
SVC: Service activities

Common Tasks

The common activity for this work package will be the development and maintenance of a knowledge base. There will be common meetings to ensure that a plan is developed for new content, that existing content is reviewed, and that a logical organization is maintained.

Each VRC coordinator must rewrite their contribution keeping in mind the objectives of the activity as well as the overall strategy.

High Energy Physics

The work will be organised around a number of tasks that are fully consistent with the overall goals of this work-package and include: (stress) testing of new middleware features and functionality in pre-production environments according to the supported VOs' requirements; providing general and specialised grid expertise for the rapid resolution of problems; development of experiment-specific add-ins/plug-ins to operational tools; support for the integration of experiment-specific critical services into the overall grid infrastructure; investigation and deployment of solutions to enable an effective user-to-user and user-to-expert support model.

Operational support for all of the tasks above for the LHC experiments will be provided by CERN (2 FTEs, co-funded), INFN (2 FTEs, co-funded) and UIO (0.5 FTEs, co-funded).
GSI will provide operational support for all of the above tasks for the FAIR community (2 FTE co-funded)
DESY will provide operational support for all of the above tasks for the ILC community (1 FTE co-funded)
CESNET will investigate generic and sustainable implementation of LHC data analysis (Tier3) support (1 FTE, co-funded)

Life Sciences

Users in the biomedical community range from technology experts porting applications and developing specific services to scientific researchers purely using the tools. The objective of this task is to guide, ease and assist users on efficiently exploiting the infrastructure, in order to achieve a large adoption of scientific gateways both from users and service providers in the life sciences user community and promoting and encouraging the use of grid-enabled bioinformatics and medical informatics services in the research community. This activity will focus on collecting and structuring information and providing first line user support for the access to VRC services in collaboration with EGI user support teams. As shown below, tasks of the LSVRC address the general goals of SA1.

SA1 objective 1: "Create and maintain targeted documentation"

- A knowledge base will be setup. It will contain references to other general-purpose documentation sources about grid application porting. In addition it will develop new use cases based on the specific scenario and requirements of LS. This subtask will also generate a list of requirements coordinated with other VRCs for driving developers on new generation components.

SA1 objective 2,3, 4 & 5: "Provide support concerning (i) use of the grid infrastructure, (ii) domain-specific services and applications and (iii) operational problems" and "contribute to the treatment of user support tickets"

- Tickets on unexpected behaviour, failures or usage doubts are a very powerful tool to address critical operational issues. Advanced members of the LSVRC will contribute to the treatment of those tickets to help particular users and to contribute to the knowledge base.

SA1 objective 6: "Investigate novel mechanisms for providing user support"

- Ticket-based support request has been inefficient in many past experiences, mainly due to the lack of organisation and reward. This task proposes reducing those barriers by means of creating an explicit list of expertises and people, the creation of the figure of the ticket dispatcher, according to this expertise, and the implementation of a rewarding mechanism for the most active ticket-solvers with, for example, covering the registration to conferences in the field (such as Healthgrid).

- LS partners involved in this workpackage are HealthGrid (1PM), UPV (18PM), CSC (4PM), AMC (3PM) and UMAN (3PM).

Computational Chemistry & Materials Science Technology
The knowledge concerning grid computing and its application in chemistry and materials science and technology is unequally structured among community members. Thus certain tasks are required to fulfil different community needs at different levels. At first VRC supporters will focus on gathering and completing of an existing documentation spread among partners. To ensure wide availability and easy access to the documentation a single point of entry at CCMST gateway is planned.
Two main areas of User's Support can be distinguished for CCMST: technical and direct respectively.
Technical Support will deal with all the services and tools related to CCMST needed to keep the infrastructure ready for utilization by community members. Duties performed within tasks will deal with intensive debugging of accidentally occurring operational problems related to CCMST discipline. Specialized teams of experts will contribute to the treatment of user support tickets. Special attention will be devoted for provision of application specific as well as service specific automated validation tests integrated with EGI operations tools. Coordinators oft these tasks will also closely cooperate with EGI User Technical Support Group as well as with middleware developers for the evaluation and testing of pre-production service useful for CCMST community.
Direct Support will be responsible for interactions with users and dealing with their accidently occurring problems. In the past we have found support services like GGUS very useful however, in many cases direct interaction with specific user is needed. For this we will investigate novel mechanisms of providing user support via widely accepted technologies like live chats via frequently used protocols (i.e. jabber or ICQ) as well as via user forums.

Computer Science & Engineering
User support for the CSE VRC has two faces. The first one is related to the usage of the gateway, and will be assured by CNRS/HG. Given the fact that the community is experienced in computer technology, this activity will be limited to the interaction with the overall support system (ticketing and possibly more advanced tools) for issues related to the gateway interface with EGI. The second aspect is documentation, and will be assured by CNRS/LRI. Documenting both the data organization and the analysis facilities is essential for facilitating expert usage of the gateway.

Complexity Science
The Complexity Science VRC will set up a specialized support team that will provide user support services to the wider Complexity Science community. The Support Team will make use of the helpdesk infrastructure provided by the EGI, namely the GGUS ticketing system, and will act as a support unit specific to this VRC. The main task of the Support Team will be handle trouble tickets coming from the Complexity Science users, to provide answers and fixes to user questions and problems and to escalate requests to the other appropriate units within the helpdesk service, whenever such an action is needed (i.e. a problem or query is too generic to be considered CS VRC specific or is out of the scope of the CS VRC Support Team). The Support Team will also provide answers to queries related to application porting so that best practices are applied. Overall the CS VRC Support Team will aim to deliver a high quality service by responding quickly to user queries, by solving application porting related problems (not just providing quick fixes) and by being proactive and predictive with respect to the user needs

In addition, to solving specific user or application related problems the Support Team will also maintain the Projects Knowledge Base making sure that related material is up-to-date and that new CS VRC services are properly documented. Solutions to user or application related problems that will be specific to the CS VRC community will also be collected and archived in the Projects Knowledge Base.
AUTH will supervise the Support Team Operation and BIU along with UA will participate in the CS VRC Support team.
The CS VRC will be responsible for managing and maintaining the information stored under the VOMS interface(s) supporting the Complex Systems VO(s). Thus the CS VRC will control registration and removal of physical entities with/from the VOMS interface. In addition, roles and attributes of the CS VO(s) on the VOMS interface will be determined and controlled by the CS VRC VO Manager(s). The VO Manager(s) will in addition be responsible for the definition and the maintenance of the policies related to the VO resources usage. AUTH will lead this sub task.

Photon Science

End-User support for PS communities
The PS user communities are extremely volatile users. A large fraction of the researchers performing experiments at light source facilities are first time user, being novices to the instrument as well as the Grid. User support is hence an essential and ongoing effort. Planned tasks involve

- Investigation and deployment of tools which enable effective interaction between facilities, users and experts.
- Most facilities have an in-house support infrastructure like an issue tracking system and most communities have their own bulletin boards to post issues specific to the community. There is however no way to exchange between facilities and/or communities and no direct integration of the GGUS system. Interfacing between the systems will improve the user experience and is essential for users performing analysis in a Grid environment.
- Coordination of support providers, namely experts from the VO taking responsibilities for specific user communities.
- Coordination of general and VO-specific training for end-users and support providers.
- Investigate novel mechanisms for providing user support, setup a moodle (http://moodle.org) based virtual learning environment.

Humanities

Deliverables (brief description and month of delivery)

Table 15: Scientific Gateways (SA2)

Work package number	SA2	**Start date or starting event:**			M01	
Work package title	Scientific Gateways					
Activity Type[25]	SVC					
Participant number						
Participant short name						
Person-months per participant:						

Objectives

- Operate existing portals during evolution to accepted portal implementation(s)
- Maintain documentation, information, and news on VRC portal
- Ensure that the VRC portal functions effectively for the target community
- Extend functionality of VRC portal to meet needs of the community
- Manage transition to common scientific gateway implementation

Description of work (possibly broken down into tasks) and role of participants

Each VRC has identified that a Scientific Gateway is important for coordinating activities within the targeted community, informing the community of events, and providing access to the grid infrastructure. A gateway should encompass the following functionality:

- Documentation, information, and contacts
- Events/News
- Monitoring view of activity within the VRC/VO
- Monitoring of services
- Access to and management of data
- Access to grid services

Although the scope for each VRC will be different depending on the needs of its community. EGI in cooperation with the NGIs is expected to operate the Scientific Gateway machines. Another project, EGI-SGI, will analyze requirements for the Scientific Gateways, analyze existing portal implementations, and work towards convergence to a single implementation or a few implementations.

VRCs with existing portals will continue to operate them until they can be migrated to the common accepted implementation(s). All VRCs will maintain documentation, information, and news that reside on the portal. The VRC will also ensure that the gateway functions well for the community and meets its needs. When used to access grid resources for an VRC, the VRC may need to develop specialized plug-ins to allow access to domain-specific resource or data. Where possible those developments will be as general as possible to permit reuse by other communities.

Phases

Phases: initial setup, operation, requirements gathering, evaluation, transition.

Common Tasks

Requirements gathering for common implementation(s). Common evaluation of new imple-

[25] Please indicate one activity per work package:
RTD: Research and technological development; COORD: Co-ordination; MGT: Management of the consortium;
SVC: Service activities

mentations.

Each VRC coordinator must rewrite their section keeping in mind the objectives of the work package and the general strategy. The description for each VRC should include a general description of how the scientific gateway will be used in the community and how it will be deployed. If a data repository will be created/maintained, this should be clearly described. The length of each section should be 0.50-0.75 pages.

High Energy Physics

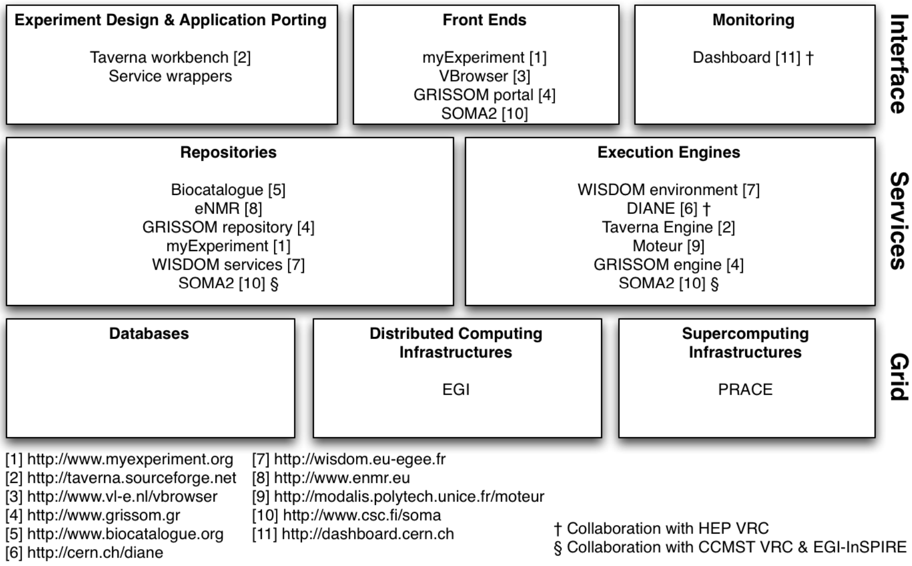
In order to provide a global view of the status of HEP distributed computing, VO-specific workload and data management systems and the generic monitoring frameworks need to be integrated. The goal of this integration continues to be to create a global monitoring infrastructure with dedicated portals providing a complete and reliable picture of the status of the VOs computing activities on the grid and of the status of the distributed sites and services (multi-VO). The general strategy consists of providing common solutions for handling monitoring data flow, from the data producer to the end user. This will reduce development and maintenance effort, decrease time required for enabling of the new functionality, and facilitate deployment of the monitoring systems.

Tasks: Development of experiment-specific plug-ins for existing generic frameworks (e.g. Service Availability Monitor (SAM), Service Level Status (SLS) and Nagios); Enabling common communicating between various components of the monitoring infrastructure via the Messaging System for the Grid based on the Apache ActiveMQ message broker; Handling monitoring data in central repositories (e.g. using an Oracle backend at CERN); Development of monitoring portals serving the needs of various information consumers including the HEP user communities, support teams at the distributed sites and management; Evolution of portals, adapting to new technologies and following user feedback and requirements.

CERN (2 FTEs co-funded]

Life Sciences

A major goal of the LSVRC is to enlarge the community of grid users in life sciences. As shown on the following figure, we propose to build a gateway that will federate services and interfaces used by the various communities (including but not limited to ELIXIR, LIFEWATCH, INSTRUCT and BBMRI ESFRIs) to access grid resources. Scientists of the VRC will be provided with a single interface from which they could assemble, present and run components to design and run experiments. It is crucial that users are able to access grid resources using their preferred tools. Therefore, the gateway interface will be fully customizable to adjust to the various user communities.



The corresponding tasks address the general objectives of SA2 as shown below.

SA2 objective 1: "Operate existing portals during evolution to accepted portal implementation(s)"

- Operation and maintenance of the existing gateway components

SA2 objective 2: "Maintain documentation, information, and news on VRC portal"

- Development and documentation of the gateway interface
- Graphical design of LSVRC web portal and scientific gateways

SA2 objective 3: "Ensure that the VRC portal functions effectively for the target community"

- Requirement collection to ensure adequation to VRC needs

SA2 objective 4: "Extend functionality of VRC portal to meet needs of the community"

- Customization of the gateway components when required
- Design and implementation of service wrappers for application tools

SA2 objective 5: "Manage transition to common scientific gateway implementation"

- Harmonization of component interfaces and sharing of experience to avoid duplication of work and manage transition to common scientific gateway implementation

The various LS partners will ensure that components from their respective communities are integrated in the common gateway. For molecular biology: INFN (3PM), SIB (12PM), UMAN (12PM); structural biology: CERM (9PM) & INFN (3PM); biodiversity: HealthGrid (18PM); healthcare: HealthGrid, HES-SO (3PM); medical imaging: AMC (5PM), CNRS (3PM); genetic population: CNR (4PM); microarray analysis: IBRB (12PM). CSC (2PM) will also be involved in this work package.

Computational Chemistry & Materials Science Technology

Scientific portals are gathering a much attention recently because of their integrating future between different tools, applications and data. Build on top of web interface are a valuable tools especially for new users having no experience with grid technology. The available portals utilized by CCMST community range in their functionality. None of them is however completed and the main goal of this activity is to drive their evolution in to a widely accepted (preferably by most communities) implementation of scientific gateway. As community portal is usually the main entry point for the users it is best place to provide all the news concerning all aspects of grid computing related to CCMST community. In parallel the documentation gathered and maintained by NA3 activity will be integrated with the portal.
The two main functionalities highly required by the community include provision of CCMST applications as a web services together with workflow engine allowing execution of compli-

cated numerical experiments. The successful implementations (including ABCTraj with P-Grade and Chempo with Gaussian, Turbomole and GAMESS) proven high usefulness of such approaches enabling not only easy access to the specific resources by non specialists but also possibility to run batches of jobs during computational campaigns. The further development is needed to extend number of plug-ins for other needed by community software packages. An essential part of work will also cover implementation of specific automated analysis tools needed during computational campaigns as well as integration with visualisation software for better result analysis.
Execution of complicated workflows for CCMST community is second request, which will be of our particular interest. As this kind functionality is required by other VRC we are going to closely cooperate with them to adjust existing implementations to specific CCMST community needs.

Computer Science & Engineering
The core CSE VRC contribution to this WP targets the successful transition from the Grid Observatory portal developed in EGEE-III to

- A comprehensive data curation model;
- Scaling access to much larger communities.

Full-fledged curation involves good data creation and management, including data representations independent of operational formats, and a rich indexation schema allowing versatile presentations along the needs of varied users. These goals will be implemented through three tasks.

Establishing and developing long-term repositories of digital assets for current and future references.
This task scales up the process, started with the GO cluster in EGEE-III of acquisition and long-term conservation of the monitoring data produced by the EGI grid about its own behaviour. The approach EGEE-III GO will be continued: building on the rich ecosystem of monitoring tools developed in gLite and by the users community, as well as the operations team with Nagios deployment. We plan to

- Deploy the acquisition tools as a certified component of the gLite middleware;
- Integrate new sources describing alternative of complementary execution models.

The only exception to this approach targets the acquisition of power consumption indicators for which new sensors must be created. The resulting traces will be of particular interest for interaction with the Cloud computing community.
Raw traces, as provided by the present GO, must be conserved as digital assets. However, their idiosyncrasies and inconsistencies are a serious impediment to large-scale usage. The first necessary step to facilitate usage is to re-structure the datasets according to standards, e.g. the Common Base Event for event-oriented data.
CNRS/LAL and CNRS/LRI will contribute to this subtask.

Providing digital asset search and retrieval facilities to scientific communities through its gateway.
The next and much more ambitious step is building the knowledge representation and management. State-of-the art knowledge management is based on ontology. We will thus:

- Develop a grid ontology;
- Exploit it for data consolidation.

The developed ontology integrates, when appropriate, existing grid ontologies (structural concepts), and will extend them in the directions of new concepts describing grid inputs and dynamics. A by-product is to address the risks for the GO associated with format evolution and obsolescence.
Instead of simple catalogue of concepts (thesaurus) we will propose a heavyweight ontology, whose semantic relations will be the foundations of knowledge discovery. One important application area for this activity is to provide explicit and exploitable foundations for reliable

operation- or user- oriented metrics, and remediation for erroneous/missing information.
The usage of industrial software tools (technology transfer from Mondeca, funded by Région Picardie) ensures the exploitability of the ontology and semantic engines developed along these lines.
UPJV and CNRS/LPC will contribute to this subtask.

Analysis services
Common knowledge in the area of data curation says "Services Make the Repository". Sharing analysis code avoids duplication of effort if the individual components are elementary enough to be easily evaluated by legitimately cautious users. The envisioned architecture is thus a set of elementary functions, and workflow-building facilities. Analysis facilities will initially share codes developed by the community, either within or outside the VRC. The code produced by the VRC will follow high standards of structuring and documenting, in order to engage external contributors to do so. A more ambitious scheme proposing on-line facilities is addressed in SA3.
CNRS/LRI will lead this subtask.

Gateway organization
The gateway involves the design of an ergonomic use interface for data access, a stable architecture allowing general access without grid certificate, and the capacity to build workflows from the analysis services, and the capacity of reporting on its usage. CNRS/HG, which has developed the GO portal and has strong expertise in portal technology, will be in charge of this sub task.

Complexity Science
The CS VRC will develop and maintain a use case repository in collaboration with the CS Support Team. The use case repository will serve not only the needs of new or inexperienced users and researchers but also deepen the knowledge of more advanced users by providing best practices and optimized workflow examples based on the specific needs of the Complexity Science research field. We will base this repository of knowledge on a wiki like interface, thus allowing also users to contribute with their thoughts and ideas as time progresses. Tools and code recipes developed in the application porting procedure (either in the form of source code or as installable packages) will also be available through the use case repository for users to download and directly use.
The use case repository will be a part of the CS VRC Scientific Gateway.
UNIPA and JLU will develop and maintain the use case repository

We plan to design and deploy the CS VRC Registration Portal that will serve as a point of entry for new users. Through this portal we plan to provide all necessary registration forms a user has to complete prior to actually using the underlying Grid resources.
The steps one should complete in order to gain access to the Grid infrastructure, like registering with a Virtual Organization and/or accessing a User Interface will also be provided in the form of modules on top of the CS VRC Registration Portal.
The Registration Portal will be accessible via http://www.complex-systems.eu and its final goal will be to serve users of the CS VRC community as a one-stop-shop mechanism where they will be able to acquire a grid personal certificate, register with a CS Virtual Organization and gain access to a User Interface in just a few steps.
The CS VRC Registration Portal will eventually become part of the CS VRC Scientific Gateway.
AUTH and UNIPA will develop and maintain the Registration Portal.

We plan to develop a VO registration module that will be used as a front-end mechanism for the CS VO(s). This module will be subsequently added to the CS VRC Registration Portal so that new users of the community may easily request for registration with a CS VO whilst more advanced users may request for specific roles and/or attributes within a specific CS VO. **AUTH** will develop and maintain this module (3 PM)

We will implement a User Interface web based front-end, which will be available through the CS VRC Registration Portal. This front-end will be implemented using the gsi-ssh mechanism alongside a proxy issuing mechanism. To be more descriptive, a registered VO user with stored credentials on the browser's cryptographic security device will be able to get mapped to a pool account onto a User Interface and have thus direct access to the Grid infrastructure through his or her browser window. New users of the CS community will benefit from this mechanism, as they will be able to submit and retrieve the output of their first jobs in only a few easy and understandable steps (a digital certificate and a valid registration with a CS VO will be sufficient to use the UI module).
Thus, once activated on top of the CS Registration Portal, the UI module will be an ideal starting point of interacting with the Grid for an inexperienced user, as the full list of production quality Grid resources will be available in the back-end.
AUTH and BIU will develop and maintain the UI module.

Within this activity we will also develop and deploy a module for interacting with the Scientific Database that will be developed in the context of JRA1 Work Package. Users will be able to query the database and access datasets based on their authorization level directly from their browser making data curation easy and robust.
BIU and UNIPA will develop the Database module

On top of the CS VRC Scientific Gateway we will implement a Resources monitoring mechanism so that users are notified at close to real time of CS VRC specific resources unavailability and downtimes. The Nagios monitoring mechanism will be implemented in the back-end of this service.
AUTH will lead this sub task

The CS VRC will design and implement a Scientific Database on top of the underlying Grid services. The database be available to users of the Complex Systems VRC and will provide a fined grained access control to the datasets. Specifically we plan to gather several collaborative datasets, starting with stock exchange, socio-economic and climate data sets. Due to the sensitivity and the possible commercial restriction of these datasets, we plan to use the Hydra encryption storage service for providing secure encrypted storage. In addition, we will use the AMGA tool to add metadata allowing for the end users to interface in a robust manner with the data on the Grid. Several levels of accessibility to these data will be put in place using VO roles and attributes so that access policies on proprietary data are properly configured. Once the Database is build on top of Grid services we will provide a module on top of the Scientific Gateway.
BIU will lead this activity and oversee the design, implementation and maintenance of the Scientific Database.
AUTH will contribute to the design of the Scientific Database.
JLUG, UNIPA and SU will contribute to the implementation phase and the maintenance phase.

Photon Science

User Interfaces
The PS communities have in contrast to HEP researchers a rather heterogeneous computing expertise. Some groups are well capable to perform complex data analysis in a distributed computing environment; some groups will fail entirely to explore the grid for their specific computing or data management tasks.
A high degree of interactivity and transparency for ongoing or pending transactions and self-

explanatory error or status reports are essential. Therefore, effort in the area of support for the integration of the Grid middleware with the user interface layer is required. This will largely consist of:

- Evaluation of existing Grid portal tools and lightweight clients (i.e. p-grade)
- Establish lightweight data export service from data archive to the user institute (for easy access to small data samples)
    - Connecting gateway to data archive
    - Connecting gateway to facility databases and metadata catalogues
    - Provide generic model and tools for definition of user specific metadata
    - Protect access to metadata using Grid authentication and authorization schemes
- Improve on existing user interfaces, trying to improve ease of use.
- Improve modularity of UIs. Since the PS provide services to a number of vastly different experiments, a unique interface might not be sufficient to satisfy all users' needs. A stronger modularization and plugability of the interface is therefore desired.
- Further develop, extend, customize, deploy and support the VCR (Virtual Control room) portal for the PS community.
- A number of standard software packages are capable to submit jobs to predefined remote compute hosts, clusters or MPI environments. Integration of seamless Grid job submission can greatly improve computing opportunities for a number of applications, and allows exploring local and distributed computing infrastructure likewise.

Cross facility annotation and exploration of data

On modern beam lines such as available at ESRF, DIAMOND, SLS, and PETRA III, individual crystallographic and Small-Angle scattering experiments take place on time-scale of few minutes. The data generated by these experiments need to be kept in an organized and easily analyzable form. At ESRF, iSpyB has been developed for this purpose over recent years. And is currently being used at world leading synchrotron in the field of macromolecular crystallography. iSpyB offers a complete meta-data recording mechanism, which can presumably be integrated into a grid-environment. It also has the potential to be adapted to all ranges of light source experiments.

- Further development of iSpyB to facilitate the handling of data at different synchrotrons in an integrated fashion, involving data base design, deployment of software and, harmonized credentials.
- Further development of iSpyB to record meta-data for a wide class of light source experiments
- Integration of iSpyB into a analysis framework.
- Integration of iSpyB into a grid environment.
- Validation through typical user group.

Humanities

Deliverables (brief description and month of delivery)

Table 16: Targeted Application Porting (SA3)

Work package number	SA3	**Start date or starting event:**			M03	
Work package title	Targeted Application Porting					
Activity Type[26]	SVC					
Participant number						
Participant short name						
Person-months per participant:						

Objectives

- Port example applications covering common use cases.
- Port strategic applications with high scientific, social, or economic impact.
- Interface common analysis frameworks or APIs with grid infrastructure.
- Optimize and maintain common scientific libraries for the grid infrastructure.

Description of work (possibly broken down into tasks) and role of participants

The number of different applications in a particular scientific domain is nearly as large as the number of participating researchers. Nonetheless there are often commonalities between those applications and how they interact with the grid services that can be packaged for reuse to avoid unnecessary duplication of effort and to speed the development of applications for the grid.

The scientific analyses in a particular domain usually can be grouped into a small number of distinct use cases. In this case, the application porting teams will select prototypical applications and help port them to the grid infrastructure. The principles and techniques used to port the application will be captured via "case studies" that will be made available to others within the discipline. In addition, an VRC may identify particular applications that have high scientific, social, or economic impact. The team will help port those strategic applications in order to motivate people within the community, to encourage more people to use the grid, and to publicize the utility of the grid infrastructure.

Many scientific domains maintain standard analysis software and APIs that encapsulate common analysis workflows or provide access to standard data repositories. These frameworks are often the foundation for many applications within the domain and hence interfacing them to the grid infrastructure can profoundly increase grid use within the domain with few inconveniences for users. Consequently, the porting teams will work to interface these frameworks to the infrastructure.

Similarly, there are many scientific libraries (BLAS, LAPACK, etc.) that are in common use but need to be adapted and maintained for the grid infrastructure to ensure that they function correctly and efficiently. Porting and maintaining those libraries lowers the entry barrier for scientists and will increase the number of grid users.

Mention that all ported applications must be registered in the application database.

The work on scientific gateways may also provide a common mechanism for integrating tools/libraries and thus may make it easier to share the effort and speed the availability of new packages.

[26] Please indicate one activity per work package:
RTD: Research and technological development; COORD: Co-ordination; MGT: Management of the consortium;
SVC: Service activities

Phases
Phases: setup, three phases of plans/evaluation/impact.

Common Tasks
The common tasks for this activity will be to create a RESPECT-like repository of tools and libraries that can be used easily in conjunction with the EGI UMD. There will be common meetings to evaluate suggested products and to coordinate porting of libraries between VRCs. Include here must also be discussions with TAPAS. In certain cases there may be combined effort to port/maintain a tool/library, e.g. MPI support.

Rewrite the each section keeping in mind the objectives of the work package and the general strategy. You should highlight concrete packages that you'll focus on porting. This should be written as examples rather than definitive engagements to port specific things. The length of each section should be 0.50-0.75 pages.

High Energy Physics
Application porting to the grid was successfully accomplished in several cases using Ganga, a user interface developed to provide a homogeneous environment for processing data on heterogeneous resources. Ganga is used by several projects ranging from image processing to medical physics in addition to its main use within the LHCb and ATLAS collaborations in HEP.
Tasks: (partners in parenthesis): Integration of the Ganga documentation and release process (CERN); Adaptation of Ganga to fully exploit multi-core processors, cloud computing, parallel systems and a unified middleware stack. Integration of data management aspects to interact with grid storage solutions. Inclusion of support for web based interfaces (BHAM, CERN, IC, UIO, DESY); Creation of sustainable Ganga user communities via support forums and tutorials. This also includes the implementation in Ganga of new functionality required for the evolution of their use cases. (BHAM, IC, UIO); Adoption of existing Grid components for user analysis (Ganga, AMGA) and integration of job submission and monitoring framework into the grid (DESY).
A continuous integration effort will also be devoted to maintain and adapt the experiment frameworks to new middleware components, to optimize resource usage by production and analysis workflows and to improve fault tolerance in data management.

Effort (Ganga support and evolution): CERN (1 FTE co-funded), IC (1 FTE co-funded), BHAM (1 FTE, co-funded), UIO (1 FTE, co-funded), DESY (1FTE co-funded)
Effort (Integration of LHC experiments frameworks): INFN (2 FTEs, co-funded), CERN (1 FTE, co-funded)

Life Sciences
The VRC cannot manage by itself the migration and porting of new applications in the Grid. The VRC focuses on coordinating help and providing first line user support to application porting in collaboration with application porting VRC. This first line support will be the catalyst to start-up collaborations and to undertake the application porting.

Actors in the LS VRC can be classified among different criteria. If we consider Grid-usage awareness, we can identify clearly from top-level users with challenging scientific problems and low capabilities on Grid usage and exploitation, to research groups with large expertise on the migration and exploitation of such infrastructures. There is a inherent need for collaboration, which should be the target of this subtask.

SA3 objective 1: "Port example applications covering common use cases."
SA3 objective 2: "Port strategic applications with high scientific, social, or economic im"pact.

SA3 objective 3: "Interface common analysis frameworks or APIs with grid infrastructure."
SA3 objective 4: "Optimize and maintain common scientific libraries for the grid infrastructure."

In order to make this collaboration happen, several issues must be faced:

- Awareness. An inventory of expertises and problems will be performed to identify the potential synergies. This inventory will be available and organised at different levels of detail. This will also include inventories of components, tools and even data.
- Communication. The LS VRC will foster the collaboration among groups creating thematic groups and communication tools.
- Confidence. Collaboration is based on mutual confidence. Mutual confidence cannot be imposed, but can be constructed more easily on top of signed agreements. The task will propose templates for IPR management, scientific cooperation agreements and other basic regulations. This could avoid medium-term misunderstandings and improve the quality of collaboration.

Support to the addition of plug-ins on the scientific gateways, targeted towards service providers in the life sciences user community
Support to the provision of grid-enabled bioinformatics and medical informatics web services, targeted towards service providers in the life sciences user community
Support to application porting through the scientific gateways
Support to application porting using grid-enabled bioinformatics and medical informatics web services
Support to application porting through the population genetics analysis scientific gateway

LS partners involved in this work package are CNRS (8PM), SIB (8PM), UPV (14PM), CSC (3PM), CNR (2PM), INFN (5PM) and AMC (5PM).

Computational Chemistry & Materials Science Technology
Tasks within this activity will focus on porting applications to the Grid, maintaining existing software ports and providing support for MPI environments on clusters and grid-enabled supercomputers.

Parallel version of software packages
Molecular and materials science applications are very well known of their high demands concerning hardware as well as computations time. This is why our main targets within this activity are parallel versions of VRC software packages crucial in achieving results within feasible time. Over the few years of development a predecessor of CCMST VRC – Computational Chemistry Specialised Discipline Cluster ported several, most frequently used (freely available and commercial) software codes to the grid. Majority of those packages were not enabled for parallel execution. Taking in to account parallel computing aspect, which has so far not been in focus of distributed computing, despite development of multi-core processors our goal is a natural extension of previous works required by CCMST community. The applications ported to the Grid will be selected in such a way that they are of particular use for community members and that they require large computing resources.

Maintenance of existing ports and grid version of new packages
Grid ports of CCMST applications require maintenance as versions with new functionality appear periodically. Such tasks are required to keep grid architecture attractive for users. In parallel grid ports of new software codes will be provided. Those mostly include specialized pre- and post-processing tools enabling extraction of important data from scientific output. Such tools are of particular importance from the point of view of application of the workflows in computational campaigns allowing execution of complicated batches of jobs from CCMST

portal.

Software ports library
CCMST software is not exclusive for members of our community only. Others VRC can benefit from using it as well. It is therefore to gather in one place all the techniques and tricks our specialists apply during porting procedure. Such a library including scripts will be particularly useful for users porting certain software packages for different middleware. For this case close cooperation link with Application porting VRC will be established to utilize their resources in order to jointly support CCMST codes on all middleware platforms supported by EGI.

Computer Science & Engineering
The specificities of the CSE community and of its interest in the grid preclude to couple gateway usage and computing on the grid infrastructure. Nonetheless, the VRC will devote effort to demonstrate the benefits of grid computing. In a complicated landscape of tools and software featured by the community, many of them being home-made or limited to a small sub-community, a significant commonality does exist, which is the numerical software suite provided by MathWorks (Matlab and toolboxes). We will thus exploit the EGEE/EGI implementation of the Matlab Distributed Computing Toolbox and Parallel Computing Toolbox [Ref. 24]. Two actions will be pursued.

- Develop codes associated to the specific needs of the CSE VRC. These will be showcases of the effectiveness of grid usage.
- Investigate with the wider community the grid usage model created by these new tools.

It would be very desirable to propose access to open-source equivalent software resources, e.g. Scilab/ProActive []. The CSE VRC will closely follow any opportunity to in direction that might appear in the timeframe of the project.
CNRS/LRI will participate in this subtask.

Complexity Science
We plan to design basic workflows specifically fitted to the needs of the Complexity Science community. Using these, users will be able to make robust and optimal usage of the underlying Grid resources. Understanding the needs of the Complexity Science community, not just with respect to the main processing parts but also with respect to the pre-processing and the post-processing parts, will allow us to design test cases of workflows making optimal use of the underlying middleware components and services.
A common example of a CS workflow would involve the creation or the retrieval of the complex system or complex network under study and the application of a successive series of stochastic or numerical algorithms on top of that dataset. Due to statistical deviations that arise in these sorts of studies the re-application of the algorithms on top of the same or similar initial datasets is required to access the dynamics of the systems so having the ability to design the workflow in advance would largely facilitate the whole procedure. The subsequent post processing of the results and visualization thereof could and should also be considered as the final part of such a workflow.
The collection of such usage scenarios in the field of Complexity Science and getting continuous feedback from CS research groups on common usage patterns of the underlying Grid resources is deemed crucial for the successful completion of this task.
UA and BIU will collect feedback and requirements on usage scenarios collection from CS research groups.
AUTH will design and produce use case examples of optimal workflows based on the feedback gathered.

The CS VRC will develop and produce optimized computational libraries of commonly used algorithms in the field of Complexity Science such as the Network Analysis, the Detrended

Fluctuations Analysis, the Wavelet and the multifractal DFA algorithms etc. Such algorithms that consume a large amount of computational time will be studied and parallel counterparts (using the MPI and/or the OpenMP libraries) will be produced and put in place for the researchers to use. These counterparts will be provided in the form of computational libraries so that Applications making use of these algorithms may benefit directly by performing the appropriate library calls.
AUTH will manage this sub task.

Photon Science

Operational Support
The PS VRC members are involved in several grid activities, e.g. serving as Tier 2/3 centres. However, integration into the PS experiments is still limited; recording of data and metadata for example is commonly not connected to existing grid infrastructure.

- Offer general Grid expertise for identification and solution of grid issues as well as site configuration and setup. This could include for example automatic cross-site network optimization to improve remote users' experience and cross-facility data exchange.
- Offer support on experiment specific integration.
- Adaptation and integration of HEP VRC developed operational tools, e.g. for workload and data management, to meet PS specific requirements.
- Interfacing site or experiment specific issue tracking systems with grid systems.

Data processing in the PS communities uses a good deal of closed source or proprietary software, various operating systems, MPI implementations and a variety of data formats. Data processing and analysis frameworks are hence complex and heterogeneous. Adaptation of these frameworks to grid infrastructure will require substantial support both from the user communities as well as the service provider. Fortunately, several projects, for example within the ESFRII EuroFEL project, are aiming to collect and define specific requirements in software repositories, or define standards for device definitions and exchange formats, upon which the PS VRC can base on. EMBL for example has already developed fast data evaluation frameworks for both SAXS and MX experiments.

Standard formats SE compliant
PS communities use a large number of different file formats. There is however a limited number of de facto standards defined, which are CIF or HDF5 based. HDF5 is on the verge becoming the standard format in photon sciences and is for example used by the LCLS free electron laser. The European laboratories are currently discussing to implement NeXus as a standard. NeXus is hitherto a HDF5 and XML based format and therefore fully compliant with HDF5. HDF5 has furthermore the intriguing advantage to be able acting as a mounted file system, which can greatly facilitate management and analysis of data collected at sources like XFEL. However, none of the standard format is capable to work directly on a dCache SE. We therefore aim to integration dcap/gsidcap IO into HDF5 and CIF.
In line with JRA1, we will also support incorporation of existing data and metadata into the data infrastructure, namely support beam lines and users to convert their old datasets to such standardized formats. To minimize the need for data conversion at a very stage of the data production pipelines, we will assist the beam lines enabling instruments to directly create data in standardized formats without the need of converters. Instrument Elements as successfully deployed within DORII is a promising path to achieve this goal. Implementations and developments will consequently aim to maximally profit from the ongoing DORII+ activities.

Analysis framework for SAXS
EMBL Hamburg has implemented a fast data evaluation pipelines for biological Small Angle X-ray Scattering (SAXS) based on 'ATSAS-Online'. It has similar scope and drawbacks like the MX framework. The SAXS analysis framework will be adjusted and ported for grid de-

ployment.

Deployment and Integration of SAXS application
Both analysis frameworks will be deployed. This will allow several thousand of users to use the frameworks for a wide range of structure determination experiments. It will serve as a showcase for other types of PS experiments. Essential component of this task is the documentation and dissemination of the frameworks in the grid context, to support additional user communities implementing analysis frameworks and deploying analysis software.

Crystallization as an integrated remote service
EMBL operates a crystallization facility as a service available to the European MX community. The facility generates millions of images per month, which have to be investigated and analyzed by the users. Currently all these operations are performed manually on a local computing infrastructure, which is inefficient both for the service provider as well as the user. A remote operation, automatic delivery of the images and distributed analysis could tremendously increase the usability and efficiency of the crystallization facility. In the long term it is envisaged to integrate the facility with upstream experiments, namely SAXS to analyze the crystallization trials, and MX to perform the experiment on the successful candidates, which is however beyond the scope of this proposal. This project serves also as a user showcase for a number of different aspects of the PS VRC.

Adaptation and maintenance
The PS VRC will support user communities beyond the SAXS case studies:

- Investigate possibilities to abstract from specific OS requirements for example through virtualization. Emerging new open source projects like RedHat's deltacloud might offer new opportunities and API's for multi-disciplinary computation in a heterogeneous environment.
- Adapt user interfaces and pluggable middleware components to meet the experiment specific requirements.
- Support maintenance of end-user distributed analysis tools/frameworks and their related VO-specific plug-ins.

Humanities
A 'watching brief' will be kept on application porting issues for Humanities grid users. This will be carried out in close alignment with SA2 on mapping user needs. Our policy will be to use this map to define possible areas for the applicability of application porting. We will work to understand the 'engaged' community's user needs for application porting services, both present and future and, if a suitable opportunity arises, develop a case study.

Deliverables (brief description and month of delivery)

Table 17: Summary of staff effort

Participant no./short name	**SA1**	**SA2**	**SA3**	**person months**
Part.1 short name				
…				
Total				

Table 18: List of milestones

Milestone number	**Milestone name**	**Work package(s) involved**	**Expected date**[27]	**Means of verification**[28]

1.8.4. Graphical presentation of component interdependencies

Provide a graphical presentation of the components showing their interdependencies with a Pert diagram or similar.

1.8.5. Significant risks and associated contingency plans

Table 19: Risks for User Support (SA1)

Risk SA1	**Impact**	**Occurrence Probability**	**Mitigation**
Low impact of novel documentation content to CS VRC	Only a few users benefiting from novel documentation	Low to medium	Disseminate upon the novel documentation material within the commu-

[27] Measured in months from the project start date (month 1).

[28] Show how you will confirm that the milestone has been attained. Refer to indicators if appropriate. For example: a laboratory prototype completed and running flawlessly; software released and validated by a user group; field survey complete and data quality validated.

community			nity

Table 20: Risks for Scientific Gateways (SA2)

Risk SA2	Impact	Occurrence Probability	Mitigation
Scientific Gateway provided by collaborating projects will not meet with the CS VRC technical needs	Work package progress will likely be slowed down	Medium	Invest more effort into interfacing with the developers of the generic portal so that the required specifications are met.
In the beginning of the Project it is expected that only a small number of people close to the Project will participate in setting up the Support Infrastructure. Thus there is a risk that only an even smaller number of people will be contributing their experiences in the Projects Knowledge Base.	Small impact of Knowledge Base on the CS community due to poor content	Medium	In the occurrence of such a risk we would firstly try to enrich the contents of the Knowledge Base with topics from the state of the art in Complex Systems research and secondly reach out to the community in order to get more people involved in the accumulation of experiences and use cases within the Knowledge Base.
EGI-SGI project is not funded.	Will have to rework plans around the fact that a common platform will not exist. Will increase long-term costs and complexity for the infrastructure. Risks to have different communities develop similar functionality.	Medium	The work plan includes activities to identify common requirements. This will hopefully lead to sharing of algorithms and techniques even if detailed implementations are different for different scientific gateways.

Table 21: Risks for Targeted Application Porting (SA3)

Risk SA3	Impact	Occurrence Probability	Mitigation
Low scaling or no benefit from parallel versions of CS VRC algorithms	Low or limited optimization of CS related applications with respect to the usage of resources	Medium to high	Investigate other options of accelerating algorithm execution such as CUDA, OpenCL, RapidMind as well as mixed versions of MPI with the above if applicable.

Shortage of interest (LS)	No work	Low	Dissemination
Too much interest (LS)	Cannot fulfil users expectations	High	Coordinate with external parties (regional programmes) for training of additional manpower
TAPAS project is not funded.	This project may have to concentrate porting effort on common application libraries that would be ported and supported by TAPAS. This reduces effort for flagship applications and requires additional effort for maintenance. This also risks to have different port of the same library in different VRCs.	Medium	Cooperation between VRCs within the project will ensure that common libraries are shared between VRCs. Communication with other VRCs outside the project will also help to avoid duplication.

1.9. Joint Research Activities and associated work plan

A detailed work plan should be presented, broken down into work packages (WPs) which should follow the logical phases of the implementation of the project's Joint Research Activities, and include assessment of progress and results.

1.9.1. Overall strategy of the work plan

(Maximum length — one page)

The "golden rule" for these activities is to ensure an evolution of the grid in a direction necessary to serve the user community. A couple areas have been identified as critical issues for the communities and ROSCOE will do core requirements gathering and testing of services for leading this evolution.

VRCs most often make use of and create large datasets that are on one side indispensable for their activities and on the other side represent a valuable outcome of their knowledge production. An important component of the VRCs grid activities is in this respect the handling of the outcomes of the Europe's recent mass digitization and metadata creation initiatives aimed at preserving heritage, fostering related research and consolidating the ground on which future advances are saved and made available for shared usage. For this reason ROSCOE will address significant efforts to the enhancing of collaboration and mediation inside individual domains and among the various domains. In particular the work will focus on articulating areas of common interest and identifying common approaches to the handling of complex data-driven applications as a key to the design of the activities of VRCs. This will be carried out in particular for those domains in which both experimental acquisitions and theoretical simulations produce a large and continuous stream of information to make the appropriate representation, storing, accessing, retrieving and handling of data a challenging problem and a key step to sustainability ………………………

Equally important for sustainability will be the acquisition and handling of data related to traces of the grid activities and to the characteristics of the various domains to model the dynamics of the grid using techniques of Complex Network Analysis, Machine Learning and Data Mining. The objective of the associated work will be the offering of a better global picture of the system and its behaviour as well as the stimulation of the development of models for its sustainable management. Two aspects of sustainability become pertinent in this context: (i) how can existing grid systems be made more reliable and self-adapting (based on automated adaptation to known faults); (ii) how can grid systems be made accessible to a wider user community, thereby encouraging greater usage of these existing systems. Both of these aspects rely on improving usability of existing grid systems, and making them reliable, so that a wider community can consider use of such infrastructures. Sustainable management will pursued in ROSCOE by using the information produced by the sensors of the monitoring information and by handling related data by means of scalable algorithms to the end of building suitable representations and models of the grid and its VRCs. This activity will find different specific implementations in the various domains that will be used as a test case for developing realistic simulations of the grid behaviour, for predicting oncoming events, and for optimizing the scheduling and workload distributions. Some experiments will be conducted to foster cooperation among grid users to better deal with the rapidly growing complexity of computing systems management, and to reduce the barrier that complexity poses to its further growth. To this end a more intelligent use will be made of present sensors through the support of learning agents and the formulation of new sensors able to provide information on the quality of the service through objective and subjective evaluations will be stimulated. The aim will be not only the enhancing of efficient competitive cooperation inside a single VRC but also the fostering of a similar attitude among different domains targeting so far one of the key objectives of EGI and contributing to the development of a grid economy.

1.9.2. Timing of the different WPs and their components

(Gantt chart or similar)

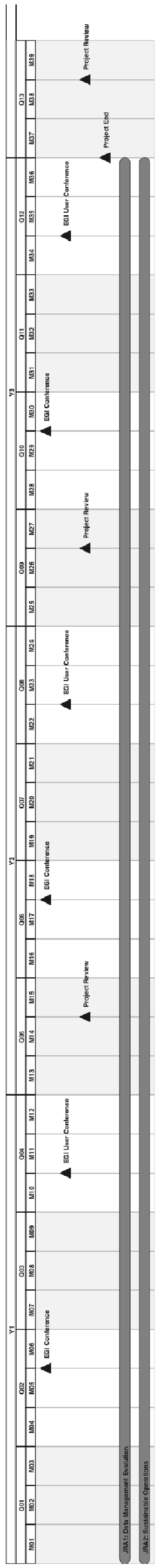


Figure 7: Gantt Diagram for Joint Research Activities

1.9.3. Detailed work description broken down into work packages

Table 22: Work package list

Work package No[29]	Work package title	Type of activity[30]	Lead Part. No[31]	Lead Part. short name	Person / months[32]	Start month[33]	End month
JRA1	Data Management Evolution	RTD				M01	M36
JRA2	A Sustainable Operations Model	RTD				M01	M36
	TOTAL						

Table 23: Deliverables list

Del. no.34	Deliverable name	WP no.	Nature35	Dissemination level36	Delivery date37 (proj. month)

[29] Work package number: WP 1 – WP n.
[30] Please indicate one activity per work package:
RTD = Research and technological development; COORD = Co-ordination; MGT = Management of the consortium; SVC = Service activities
[31] Number of the participant leading the work in this work package.
[32] The total number of person-months allocated to each work package.
[33] Measured in months from the project start date (month 1).
[34] Deliverable numbers in order of delivery dates. Please use the numbering convention <WP number>.<number of deliverable within that WP>. For example, deliverable 4.2 would be the second deliverable from work package 4.
[35] Please indicate the nature of the deliverable using one of the following codes:
R = Report, **P** = Prototype, **D** = Demonstrator, **O** = Other
[36] Please indicate the dissemination level using one of the following codes:
PU = Public
PP = Restricted to other programme participants (including the Commission Services).
RE = Restricted to a group specified by the consortium (including the Commission Services).
CO = Confidential, only for members of the consortium (including the Commission Services).
[37] Measured in months from the project start date (month 1).

ROSCOE

Table 24: Data Management Evolution (JRA1)

Work package number	JRA1	**Start date or starting event:**		M01	
Work package title	Data Management Evolution				
Activity Type[38]	RTD				
Participant number					
Participant short name					
Person-months per participant:					

Objectives

- Development of distributed storage strategies and solutions for data intensive (ES-FRI-) projects
- Development of enhanced access to distributed data and metadata
- Development of security fine grain access control schemes
- Development of grid based long term data storage concepts and solutions
- Cooperation with industry

Description of work (possibly broken down into tasks) and role of participants

Need to provide the common description of this activity. Break down into tasks that are linked to specific groups but also common tasks.

The core storage management solutions that are in use today have their roots in a different era – some 15 to 20 years ago. Since that time not only have relative costs and capacities (such as storage and network throughput) changed enormously but also the entire IT landscape. Attempts to rationalize the inevitable diversity via standards such as the Storage Resource Manager (SRM) have had debated success: if a concept does not exist in the backend it is hard to make it 'appear' via the front-end interface. Furthermore, the available implementations vary widely in their interpretation of the agreed standard, leading to additional confusion. Finally, as the individual components have been designed and implemented almost entirely independently, large opportunities for optimization and rationalization have been lost. For example in HEP, the LHC VOs deal with sets of files (depending on their computing models) which have some strong logical connection: typically the full set is treated together in various operations ranging from transfer through to data processing. However, such concepts are not implemented in the component data management solutions – even though they would allow many operations, such as bulk network transfer or retrieval from tape, to be greatly optimized. They are typically 'unpacked' – possibly by catalogue lookups – handed to the subsystems one by one and then reassembled at the target system. Such operations may occur multiple times: at the source storage system, at the file transfer stage and again at the target system. Thus an investigation of the end-to-end data management problem is long overdue. This would take into account not only the advances of recent years but also take a higher level view, covering at least catalogues, data transfer and storage / access issues. Again, although of particularly pressing concern for the supported communities, the requirement is highly generic meaning that advances in this field would benefit a range of other disciplines – as has been demonstrated on numerous occasions in the past.
Much of the successful data management on today's grid infrastructures is based around crea-

[38] Please indicate one activity per work package:
RTD: Research and technological development; COORD: Co-ordination; MGT: Management of the consortium;
SVC: Service activities

tion and access to new datasets. There has to date been little direct integration with existing data sources except where files can be extracted from these and used independently in the grid environment. This approach will not scale to support full and effective use of the key communities represented in this project. Therefore, it is essential to develop more effective integration into existing and possible future data sources and data services. This issue is particularly relevant to the Photon Science, Complexity Science and the Humanities [and Biomedical Science] communities.

Common Tasks
The management of large amounts of data or moreover information, which today is distributed not only between sites but also between communities is a important issue for many communities. Therefore all of the proposed tasks do have a "generic" relevance for many communities, the proposed implementations do have a "show case" character. The same holds for "long term data storage strategies", which do ranges from developing concepts for long-term experimental data storage together with the analysis software to "information storage" in Humanities.
The HEP community itself is very much interested in these issues and therefore will participate with unfunded resources.
In all cases it is planned to strongly cooperate with other VRC's like Astro-Physics and Earth Science, which are not part of the ROSCOE proposal.

Photon Science
In Photon Science data and storage management solutions that are in use today are based upon more or less standard industrial solutions. The reason for this is, that past and today's experiments are carried out by small groups within a couple of day's at the facilities. Data, after a short period are transferred to the home institutes either by physically sending disks or using the European network infrastructure.
With the new light sources this behaviour will not be working at all any longer. The mere data volumes and burst rates in combination with the privacy of raw experimental data as well as metadata require careful selection of appropriate, scalable data management frameworks. Existing solutions like e.g. the HealthGrid model might prove to fulfil the basic requirements for some PS specific applications, but might also create significant bottlenecks in the analysis framework for envisaged burst data rates of 1PB/week.
Due to the development of the detectors and the brilliance of the light sources the amount of data is increasing drastically. For the ESFRI-project XFel and based on today's megapixel detector technology the amount of data per year exceeds those, expected for the LHC. Going to higher resolution of the pixel detectors makes the situation even worse. One of the key problems herein clearly is the data suppression.
From the technical point of view, one has to face a couple of problems.
First there are the enormous data rates on site, these facilities are running in a burst mode fashion. The other major issue to be tackled is related to the special data access and transfer pattern create by the PS user communities. Experimental data are often created and exclusively analyzed by small but distributed research groups requiring access to data across facilities and disciplines. Replication of massive data volumes to individual users needs to be combined with replication to cooperating facilities and/or national data centers. Particularly the European XFEL has to cope with a wide variety of protocols and replication schemes, so that it might even become necessary to optimize routing of the data through the Geant networking infrastructure. So, the whole end-to-end management of the data in terms of catalogues, security, etc. have carefully to be investigated.

Security and fine grained access control
Experiments at light sources are often highly competitive, data are exclusively owned by the individual research collaboration and data as well as metadata need to remain fully protected

for a time that frequently exceeds the duration of the data in an archive. Consequently, fine-grained authorization schemes and ACL's are indispensable.
As long as data management and analysis is performed locally, data protection can easily be achieved through authentication/authorization schemes already implemented at most light sources. However, secure data analysis in a grid environment is still a non-trivial task. A number of solutions have already been developed and interfaced for example to gLite middleware, like in grid projects focusing on medicinal data.
Within the EuroFEL ESFRI project a number of central authentication and authorization schemes are currently being discussed, particularly in view of cross facility and cross-national authentication and authorization. Systems currently favoured are Shibboleth or OpenID based. Although it appears trivial to map a federated ID to a short-lived grid-certificate, or to use a personal grid-certificate to authenticate against a Shibboleth-System, a number of issues seem still open.
The diversity of the Photon Science Community, the security requirements and replication schemes requires to primarily focus on industry standards and solutions.
Due to the diversity of the Photon Science Community, on one hand one has to learn from other communities facing similar problem like HEP and on the other hand one primarily has to look for industrial standards and solutions, since a model is needed which is suitable for small and large sites at the user end.
The following tasks are proposed:

- Task 1: Investigation of the readiness of the industrial and open source technology development in storage standards and solutions (M01-M06)
- Task 2:Cooperation with industry and contribution to open source projects in developing standards and solutions like NFS, AFS, Lustre etc for large data volumes and high throughput access. (M01-M36)
- Task 3:Cooperation with other communities, facing similar problems (M01-M36)
- Task 4: Development of common end-to-end data models for photon science anticipating the highly distributed structure of the grid and the special security requirements imposed by the photon science community. (M01-M12)
- Task 5: Implementation of data models developed in Task 4. (M06-M36)

The whole work program for the Photon Science proposal will be carried out in strong collaboration with the HEP VRC, in particular LHC, because of the similarities in technical problems.

Complexity Science
The *Complexity Science* VRC will implement and deploy a Hydra/AMGA based database model as outlined in SA2. Hydra will serve as the data protection engine, whereas AMGA provides secure access to metadata and acts as a key ring of encryption keys. Such a model is expected to be appropriate for other VRC's and disciplines as well. However, the scalability of such a model needs to be investigated to clearly define the inherent limitations and identify the appropriate use cases.

- Task 6: Development of a test suite for a generic Hydra/AMG based database model

Humanities
The effort in this JRA will be focused on existing humanities data collections. The 'complexity deluge' imposes multifaceted and highly distinctive demands on e-infrastructures. The H-VRC will develop a programme of activities to address the core concerns of the humanities grid early adopter community which are affected by this deluge; which we identify as being concerned primarily with massive increases in the volume and complexity of humanities data available. The programme will include:

- Gathering knowledge on Europe's recent mass digitization activities and metadata creation initiatives, and developing appropriate responses for curating and creating conditions for sustaining very large digital collections and libraries. The EU's digital libraries initiative Europeana claims nothing less but total and sustainable access to heritage research records (http://www.europeana.eu/).

- Documenting and addressing the new challenges this presents. For example, there is now a need to consider federated digitization workflows, e.g. as supported by Goobi http://gdz.sub.uni-goettingen.de/index.php?id=20&L=1; and grid-based workflows can be used to support distributed workflows for OCR and other forms of multimedia data capture.
- Where possible and appropriate, H-VRC will seek to collaborate with the other VRCs to articulate areas of common interest and, identify common approaches to complex data-driven applications; although its focus will be on humanities datasets. Where projects/issue etc arise which are of potential interest to other domains, it will act proactively to be the primary point of contact for humanities researchers.
- Scoping the use of Grid services which cooperate with existing repository systems, focusing in particular on humanities applications pertaining to distributed datasets and virtual collections; and the metadata catalogues and interoperability schemas needed to support those applications. In doing so, it will seek to define the role of the Grid in data curation, which, in this case, constitutes managing and analyzing data, rather than 'just' moving or storing it.

The nature of humanities data is such that both particular and general challenges will be encountered: as such, the JRA activity will feed logically back into the Service activities, especially the formulation of the user support knowledge base,
The H-VRC will produce a scientific report, identifying which existing digital humanities resources and/or European digitization programmes have most value as exemplar material for grid-based research, and what form that research takes. It will specify what kind of services are needed to interoperate to best serve the humanities community's needs (e.g. digitization, image processing and interpretation support services), and how this can be achieved.

- Task 7: comprehensive analysis of current digitization activities
- Task 8: Development of strategies and policies for Grid usage in humanities

Computational Chemistry & Materials Science Technology
The CCMST VRC will work on the development of shared data models and its implementation (to make them almost de facto standards for the community) for Quantum Chemistry (QC) data to support collaborative work and interoperability. Typically this is the case of ab initio electronic structure outputs generated by different Quantum Chemistry packages (each of which has its own format) for different (set of) arrangements of the nuclei when they need to be gathered together or to be provided as an input for other programs in a row or in parallel in complex CCMST simulations. The work will start from the proposals quoted in the literature for developing such standards. A model format for ab initio data tailored on the DALTON package and meant to be ported on the grid has been already suggested, for example, by the EU collaborative initiative COST. Such product, named QC5, and related procedures were specifically designed having in mind their 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. Experimenting these assumptions and make the model work for an operative grid will be the challenge of CCMST. Then the goal will be to use the model to cope with the increasing availability of QC digital data (interfaced with users and codes) in order to produce a profound impact in both the quality and rate of discovery of the CCMST VRC.
The contribution of CCMST to JRA1 will deal with the evolution of chemical data management on the grid. The chemistry and materials science community will deal in general with the evolution of molecular information storage, retrieval and manipulation for various practical purposes. In particular, however, it will concentrate on the storing and handling quantum chemistry data focusing on its model representation (including efforts to develop common

standards for it), semantic applications (including the intensive use of different field-specific symbolic representations) and privacy protecting data base queries (including those related to structure-properties relationships). To this end links to data production from quantum chemistry tools of the Configuration
Interaction (CI) formalism level (with particular emphasis to Multi-reference approaches based on the use of local orbitals and to size consistency problems) as well as from Quantum Monte-Carlo, mixed CI and DFT techniques, relativity and pseudopotentials packages will be activated.
Design of models for data representation will be based on the QC5 approach already developed for the output of electronic structure information of some ab initio programs and possible extensions to molecular dynamics will be investigated. In practice, the main effort will be devoted to integrate QC5 into the more general evolution line developed by the WP for data management by taking into account the specific semantic implications of molecular data representations and the importance of security aspects intrinsic in the handling of "sensitive" information.

- Task 9: development of data management strategies on the grid

The CCMST partners will contribute to the total JRA1 effort with
UNIPG 10 PMs
CMRS IRSAMC 5 PMs

Deliverables (brief description and month of delivery)

Table 25: A Sustainable Operations Model (JRA2)

Work package number	JRA2	**Start date or starting event:**		M01	
Work package title	A Sustainable Operations Model				
Activity Type[39]	RTD				
Participant number					
Participant short name					
Person-months per participant:					

Objectives

This task contributes to the overall grid management and governance goals towards sustainable, reliable and secure grid platforms, through the following actions.

- Define advanced representations and evaluations of grid and VRC behavior
- Propose new sensors and application indicators, in order to evolve towards an information provisioning service organization.
- Develop tools for intelligent monitoring via learning agents
- Provide Quality of service and Quality of user parameters
- Develop a credit system to reward Grid work

Description of work (possibly broken down into tasks) and role of participants

Provide the common description of this task. Determine if there is a common thread that can be used to tie the two activities together.

As pointed out in the call, the achievement of sustainability by the VRCs is a key objective. This prompts VRCs to design algorithms and tools suited to evaluate the relevance and the importance of the contributions of the various members of the community to the achievement of their goals and objectives. The basis on which such an evaluation can be built is the traces of the grid activities. At the same time, at a more general level, the mission of the Computer Science and engineering community in ROSCOE lies in designing grid models and validating them on challenging operational issues. Cross-fertilization of these two activities will contribute to transfer expertise in fostering cross-community collaboration from current projects to EGI.eu (general goal 4), by advancing the realization of a more sustainable operation model. For this purpose, use will made of advanced methods based on Machine Learning, Complex Systems, and Data Mining to analyze, represent, and correlate the quantities monitored by the grid sensors to the end of modelling the dynamics of the various domains. The work will therefore require the handling of a massive amount of sensors' data, the development of suitable scalable and/or grid-enabled algorithms, the evaluation of the representativity of the data, and the assessment of the quality of the representation as well as the formulation of a model grid VRC.

Besides the general-purpose monitoring data collected by the Grid Observatory (SA2), specific information is or is intended to be collected by many VRCs along the needs of its particular domain. This information is precious, because it captures the expertise and evaluation of the VRCs.

In the JRA, the main aim will be the fostering of cooperation between grid users to reduce the

[39] Please indicate one activity per work package:
RTD: Research and technological development; COORD: Co-ordination; MGT: Management of the consortium;
SVC: Service activities

barrier that complexity poses to its further growth. To this end presently available sensors will be used through learning agents and new sensors will be formulated in order to provide information on the quality of the service considered using both objective and subjective parameters. Behavioural models of the grids will be created, thus enabling to develop principled solutions based on the valuable achievements of Autonomic Computing, whose aim is to develop computer systems capable of self-management, to overcome the rapidly growing complexity of computing systems management, and to reduce the barrier that complexity poses to its further growth.
The aim will be that of reaching sustainability through both an enhancement of efficiency within the single VRC and the strengthening of collaborative efforts between different domains.

Phases
Provide the phases of this activity for the Gantt chart. This may require providing separate phases for the two subtasks.

High Energy Physics

Life Sciences

Computational Chemistry & Materials Science Technology
One of the main commitments of the CCMST VRC is the structuring of domain activities as true cooperative endeavours (especially when tackling the computationally demanding molecular simulations using quantum means) through the creation of instances and mechanisms inducing effective collaboration. This will be achieved by designing and implementing some procedures as SOA applications structured in terms of collaborative services offered on the grid. This choice implies that each application is usually made of a set of services (which may be themselves made in turn of other (simpler) services and that the contribution of each component of the service needs to be evaluated not only in terms of the resources engagement but also in terms of its quality (QoS). The evaluation of the quality of a service (as well as of all its components) will be carried out using the outcome of the already implemented sensors of the grid (like those provided by DGAS) based on objective parameters whenever this is possible. The analysis of the related set of data will be carried out jointly with other laboratories of the work package and will adopt the strategies developed in it. In addition CCMST will develop new parameters (for which a proposal of new sensors to be inserted in the monitoring software will be made) of the subjective type. Subjective evaluation parameters will be based also on procedures quantifying the quality of user (QoU). In this case it is likely that new parameters and sensors will have to be designed and developed. Both QoS and QoU quality indices will be employed to drive the activities of the CCMST VRC towards its objectives and in particular to enhance the collaborative efforts mentioned above. To this end the determination of the quality indices will be linked to a system of credits awarded depending on the contribution to the VRC activities and redeemed in exchange for the resources available to the VRC.

Computer Science & Engineering
The CSE VRC contributes to the work package through the following actions:

- Models of the grid dynamics: define advanced methods for analyzing, representing, and correlating the data collected by the Grid Observatory Information provisioning services: propose operational on-line usage of these methods, in order to evolve towards an organization that provides information-based services.

Models of the grid dynamics (LRI, UNIPMN, Cardiff)
Based on the acquisition of grid traces and the representation of the grid domain (SA2), the

objective of this task is to model the dynamics of the grid. Grid, seen as a complex structure, has its own emergent behaviour that we examine using techniques from the areas of Complex Network Analysis, Machine Learning and Data Mining. The final goal is to provide an understandable model of the grid, or more realistically, of relevant subsystems, allowing the system administrators and end users to make use of it; therefore the model should be able to "explain" its output, or provide some insights into the typical uses of the system.
In this perspective the activities will focused on

a) Complex networks as a general model,
b) Application of statistical inference to fault detection, diagnosis and explanation.

The input from the user VRC will contribute to provide informative labels capturing the user interpretation ("expert" labels in Machine Learning), which are generally not available in general purpose trace. When such labels are available, more powerful methods (supervised learning) can be exploited. When they are not available, the unsupervised learning, or data mining methods have proved their efficiency. Moreover, user VRC input will streamline this process by providing expert advice and validation
The Computer Science and Engineering community will exploit the models created in this activity to generate test data, for prediction of oncoming events in order to optimize the scheduling and workload distribution, as well as for detection of outliers, intrusion or other anomalous behaviours in the system. These models will also contribute to create autonomic use cases (NA2), where the target self-management is based on feedback loops that instantiate the model.
As the modelling task will progress, the implicit topology structure defined on the space of grid events through grid models will be embodied in the ontology developed in SA2. Conversely, the navigation tools constructed will be tested against the following criterion: acquire information that are at the same time sufficiently precise to ground the model and parsimonious to allow for a wide range of experimentations, both in the Computer Science and Engineering community and in the chosen end-user community.

Information provisioning services (LOGICA)
Information Technology (IT) Governance focuses on performance of IT systems and risk management. Industrial governance standards are captured in ISO 38500 [Ref. 17], which strongly focuses on managing the IT resources on behalf of stakeholders who expect a return from their investment.
For the complex grid infrastructures, automated support to processes of information retrieval and analysis is become necessary. The specific focus of this sub-task is *governance support by intelligent monitoring and learning agents*.
Fortunately in grid an extensive monitoring infrastructure is already operative (see for example gLite logs, user-level software) including the generic monitoring environment Nagios. Therefore the task confines itself to the exploitation of the output of these monitoring tools, in an operation-oriented perspective. To correlate information between domains we shall develop methods to correlate event-information between sites and try to automate the retrieval of application-level metrics.
We will demonstrate tools that allow feeding back the results of these metrics through both automatic and administrative means to the site operations. Of primary interest are automatic feedback loops that enable near-real time failure identification and remediation. Our commitment is to develop Nagios plug-ins. A technical challenge thereby is that Nagios plug-ins are not statefull, while state-of-the-art methods exploit historical data.
In the perspective of real-time real time pro-active operational management, the activities will focused on:

a) Modelling the grammar of grid behaviour, and use these models to predict the behaviour;
b) Building a demonstrator.

Pro-active grid management also implies a sound communication layer between the human administrators and software agents. The agents should uses the same ontology as the human administrators and be able to learn from the strategy that human administrators follow in their analysis tasks. Collaboration and interfacing are therefore a key element of this contribution.

Complexity Science

Photon Science

Humanities

Deliverables (brief description and month of delivery)

Table 26: Summary of staff effort

Participant no./short name	JRA1	JRA2	person months
Part.1 short name			
…			
Total			

Table 27: List of milestones

Milestone number	Milestone name	Work package(s) involved	Expected date[40]	Means of verification[41]

1.9.4. Graphical presentation of component interdependencies

Provide a graphical presentation of the components showing their interdependencies with a Pert diagram or similar.

1.9.5. Significant risks and associated contingency plans

Table 28: Risks for JRA1

Risk JRA1	Impact	Occurrence Probability	Mitigation

Table 29: Risks for JRA2

Risk JRA2	Impact	Occurrence Probability	Mitigation
Difficulty in acquir-	Models will lack	Low	We can apply data trans-

[40] Measured in months from the project start date (month 1).
[41] Show how you will confirm that the milestone has been attained. Refer to indicators if appropriate. For example: a laboratory prototype completed and running flawlessly; software released and validated by a user group; field survey complete and data quality validated.

ing the necessary data for learning, at the right level of granularity (too detailed data)	generalization power		formation and abstraction techniques to change the level of granularity in the data. In this way at least approximate models can be built.
Lack of expert interpretation of the data.	The practical impacts of the results will not be demonstrable	Low	Initiate a virtuous circle with experts by addressing real problems of progressive difficulty with respect to the interpretation issues

2. Implementation

2.1. Management structure and procedures

Describe the organisational structure and decision-making mechanisms of the project. Show how they are matched to the complexity and scale of the project.
(Maximum length for Section 2.1 – five pages)

The ROSCOE consortium is composed of 48 partners, from 14 European countries consisting of 45 academic institutes, 1 SME, and 2 large enterprises. Two American partners also contribute to the project on an unfunded basis.
For this large project, a clear management structure has been devised that contains representatives for all of the project's stakeholders and that ensures the technical, financial, and administrative challenges are met. The diagram in Figure 8 provides an overview of the ROSCOE management structure.

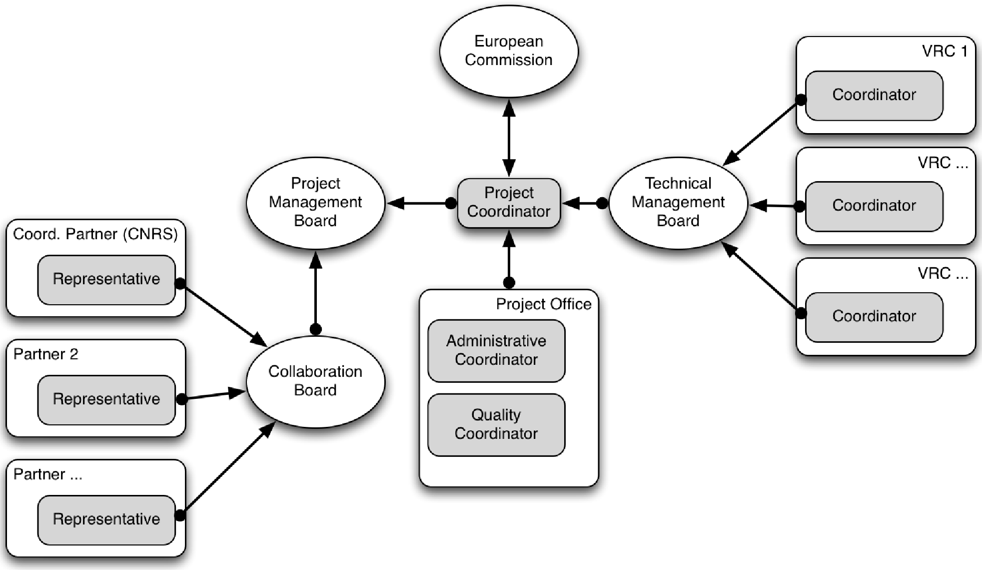


Figure 8: Management Overview Diagram

The largest body in the project is the Collaboration Board (CB), which consists of one representative from each signatory. This board selects the members of the Project Management Board (PMB) subject to the policies and procedures described in the Consortium Agreement. The PMB oversees the project. The PMB is responsible for major decisions concerning the project's work plan, allocation of project resources, and resolution of conflicts.
The Project Coordinator is the sole intermediary between the parties and the European Commission. He regularly reports to the PMB and is responsible for the daily operation of the project. He ensures that the project follows the defined work plan to achieve its goals. He chairs the Technical Management Board.
The Technical Management Board (TMB) consists of the coordinators of each VRC. Members of the Project Office may attend to facilitate the meetings and to ensure administrative information is passed to the VRC coordinators. The TMB ensures effective cross-activity communication, follows the progress of project with respect to the defined work plan, and resolves any issues encountered.
The Project Office with the Project Coordinator is responsible for financial reporting, general project administration, quality control, coordinating common services, and furnishing project collaborative tools such as mailing list servers, meeting management software, chat servers, and the like. The Administrative Coordinator heads the Project Office.
The Quality Coordinator will oversee the quality control aspects of the project, ensuring that deliverables and milestones are of high quality and that they are achieve in a timely fashion.

The following section provides detailed information about each identified management role or body.

2.1.1. Detailed Description of Management Roles and Bodies

2.1.1.1. Project Coordinator

The Project Coordinator will ensure that the project meets all its contractual obligations (including prompt delivery of all reports and deliverables), that the participants execute the defined work plans, and that the project ultimately achieves its goals. The Project Coordinator interacts with the following bodies:

- European Commission: The Project Coordinator will be the sole liaison with the European Commission for the project.
- Technical Management Board: The Project Coordinator will chair the Technical Management Board (TMB) and define the agendas for each meeting.
- Project Management Board: At each Project Management Board (PMB) meeting, he will present the achievements of the project to date, any issues that have been encountered, and any recommended actions that need to be done by the PMB.

Dr. Charles Loomis is proposed to be the ROSCOE Project Coordinator. He has participated in the flagship European grid projects—European DataGrid, EGEE, EGEE-II, and EGEE-III—effectively leading the activity related to user community expansion and support in the EGEE-II and EGEE-III projects. Because of this experience his has close ties with the scientific user communities using the European Grid Infrastructure and also with those operating it.

2.1.1.2. Coordinating Partner

The Coordinating Partner is responsible for the scientific coordination, administration, and financial management of the project through personnel provided to the project. The Coordinating Partner will be responsible for the distribution of the EC financial contribution to the project's partners.
CNRS is the Coordinating Partner for ROSCOE. As the largest fundamental research organisation in Europe, it has extensive experience in European Framework Programme projects and in performing the administrative, legal and financial services necessary to ensure the effective management of the project and the coordination of the Consortium.

2.1.1.3. Administrative Coordinator

The Administrative Coordinator will manage the Project Office and will help the Coordinator, either directly or with others in the Project Office, with everyday tasks related to all the management, administrative, and financial reporting aspects with respect to the coordination of the project. Specifically, these tasks include arranging meetings, taking minutes, and disseminating information to the project participants and partners. The administrative coordinator interacts with the following bodies:

- Project Coordinator: The Administrative Coordinator will reports directly to the Project Coordinator.
- Project Office Personnel: The Administrative Coordinator will manage the Project Office and interact with all its personnel.

2.1.1.4. Project Office

The Project Office is collectively responsible for the management, administrative, and financial reporting aspects of the project.
The tasks will be:

- Meeting Organization: Organising and animating along with the Coordinator all the meetings scheduled for the project.
- Maintaining Contacts: Keeping the contact information of members and other useful people updated and available. This includes the maintenance of defined mailing lists.
- Internal Communication: Disseminating the minutes and decisions of the meetings. Managing the communication with all the Partners, providing the Consortium with all the necessary information concerning the management of the project.
- External Communication: Helping communicate with the European Commission and with other EGI-related projects.
- Monitoring Progress: The Project Office will help monitor the project's progress with respect to the work plan and with respect to any feedback received from external reviewers or from the EC.
- Project reporting: People in the Project Office will assist the Coordinator in preparing the project's periodic and final reports to the EC. It will also ensure that the deliverables are sent to the EC on time and are of high quality.

Financial issues: Collecting cost statements and certificate on financial statements from all partners, completing and monitoring the Coordinator's budget, checking the eligibility of all the expenses.

2.1.1.5. Collaboration Board

The Collaboration Board (CB) is the largest body in the ROSCOE project, consisting of one representative from each Signatory to the Grant Agreement. Each representative will have a vote weighted by the number of participating institutes he represents. In accordance with the policies and procedures that will be detailed in the Consortium Agreement, the CB selects the members of the Project Management Board. The CB will meet at least once every year.
Should a limit on the number of members in the PMB be included here? Should the coordinating partner be guaranteed at least once place on the PMB?

2.1.1.6. Project Management Board

The members of the Project Management Board are selected by the Collaboration Board. The Chair of the PMB will be chosen from among the members of the PMB. The Consortium Agreement will specify how PMB representatives are appointed, how they are replaced, and how the chair is selected.
Meetings will take place at least once every three months and may take place physically or via telephone or video conferencing. Additional meetings can be called by the Chair of the PMB as necessary, in consultation with the Project Coordinator. The time, location, and agenda of the meeting will be determined by the Chair of the PMB, in consultation with the Project Coordinator and members of the PMB. The time and location must be announced at least one month in advance. The agenda must be provided at least one week prior to the meeting and must include a project status report from the Project Coordinator.
All Parties shall agree to abide by all decisions of the PMB. All disputes shall be submitted in accordance with the provisions of the Grant Agreement and Consortium Agreement.
The PMB must oversee and make decisions related to issues such as:

- Scientific guidance of the project
- Significant modifications of the work plan
- Allocation of resources
- Resolution of conflicts
- Intellectual Property issues

- Gender equality
- Implementation or modification of the Grant Agreement or Consortium Agreement

The PMB may make advisory statements about any aspect of the project.
The Chair will act as a liaison between the PMB and the Project Coordinator, working to achieve consensus among the PMB representatives on outstanding issues.
The Chair of the PMB will prepare minutes of the meeting with the help of the Project Office. These must be made available to the PMB within one month of the meeting.
The PMB interacts with the following bodies:

- Project Coordinator: The Project Coordinator will present a report to the PMB at each meeting. The Chair will act as the liaison between the Project Coordinator and the PMB.
- Partner Constituencies: The representatives will make their constituencies aware of the issues presented to the PMB and the outcome of the meetings.
- Project Office: The Chair of the PMB will interact with the Project Office in the preparation of meeting minutes.

2.1.1.7. Technical Management Board

The Technical Management Board (TMB) will consist of the Project Coordinator and representatives from the VRCs. Members of the Project Office may participate to facilitate the meetings or to communicate administrative information. The Project Coordinator will be the Chair of the TMB, although a replacement from the TMB members can be appointed if necessary.
Meetings will be held regularly and may either take place physically or via telephone or video conferencing. The Chair of the TMB will prepare the agenda of the meeting in consultation with the members of the TMB. Minutes and actions from the meeting will be taken by the Administrative Coordinator and made available to participants before the next meeting.
The TMB is responsible for following the progress of the project with respect to the defined work plan, raising any issues (internal and external) encountered, and ensuring other members are aware of significant events in each activity. The TMB is responsible for the approval of deliverables and milestones.
Decisions will generally be made by consensus. Where no consensus can be reached the issue will be forwarded to the PMB for discussion and a decision.
The TMB interacts indirectly with the EC and the PMB through the Project Coordinator. The TMB also interacts indirectly with individual participants through the Activity coordinators.

2.1.1.8. VRC Coordinator

Each VRC Coordinator is responsible for managing the people within his activity or activities to ensure that they carry out the defined work programme. He must ensure that deliverables and milestones are prepared on schedule and meet the high standards of the project. He is responsible for naming contacts with external projects and activities (for example with EGI's User Forum Steering Committee or Middleware Coordination Board). He is responsible for collecting information from these contacts and reporting that information to members of his VRC. The TMB must be kept informed of any significant advances or issues.
The coordinator manages his activity and interacts with all its participants. The coordinator is a member of the TMB and interacts with other coordinators through that body or directly.

2.1.2. Interactions with the European Grid Infrastructure (EGI)

This project is only one part of a full EGI ecosystem. Primarily it will interact with EGI and EGI.eu (the EGI management body) to ensure that our users computing needs are met by the infrastructure, that they have adequate support, that ROSCOE partners contribute to European-wide grid conferences, and that they help steer the evolution of the middleware used on EGI.

Because we wish to promote our VRCs as independent, stable, and sustainable entities, the VRCs will in most cases interact directly with structures within EGI. The VRCs are expected to directly participate in the User Forum Steering Committee, Grid Planning Board, and User Community Services. Similarly, common service activities like dissemination, training, and user support coordinators will directly interact with the appropriate personnel in EGI.eu. The Project Coordinator will interact directly with the EGI director.
Add a diagram showing the interactions between the VRCs and EGI.

2.1.3. Interactions with Other Projects

Additionally, we know of a large number of EGI-related projects that will be providing user support, application porting support, and middleware services. There will also be other projects in the Virtual Research Communities call that will have common interests and goals and may result in fruitful collaboration. Preliminary discussions have already been held with the following proposals:

- CUE: Creating Users of E-infrastructures
- TAPAS: Team to Assist Porting Applications to e-Science infrastructures
- SAFE: SSCs for Astronomy, Astrophysics, Fusion and Earth Science
- EMI: European Middleware Initiative
- EGI-SGI: EGI Science Gateway Initiative
- StratusLab: Combining grid and cloud technologies

Discussions on concrete collaboration between the projects will be held once it is known which are approved.
The ESFRI table earlier in the proposal should be duplicated here.

2.1.4. Consortium Agreement

The Consortium Agreement will define the general administrative and legal aspects: management structure, decision process, liabilities of the contractors, defaults and remedies, confidentiality, severability, disputes, and intellectual property rights provisions. Through specific annexes, it will address the specific technical and financial content of the project (list of affiliates, project plan, allocation of resources and background).

2.2. Individual participants

For each participant in the proposed project, provide a brief description of the legal entity, the main tasks they have been attributed, and the previous experience relevant to those tasks. Provide also a short profile of the individuals who will be undertaking the work.
(Maximum length for Section 2.2 – one page per participant. However, where two or more departments within an organisation have quite distinct roles within the proposal, one page per department is acceptable. The maximum length applying to a legal entity composed of several members, each of which is a separate legal entity (for example an EEIG), is one page per member, provided that the members have quite distinct roles within the proposal.)

2.3. Consortium as a whole

Describe how the participants collectively constitute a consortium capable of achieving the project objectives, and how they are suited and are committed to the tasks assigned to them. Show the complementarity between participants. Explain how the composition of the consortium is well balanced in relation to the objectives of the project.
If appropriate describe the industrial/commercial involvement to ensure exploitation of the results. Show how the opportunity of involving SMEs has been addressed.

i) Sub-contracting: *If any part of the work is to be sub-contracted by the participant responsible for it, describe the work involved and explain why a sub-contract approach has been chosen for it.*
ii) Other countries: *If a one or more of the participants requesting EU funding is based outside of the EU Member states, Associated countries and the list of International Cooperation Partner Countries, explain in terms of the project's objectives why such funding would be essential.*
iii) Additional partners: *If there are as-yet-unidentified participants in the project, the expected competences, the role of the potential participants and their integration into the running project should be described. (These as-yet-unidentified participants will not be counted in the minimum number of participants for the eligibility of the proposal).*

(No maximum length for Section 2.3 – depends on the size and complexity of the consortium)

This section must include a description of each VRCs consortium as well as the management. The consortium is intended to bring together the best people in each community to advance grid usage. Emphasize that the management is critical for forging cross-community collaboration and for making the transition to a more permanent core within EGI.eu.

Provide a list of the NGIs that are associated with your VRC. If you have a formal contact with the NGI, please indicate the person's name and email address.

VRC	**NGIs**
HEP	France (F. Malek), Germany (K.-P. Mickel), Italy (M. Mazzucato), Netherlands (J. Templon), NDGF (L. Fischer), Spain (M. Delfino), UK (N. Geddes)
LS	
CCMST	
CSE	
PS	Gauß Allianz e.V., Germany (Prof. Dr. Dieter Kranzlmüller, Kranzlmueller@ifi.lmu.de)
CS	HellasGrid, IAG
HUM	

Provide a concise description of your VRC's consortium and how those partners are optimal for achieving the goals of the VRC. Use the above comments to get an idea of what you need to describe here. Please keep the description as brief as possible, but you MUST convince the reviewers that your consortium is manageable and that all of the partners are useful.

2.3.1. HEP

The consortium consists of the following members: CERN, CESNET, DESY, the University of Birmingham and Imperial College, London, GSI, INFN and the University of Oslo. These members have either made significant contributions to the development and adaptation of the grid by HEP and other communities and/or are involved in future projects for which the use of grid technology is a cornerstone. The roles of GSI and DESY are focused on the use of grid for activities in which they play a leading part (FAIR and the International Linear Collider respectively), whilst all partners are directly involved in the LHC programme and specific support aspects that are related. For example, CERN, the Universities of Birmingham and Oslo and Imperial College are currently involved in distributed analysis support for two of the LHC experiments (ATLAS & LHCb) through their use of Ganga, whereas CERN and INFN are also involved in support for CMS and ALICE. This "distributed support" model reflects not only the nature of the user community but also is a solution that is likely to be sustainable for the long-term.

2.3.1. Life Science

Life Sciences community involves tens of thousands of researchers in Europe. From biology to healthcare, every single European country is involved and the big challenge for the LSVRC is to offer services and reach out to such a wide, heterogeneous and expanded community.
The LSVRC consortium aims at involving the institutes leading the deployment of biomedical grid applications in Europe together with key stakeholders of the Research Infrastructures that will structure the community.
The European Molecular Biology Laboratory (EMBL), member of EGI council, is a major partner of the LSVRC through the European Bioinformatics Institute (EBI), coordinator of the ELIXIR ESFRI. Around EBI, the Swiss Institute of Bioinformatics (SIB), the EMBnet network of bioinformatics laboratories and IBRB will develop a collaboration model between ELIXIR and EGI that will be used as a template for other ESFRIs.
In the field of structural biology, the e-NMR project is laying the foundations for the adoption of grid technology. The LSVRC involves the Magnetic Resonance (CIRMMP), major partner of e-NMR, which together with INFN will design a scientific gateway for structural biology.
In the field of medical informatics, there is no international centre of gravity comparable to EMBL so the strategy is to involve high profile institutes which are promoting the adoption of grids for healthcare: among them are University Politecnica of Valencia (UPV), the Health-Grid Association, Hôpitaux Universitaires de Genève (HES-SO) and the Academic Medical Centre of the University of Amsterdam (AMC).
The LSVRC also involves leading European institutes developing e-science environments already adopted by the life sciences community: among them is University of Manchester (Taverna, Myexperiment), EBI and CNRS (Moteur, Medical Data Manager, WISDOM).
Finally, the consortium involves partners with specific expertise for biomedical grid training (Consiglio Nazionale delle Ricerche) and relationship to supercomputers (CSC).
LSVRC consortium was designed to reach out to the life sciences community in Europe but its vision is worldwide: indeed, CNRS, HealthGrid, INFN and UPV are actively involved in regional grid projects in Africa, Asia and South America so that the services it will propose can serve at a broad international level.

2.3.2. CCMST

The CCMST Consortium is made by a group of laboratories, which have a tradition in supporting computational chemistry and materials science technology scientists. This type of organization is a direct evolution of that of the Computational Chemistry (CC) Cluster of Excellence of EGEE III that has gathered together the members of the already mentioned COMPCHEM, GAUSSIAN, TURBOMOLE and VOCE VOs coordinated by the laboratories of the University of Perugia, CYFRONET and the National Centre for Biomolecular Research of the University of Brno. Other coordination and support activities of the computational laboratories already operating on the grid are taken care by the CSC of Helsinki (M-Grid, NDGF materials science VO), Democritos of Trieste (eLab), ENEA of Rome (ENEA VO) and TUBITAK (TR-Grid). Further coordination activities in the field of quantum electronic structure, dynamics, spectroscopy and complex molecules properties are taken care by IRSAMC at the University of Toulouse, GDRQ at the University of Barcelona, the Institute of Electronic Structure and Lasers at FORTH of Heraklion and the Grid computing Competence Centre at the University of Zurich. The CCMST Consortium is made of a solid group of research laboratories (called hereinafter Support Laboratories" or SL) which, in addition to being already operating on the EGEE European grid by contributing with hardware, software, expertise and efforts, carry out management activities and provide support to grid operations by coordinating running VOs, sustaining grid efforts through their own skilled personnel and representing the CCMST field at various levels of the related National Grid Initiatives. This will ensure a robust frame on which the VRC will operate counting on an immediately operative set of infrastructures and volume of expertise. From the very beginning the CCMST VRC will count also on a set of dedicated (single or groups of) Laboratories having a variable familiarity with grid technologies. This further set of Laboratories (called hereinafter Community Laboratories or CL), which bear also a significant amount of computational chemistry and materials science technology know how, will be extremely useful for the VRC. They will in fact not only represent a useful test bed as users but they will also allow a capillary penetration of the VRC activities in the larger CCMST community because of their geographical spread. To date this group of CCMST supporters consists over 70 participants. (I don't know if we can put a whole list of them here – perhaps a map?)

The Supporting Laboratories:

- University of Perugia, Department of Chemistry
- Academic Computer Centre Cyfronet AGH
- National Center for Biomolecular Research, Brno
- IT Center for Sciente Ltd
- Democritos, ICTP, Trieste
- Department of Chemistry, University of Toulouse
- Computational laboratory for nanotechnologies and materials, ENEA, Rome
- Institute of Electronic structure and Lasers, Foundation for Research and Technology Hellas, Heraklion
- Grid Computing Competente Centre, University of Zurich
- Departamento de Quimica Fisica, University of Barcelona
- TUBITAK ULAKBIM High Performance and Grid Computing Center (unfunded)

2.3.3. Computer Science & Engineering

The CSE VRC consortium expands the Grid Observatory cluster partnership. CNRS/LRI was leading this cluster and ensures the continuity of the vision. Imperial College and CNRS/LAL bring their proven expertise in monitoring tools and software engineering to the data collection activity. CNRS/HG will contribute to providing a cutting-edge and production quality gateway. The long involvement of CNRS/LAL in EGEE governing bodies, and experience in the discipline of IT Governance of LOGICA, will contribute to addressing challenging, visible and feasible engineering issues. UNIPM and CU are major players in the area of Machine Learning and Autonomic Computing, thus are key partners in building bridges towards these community. Finally, UPJV provides the expertise, which was lacking in the Grid Observatory cluster, in the area of semantic organisation of the data, a key component of digital assets curation.

2.3.1. Photon Science

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 user, so that each European light source typically supports 10.000-15.000 within 3 years, which on one hand renders support an eternally ongoing effort. This exceptionally large number of user and research activities potentially benefiting from the PS VRC effort on the other hand implies a tremendous impact and outreach

The Photon Science (PS) VRC will support a variety of scientific communities performing experiments at light source facilities, ranging from biology, medicine, chemistry, material science to tomography or palaeontology. The support will naturally also include the institutes and bodies operating the experimental stations at the European Light Source facilities, since these will host the original experimental data and metadata, and presumably will be mostly responsible for data archiving, management and curation.

The Photon VRC aims to facilitate the creation of integrated facilities by combining instruments and algorithms both upstream and downstream of the actual experiment, namely the preparatory phase as well as the analysis framework.

The members of the Photon Science VRC are essentially all involved 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.

Table 30: Photon Science VRC and Relations to ESFRI Projects

Partner	Facilities	Services	Users/yr	PB/year	Communities	ESFRI Projects
DESY	31 BL at Doris3 14 BL at Petra3 6 BL at FLASH	All IT Services, e.g. AAA, DM, dCache, Grid Services	~ 5000	~ 5.0	Biology, Chemistry, Material & Surface Science, Drug Design, …	EuroFEL European XFEL
ELETTRA	25 BL 1 fermi@elettra	Computing, DM, AAA, Virtual Control Room	~ 4000	~ 0.5	Biology, Chemistry, Material & Surface Science, …	EuroFEL
EMBL	7 BL at Doris3 3 BL at Petra3	Crystallization, Computing, DM, Remote Access, Analysis Framework	~ 2000	~ 2.0	Biology, Chemistry, Drug Design	INSTRUCT
ESRF	51 BL	Computing, DM, Remote Access	~ 5000	~ 2.0	Biology, Chemistry, Material & Surface Science, Palaeontology, …	ESRFUp
ITQB	User Research Group (MX)	Crystallization, Analysis	-	~ 0.002	Protein Crystallography	
STFC	27 BL at Diamond 28 BL at ISIS	All IT services including DM, Meta-Data, HPC	~ 5000	~ 2.0	Biology, Chemistry, Material & Surface Science, Palaeontology, …	ESRFUp, ELI, HIPER, PRACE, European XFEL
XFEL	5 BL	Full Experiment support, incl. DM, Storage/Archive, Detector Development	-	10-100	Biology, Chemistry, Material & Surface Science, Plasma Physics	European XFEL
Total	188 beam lines		> 20,000	20-100		

2.3.2. Complexity Science

Currently, the Complexity Science VRC consortium is consisted of 6 European and 1 US partner. The European partners are Greece, Israel, Italy, Germany, Portugal and Sweden making the distribution of the consortium in geographical terms balanced across the continent. With respect to their scientific and technological background the partners stem from a variety of different fields of research forming together a unique interdisciplinary consortium focused on the advancement of Complexity Science as a whole.

In specific, the Greek partner has experience on running large scale simulations of complex systems, the Israeli partner is one of the key stakeholders in theoretical advancements of dynamics of complex systems, the Italian partner has expertise on the study of financial markets using complex systems theory, the German partner is focused on the aftershock impacts of rare and catastrophic events related to climate and earthquake studies, the Portuguese partner is one of the pioneers in the study of modern telecommunication networks and the Swedish partner specializes on social dynamics networks. The US partner is one of the pioneers in the field of networks and having written the first books on the subject he is an advisor to the US government for topics related to the themes of complex systems.

Through sharing of knowledge and the development and implementation of a series of services in the context of the Complexity Science VRC new EGI users stemming from the Complexity Science field will be identified and assisted into porting their applications and workflows on the Grid Computing Infrastructure. The collaboration of the consortium members on setting up and implementing a unified collaborative database of complex networks and structures will in addition bring together an even larger number of stakeholders of the field.
The management of the Complexity Science VRC will be performed by the Steering Committee that will include one representative from each European partner. A User Forum Representative, a person who will actively participate in the User Forum Steering Committee of EGI.eu, and a Grid Planning Officer whose responsibility will be to provide a long term technical planning will be assigned by the Steering Committee.

2.3.3. Humanities

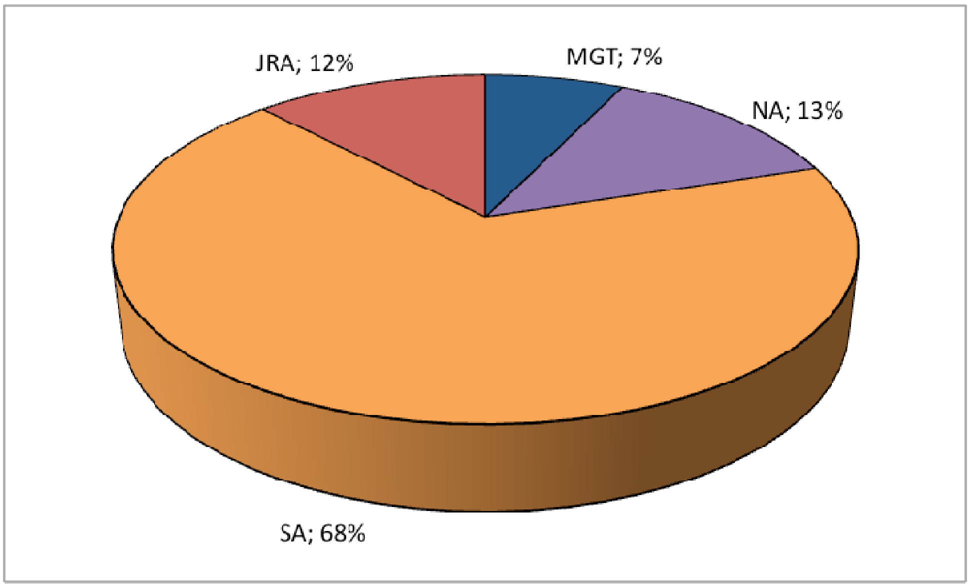
The Humanities VRC group consists of 5 European organizations and 1 with complementary interests and expertise from Taiwan. In Europe, these are Kings College London (coordinators), Goettingen State and University Library, the Max Planck Institute and the University of Catania on a funded basis, and the Vienna University of Technology unfunded. The Acadmeica Sinica Grid Centre in Taipei is likewise unfunded. Each partner brings individual strengths and connections to the consortium: KCL's Arts and Humanities e-Science Support Centre (AHeSSC) has extensive experience of collaborating with the humanities grind community in the UK and internationally. The Goettingen State and University Library (UGOE) is one of the largest scientific university libraries and one of the leading institutions in the development of digital library services in Germany. In close cooperation with national and international partners it provides infrastructure for research and education. The Max Planck Institute (MPI) is fully engaged in developing and applying advanced technologies for (distributed) data management for language resources and also taking part in associated policy development for community and EU based organizations. COMETA is a not-for-profit Organization established in Catania in 2005 whose funding members are the Universities of Catania, Messina, and Palermo, INFN (the Italian National Institute of Nuclear and Particle Physics, INAF (the Italian National Institute of Astrophysics), INGV (the Italian National Institute of Geophysics and Volcanology), and the Consorzio SCIRE. The Austrian Grid is a nationwide initiative to establish Grid computing in Austria. It combines Austria's leading researchers in advanced computing technologies with well-recognized partners in grid-dependent application areas. Academia Sinica (ASGC) is a leading scientific organization in Taiwan. AS will be presented in the project through ASGC (www.twgrid.org), one of its institutes, established as a leading high performance computing and communication centre in Taiwan, which provides advanced services for Grid Computing. These organizations have a shared history of working together at the nexus between the humanities and Grid computing.

2.4. Resources to be committed

Describe how the totality of the necessary resources will be mobilised, including any resources that will complement the EC contribution. Show how the resources will be integrated in a coherent way, and show how the overall financial plan for the project is adequate.
In addition to the costs indicated on form A3 of the proposal, and the effort shown in section 1.3 above, please identify any other major costs (e.g. equipment). Ensure that the figures stated in Part B are consistent with these.
(Maximum length for Section 2.4 – two pages)

The target for the project is to request 8.5 M€ contribution from the European Commission. The following table shows the target EU Contributions by scientific discipline (VRC). The project is following a "matching effort" policy such that each VRC may only request 50% of the maximum allowable EC contribution, reducing the cost of the project to the Commission and showing concretely the scientific communities' commitment to the project. The total cost of the project including the EU and partner contributions is approximately 28 M€. The pie chart shows the effort distribution by activity type, with the bulk of the effort going towards service activities.

Area	Total Cost	EU Contribution Current	Target	Fraction
Management	1,113,576	640,626	500 k€	6%
High Energy Physics	6,288,653	2,688,476	2700 k€	32%
Life Sciences	2,717,281	1,133,436	1000 k€	12%
Comp. Chemistry	2,257,096	1,001,095	1000 k€	12%
Grid Observatory	2,312,864	1,001,321	1000 k€	12%
Photon Science	2,944,800	1,390,188	750 k€	8%
Complexity Science	1,759,200	749,693	1000 k€	12%
Humanities	728,312	282,799	500 k€	6%
TOTAL	20,121,782	8,887,634	8450 k€	100%



3. Impact

(Maximum length for the whole of Section 3 – ten pages)

3.1. Expected impacts listed in the work programme

Describe how your project will contribute towards the expected impacts listed in the work programme in relation to the topic or topics in question. Mention the steps that will be needed to bring about these impacts. Explain why this contribution requires a European (rather than a national or local) approach. Indicate how account is taken of other national or international research activities. Mention any assumptions and external factors that may determine whether the impacts will be achieved.

Each VRC must provide input for the following impacts. Limit your response to one paragraph for each impact.

Increased effectiveness of European research through the broader use of e-Infrastructures by research communities.

(Strictly one paragraph per VRC. Be concise and to the point. Avoid vague or ambiguous propositions)

HEP: for a discipline such as HEP, centred on a few world-class accelerator laboratories, grid computing offers important advantages over politically less acceptable centralised computing. By allowing funding agencies to invest locally, whilst offering scientists working at laboratories and institutes throughout the collaborating countries previously unequalled access to data and world-wide computing resources, grid computing empowers not only the science but also these institutes – delivering significantly more value for every Euro that is invested. Whilst these are global characteristics of the grid that are to be enjoyed but all communities, they are of particular importance to HEP given its global nature and scale.

CS: One of the fundamental objectives of the Complexity Science VRC is to lower the barrier of adoption to Grid and in general Distributed Computing technologies and infrastructures through the application of well designed services and outreach activities. By achieving this goal we will further assist in increasing the usage of the e-Infrastructures. As a result collaborations among distant researchers in several fields of Complexity Science will blossom.

HUM: The barriers to using grid infrastructures are particularly high for humanists. The H-VRC will identify specific research activities being undertaken across Europe which demonstrate how use of the Grid can overcome these barriers and, through its research and service provision activities, build on these. It will seek ways link distributed humanities datasets semantically and quantitatively, and to enable humanists to approach certain fundamental research tasks without having to build e-infrastructure elements from the ground up.

PS: Broader utilization of the Grid and distributed computing environment by the PS communities will increase effectiveness in several ways: it will enhance computing opportunities for small, international research collaborations accelerating the analysis of experimental data; it will foster development and implementation of a sustainable data infrastructure; it will offer new opportunities for exchange and sharing of data between research groups and across disciplines; remote operation of instruments, standardized sampling of metadata and automatic distribution of data to the user communities will significantly increase effective usage of the PS infrastructure.

The emergence of virtual research communities of European and international dimension that cannot be achieved by national initiatives alone.

(Strictly one paragraph per VRC. Be concise and to the point. Avoid vague or ambiguous propositions)

HEP: given the global nature of High Energy Physics collaborations, this is a ***sine qua non*** condition. HEP collaborations are multi-national and require corresponding infrastructures. This includes collaboration and seamless interoperation with similar initiatives in other parts of the world (Americas, Asia etc).

GO: The gap between grid research and production grids is recurrently noted and deplored. The availability of large experimental datasets is critical to demonstrate the added value of research methods dealing with the adaptive management of uncertainty. Significant investment is needed for researchers to enter the logics of production grid, and to develop exploitation tools. This investment can be accepted only if the researcher can expect the source of data to be continued, and rejuvenated, allowing to follow the evolution of e-science practices. The insertion of the GO in the EGI ecosystem will help in building the confidence required for long-term collaborative projects, as is demonstrated by the letters of support.

CS: The CS VRC is composed of six European and one US partner. One of the Steering Committee main goals throughout its operation in the context of ROSCOE and beyond will be to expand by involving also other research partners in the decision taking and roadmap planning of the VRC.

HUM: As the H-VRC partner profiles show, there is already significant activity in this area at national level, much of which is represented by the H-VRC partners. The partners will therefore be in a unique position to identify commonalities at European level. Also, many humanities datasets which are relevant to common problems or questions reside in different countries. The Grid can help overcome such problems.

PS: most of the PS research is based on international collaborations; users are coming to light sources from all over the world. International collaborations between the facilities and the user communities are therefore indispensable.

Easier development and adoption of standards, common tools, procedures and best practices.

(Strictly one paragraph per VRC. Be concise and to the point. Avoid vague or ambiguous propositions)

HEP: the requirement stated above for global e-infrastructures is a very strong driving force for the development and use of standards and common tools. Ease of use and cost of ownership are the main driving factors for the development and adoption of common tools and best practises. This has already been shown to be of significant value in delivering reliable production services: further work in this field – within and across disciplines – is clearly required.

GO: The existence of publicly available datasets is a prerequisite for the emergence of benchmarks.

CS: Both in terms of computational workloads and data curation the CS VRC will enforce the usage of standards so that the work carried out within the ROSCOE Project will not be self contained but highly applicable and extendable to future research studies.

HUM: As expressed in the JRA1 tasks, the H-VRC will adopt a coordinated approach to metadata standards for collections and libraries of humanities data. It will seek to provide advice and good practice guides for the application of metadata in resource creation and use, and examine ways in which differing metadata schemas can be made interoperable.

PS: The PaN-data project aims to define standards and policies for the European Photon and Neutron sources. The PS VRC will support and deploy these standards for example through integration of standards and data storage solutions. Definition of analysis frameworks and extension of metadata standards offers the essential basis for the implementation of common tools, reducing parallel developments.

Use of e-Infrastructure services and tools by actors from new disciplines and scientific communities

(Strictly one paragraph per VRC. Be concise and to the point. Avoid vague or ambiguous propositions)

HEP: by lowering the cost of entry – both to service providers and to users – one can expect to attract new communities that would otherwise be deterred. This has been demonstrated in the past through the use of tools such as Ganga and Diane – both of which have been successfully used in a wide range of disciplines well beyond that of HEP in which they were born – as well as monitoring systems and gateways, such as those provided by the experiment, site and service dashboards.

CS: Through the series of services that will be developed and deployed in the context of the ROSCOE Project we plan to make the fundamental Grid services available to the wider Complexity Science community in a transparent manner.

HUM: Engaging new humanities users and promoting links with other disciplines is a key aim of H-VRC. As indicated in the engagement and training work packages, it will produce case studies and briefing papers illustrating success stories, with a practical emphasis linked to the service tasks on how that success can be replicated.

PS: Photon Science user communities hardly make use of European Grid or HPC computing infrastructure. The developments and deployment of the PS VRC and ROSCOE will hence *per definitionem* introduce the resources to a huge number of new users from a large number of disciplines.

Increased quality and attractiveness of e-Infrastructures.

(Strictly one paragraph per VRC. Be concise and to the point. Avoid vague or ambiguous propositions)

HEP: the combination of the above: tools and procedures aimed at delivering reliable, affordable productions services; user-oriented tools that simplify the porting of new applications to the grid and monitoring frameworks that provide global views to a wide range of heterogeneous activities, are all elements that will enhance the quality and hence attractiveness of e-Infrastructures. This can lead to a positive feedback situation: as communities are able to perform their work faster and more effectively the corresponding funding agencies obtain better value for their investments – both directly (via the resulting science) as well as indirectly (through longer-term socio-economic benefits). HEP provides significant added value through its requirements in terms of service reliability and availability, as well as the scale on which it pushes the limits (data storage, bandwidth, computational load).

CS: By making access to EGI and similar infrastructures transparent and more agile users will naturally be leaned towards making use of the underlying resources as they will be able to efficiently boost their research with a minimal barrier to overcome. The efforts of the CS VRC in the context of ROSCOE are targeted towards such an evolution.

GO: The GO VRC addresses the issues of (i) Improving usability of existing grid systems, and (ii) making them more reliable through its concrete objectives in JRA2, and more generally in liaising with the Autonomic Computing community. These activities will contribute to make possible that a wider community can consider use of such infrastructures, thus contributing also to the goal of Increased effectiveness of European research through the broader use of e-Infrastructures.

HUM: What constitutes both quality and attractiveness depends on the kind of researcher or research group involved. H-VRC will therefore base its approach to the humanities community on a thematic mapping of the humanities disciplines, based on a rigorous assessment of the current early adopter community's publications. It will then identify what aspects of e-Infrastructure is most likely to prove useful to each area, and tailor its coordination and service activities accordingly.

PS: Currently, the Grid infrastructure is not attractive to PS communities at all. Most users feel uncomfortable with the existing tools, portals and gateway. Integrating Grid resources into the SAXS analysis framework will certainly increase attractiveness of the e-infrastructure, which will serve as a basis to extend to other disciplines and possibly HPC infrastructures. An integrated data management and storage solution tailored for the rather special requirement from the non-HEP communities will significantly increase the quality of the infrastructure as well as the facilities themselves.

3.2. Dissemination, Exploitation, and IP Management

Describe the measures you propose for the dissemination and/or exploitation of project results, and how these will increase the impact of the project. In designing these measures, you should take into account a variety of communication means and target groups as appropriate (e.g. policy-makers, interest groups, media and the public at large).
For more information on communication guidance, see http://ec.europa.eu/research/science- society/science-communication/index_en.htm
Describe also your plans for the management of knowledge (intellectual property) acquired in the course of the project.

3.2.1. Dissemination Plan

Identify the concrete dissemination targets for each VRC. These should be in two separate areas:

- *Important institutes, user groups, or teams within the community that are not already involved in the VRC but need to be involved.*
- *Important scientific conferences that are important for having a visible ROSCOE/EGI presence and also those conferences, meetings, or schools that will be organized or supported by the VRC.*

For each entity that is identified, provide a short description (few lines) giving the level of participation and the reasons why it is important.

3.2.1.1. HEP

A key activity foreseen for this VRC will be to organize presentations of the progress and achievements of the e-infrastructure within the wider scientific and technical community and to the broader public. This would apply to the several major international conferences per year that bring together large numbers of scientists and engineers covering a wide spectrum of activities such as the conferences on Computing in High Energy Physics (CHEP). For the wider public this involves work with the CERN press office in releasing material intended for journalists and relating to progress in this area. In the last year three grid-related press releases have been made. When the LHC first started several hundred television stations worldwide participated and CERN has a high profile in the world's media so such releases have a strong impact.

The spreading of good practices, consultancy and training courses for new users are addressed through regular meetings and themed workshops – this is an on-going activity which needs to be continued, particularly during the critical early years of the LHC's operation. During these events the status of the services and the overall WLCG operations is reported and compared against the service availability expectations of the HEP communities (which in some cases can be around 99% for specific and critical services at large grid sites.) Standard operations procedures regarding service development, hardware management and maintenance have been largely discussed at several forums and are followed up on a regular basis with the grid sites that are supporting the HEP communities – constant vigilance is required to maintain the required service level. In addition, these procedures and standards have been shared with other international grid initiatives, also outside Europe. The goal of these initiatives is to spread the HEP operations requirements to other grid communities in order to establish stable collaborations as required by the corresponding user communities.

In the past the HEP community has led the creation and maintenance of grid user guides that have benefited the whole European Grid community thus contributing to dissemination of knowledge and internal / external communication. The maintenance and support of these guides as well as further introductions, FAQs and recipes will continue and will be essential as a growing number of non-expert users turn to the grid for analysis of the data produced at the LHC. The VRC will also maintain the existing level of effort in terms of presentations, participation to Grid Forums (regional and international), tutorials and courses appropriate to the tools supported by this community such as Dashboard, Ganga, storage solutions and so forth.

3.2.1.2. Life Science

The Life Sciences community is very large and heterogeneous in Europe. The LSVRC strategy is therefore not to try to go in all directions to try to answer to the needs of all the potential users but rather to target the most structured communities to serve them first in order to have success stories and experience to reach out to less structured ones. In the field of biology, the LSVRC has identified two communities, molecular and structural biology, which are now involved in the ELIXIR and INSTRUCT ESFRI design studies. The dissemination strategy is to develop the interface with these two ESFRIs and achieve a significant adoption of the research communities behind them. The success stories coming out of these collaborations will be used to reach out new groups.

The LSVRC is not going to organize its own conferences but rather contribute to the existing well-identified conferences in the targeted communities:

- In the field of bioinformatics, LSVRC will contribute to the EMBnet yearly conference
- In the field of medical informatics, LSVRC will contribute to the yearly HealthGrid conference through workshops and demos

LSVRC will support the yearly biomedical grid summer school initiated by CNR within the BioinfoGrid project.

3.2.1.3. CCMST

One of the primary goals of the CCMST VRC is the dissemination of grid know how among the members of the community of Chemists and Materials scientists and technologists.

As to dissemination the key action of the VRC will focus on mastering traditional communication channels (like presentation of lectures and communications at international conferences) and more modern communication (like those implementable using the web). As to the presentation of the VRC activities at conferences, the members of the support centre are tradi-

tionally presenting their results at the EUCO CC, WATOC as well as to the Computer science conferences like ICCSA. More formal communications will also be made through scientific journals of the field. On top of that dissemination will occur also on grid specific publications. As to the presentation of the VRC activities through more innovative channels a specific web site will be assembled and articulated in a series of pages illustrating the initiatives of the members of the community as well as those specific of the support centre. To spread the knowledge about initiatives undertaken by VRC a newsletter will be issued at least two times a year. It is also necessarily to use already available social utilities like Facebook. Such a places will not only easy spreading of information about VRC activities and achievements but also help to extend the community over existing supporting laboratories.
The CCMST VRC will also activate other two channels of dissemination to promote the innovation produced. The first one will consist in using the Theoretical Chemistry and Computational Modelling European Master to convey related information. To this end two types of presentations will be adopted. A general one via seminars addressed to all the students. A more specialized one addressed to the attendee of a specific section of the course Advanced Computing Technologies.
The most demanding commitment will be however requested by training. The training activities of the CCMST VRC will be focused on developing skills and competences of young researchers in implementing gridified versions of computer codes and packages of key importance for the chemistry and materials sciences and technologies researchers. To this end an annual hands on training school will be established. The key objective of the school is to provide young researchers with the ability of mastering on the grid advanced modelling and simulations based on multi-scale and multi-physics approaches starting from nanoscale level and consider gridification an integral part of the process of scientific discovery in the field of molecular and materials sciences. Important components of the training activity will be:

- The exploitation of collaboration through the composition on the network of different applications out of the large amount and variety of programs and suites of codes belonging to the CCMST area.
- The development of approaches based on the massive use of concurrent and distributed computing
- The establishing of a shared patrimony (library-like) of the community. This will motivate users to run their software in grid environment.
- The assemblage of a CCMST specific gateway made of scientific portals, graphical interfaces, workflows etc. suited to user friendly handle the various simulation packages, render output and intermediate data, select the optimum segment of the grid, import and link the necessary tools etc. for which the support of an appropriate organization based on specific skills and procedures is necessary.
- The standardisation of computational chemistry data and know how aimed at fostering interoperability especially in quantum chemistry.
- The creation of grid services based on stable versions of frequently used codes as black box products improving their user friendliness thanks to appropriate Application Program Interfaces (API)
- The increase of sustainability of the grid initiatives

A list of the most important of the mentioned packages is given in the Table below

#	**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	TURBOMOLE	Poland	Turbomole
14	DALTON	Czech Republic	VOCE
15	CPMD	Poland/Czech Republic	VOCE
16	GROMACS	Czech Republic	VOCE
17	NWChem	Poland	Gaussian
18	ABCD	Hungary	COMPCHEM
19	VENUS 88	Hungary	Compchem
20	ABCspli	Hungary	Compchem

A particular effort for dissemination and training will be made towards the higher education chemistry community. Special activities will be developed to enable young University teachers to develop on the grid services for a concurrent distributed management of teaching and learning units using the gridified version of the above-mentioned codes. The final goal of this activity is the establishing of a continental wide Learning Management System (LMS) in Europe.

3.2.1.4. Computer Science & Engineering

Institutes
We consider that interaction with standardization bodies and working groups, and institutions, in the area of (broadly speaking) Monitoring, Grid/Repository Integration, provenance, Grid/Web2.0 integration, and Usage Records, will be handled either by EGI, EMI, or ROSCOE common dissemination plans.

Engaging the Autonomic Computing community
The target institute is the NSF Centre for Autonomic Computing. The areas of mutual interest interactions cover the whole range of grid exploitation, from infrastructures to applications and including data quality. To be practical in addressing these issues we propose to investigate the creation of a knowledge base of Open Issues in Grid Operations and Applications, each entry including relevant and significant datasets, thereby enabling the specification *autonomic benchmarks*.

Engaging the Distributed systems community
The target organizations are

- COST Actions (2009-2013)
  - IC0804: Energy efficiency in large scale distributed systems. The objective of this action is to foster original research initiatives addressing energy awareness/saving and to increase the overall impact of European research in the field of energy efficiency in distributed systems. Excellent opportunities for synergies with the CSE VRC activity in SA2 and JRA2 do exist.
  - IC0805: Open European Network for High Performance Computing on Complex Environments. The main objective of the Action is to develop an integrated approach for tackling the challenges associated with heterogeneous and hierarchical systems for High Performance Computing. It thus addresses issues related with the CSE VRC activity in JRA2.
- OGF Working Groups
  - Grid Scheduling Architecture RG (GSA-RG), related to the CSE VRC activity in JRA2
  - GLUE Working Group (GLUE), related to the CSE VRC activity in SA2.

- The Grid Workload Archive initiative, which collects examples of execution of grid workflows, related to the CSE VRC activity in SA2.

Research grids, and more specifically the Grid'5000 research grid, and the associated Aladdin actions steered by INRIA. Multi-disciplinary projects supported by Aladdin target simulation/emulation of production grids, and specifically EGI. Such projects would constitute perfect test cases for the CSE VRC activity in SA2 and JRA2.

Strategic Conferences

Beyond common activities e.g. User Forum and (I think) SC

- IEEE International Conference on Autonomic Computing (ICAC) – the VRC should keep the associated workshop Grids Meet Autonomic Computing (GMAC) created in 2009 running. ICAC is the major conference in Autonomic Computing
- IEEE CCGRID - CCGrid is one of the major conferences in Grid research.

3.2.1.5. Photon Science

The PS VRC is prototypical for the European light sources and user communities rather than including all individual light sources. Dissemination to the European light sources will primarily be based on the participation in meetings and events of the PanData Project, which constitutes of all major European light and neutron sources. Envisaged extensions of PanData beyond Europe can enhance the outreach of the PS VRC likewise. Tight links to (ESFRI) projects like EuroFEL, Vedac, INSTRUCT or IRUVx guarantee dissemination at the level of individual beam lines or instruments. Through these cooperation dissemination to standardization activities, like data format policy and standards which play an important role in the PS VRC, and vice versa, dissemination of standards, will be possible.

Dissemination to the user communities will be based on established events. All of the PS VRC offers community specific training and workshops for the user communities. These events will be extended by the PS VRC dissemination events. Reference implementation by ITQB and for small angle scattering and the documentation of the reference installations will be an essential asset.

3.2.1.6. Complexity Science

Several important research groups that are not involved in the Complexity Science VRC but will be directly benefited by its operations have already been contacted and identified. The Complexity Science VRC will continuously be seeking new members of the community through the dissemination of its actions and the documentation of the provided services.

Addition of new members will enlarge and strengthen the community and ultimately drive it through a transition phase into a sustainable structure beyond the course of the VRC Project itself.

Three Training Events carried out by members of the Complexity Science VRC have already been planned while a constant representation of the VRC goals and achievements will be sought for during the upcoming EGI User Conferences. Dissemination regarding the usage of the underlying infrastructure and the services provided by the Complexity Science VRC will also be performed in the context of other scientific conferences that partners of the consortium will participate.

3.2.1.7. Humanities

H-VRC: The Digital Humanities communities are increasingly coming up against research challenges that could benefit from the use of grid resources. The H-VRC will therefore seek to identify and disseminate directly relevant humanities activities within EGI and, conversely, provide a focus for effectively communicating the implications and benefits of engaging with EGI to the digital humanities community. We will seek to ensure the humanities are represented at the User Forums. We will also use our existing partner networks, especially DARIAH and CLARIN, to reach out to unengaged and/or partially engaged communities who might be potentially interested in Grid usage at NGI level. Internationally, we will seek

to engage with professional bodies such as the Alliance of Digital Humanities Organizations (ADHO) and CenterNet, the international alliance of digital humanities centres (see http://digitalhumanities.org/centernet/). This will ensure that the H-VRC's activities are disseminated across Europe and beyond.
We will ensure a ROSCOE presence at the international Digital Humanities conference. This is the main international event for the DH community, and alternates every year between the US and Europe. Coincidentally, DH2010 is being held at King's College London, and we have been approached by the UK NGI for representation there: there is therefore an excellent base to build on in subsequent years. We will also represent the H-VRC's activities at EUAsia Grid, Finally, we will maintain a dissemination-oriented website, which will

3.2.2. Exploitation Plan

For ROSCOE, an important part of the exploitation plan will be the foreseen strategy for sustaining the VRCs. For this, it will be extremely important to identify the effort that will only be needed in the short term (e.g. setting up tools) and those that will have to continue indefinitely (e.g. maintenance of tools, providing user support). In addition each VRC Coordinator must provide a vision about how "sustainability" for the VRC will be achieved. This doesn't need to be definitive, but it does need to be credible. (In particular, continuous major funding by the EC is most likely not credible.)
For the exploitation, provide a list of concrete results of the ROSCOE work plan and their impact and future use within your community. Each item should be very short (maximum of a couple of sentences.)

3.2.2.1. HEP

The work described in this proposal is strongly related to the usage of WLCG: directly – as in the obvious case of the LHC experiments (as well as other data-taking experiments at CERN and elsewhere who are profiting from the same technology and support infrastructures), as well as future activities: the FAIR experiments at GSI that build in particular on the work down for the ALICE LHC experiment and studies related to the International Linear Collider that is expected to use the scientific results from the LHC as key input to its design. Thus a common priority is for the successful and smooth utilization of WLCG, building on existing infrastructures (EGEE, NDGF, OSG) and their successors in the EGI world. The WLCG operations and service model – whilst building on those of underlying infrastructures, extends significantly in both shared and experiment-specific areas. These include the daily operations meeting, on-going monitoring of services and links, escalation and reports to the WLCG management board based on Key Performance Indicators and analysis of exceptions, as well as longer-term reporting (quarterly, annual) that allow the status of the service as well as associated trends to be closely monitored. Service metrics – including targets for improvement – are established and followed up through these meetings. These have contributed measurably and significantly to service improvements and reductions in operating costs that are required for medium to long-term sustainability. An important element of the work that will be undertaken by this VRC will be to achieve further improvements in this area, benefiting not only those communities that are directly supported but others that adopt the same tools and/or service deployment and operations models.

3.2.2.2. Life Science

The proposed strategy for long-term sustainability is threefold:

- The Scientific Gateways for ESFRIs like ELIXIR and INSTRUCT will be progressively migrated to the ESFRIs for maintenance and operation
- EMBnet for bioinformatics and HealthGrid for medical informatics will promote and disseminate the LSVRC services within their respective influence spheres
- The LSVRC services will be operated by permanent staff from the National Grid Initiatives

The concrete results from ROSCOE work plan are the following:

- A technology for design and customization of scientific gateways
- Operational scientific gateways for Molecular Biology, Structural Biology and Medical informatics
- Working collaboration between ELIXIR and EGI
- Working collaboration between INSTRUCT and EGI
- An operational network of collaborators involved in NGIs and ESFRIs

3.2.2.3. CCMST

The proposed CCMST VRC is meant to push chemists and material technologists to take a significant leap towards distributed computing. The work proposed within the tasks described in previous paragraphs and ways to achieve them are based on present community needs and past experience and projects of the VRC partners in grid computing. It is worth to mention that CCMST is a high-ranking community in the utilization of distributed grid resources despite it rather marginal access to EGEE resources. For large number of community members outside of the grid computing (EGEE utilization – 8% in contrast to HPC centres – 40-70%) we have developed an extensive plan to foster collaboration between community members on European scale. We are confident that VRC is able to bring together a truly critical mass of expertise in chemistry and molecular and materials science and technologies and interface them with the ability of managing grid segments, e-infrastructures, scientific visualization and software s a service technology freeing the enormous potential frozen up to now by the lack of valid instruments. As a result of our work we expect:

- The structuring of the CCMST community in a way suitable to represent its interest in the grid community by grouping together scientific groups or laboratories having similar interest..
- The creation of a distributed support centre providing helpdesk, documentation, technical support, software database maintenance to the CCMST community for grid implementation of the CCMST codes. This is essential as such resources are widely spread among the partners. This also will help users utilizing different middleware resources get quick answer to the problems they experience
- A collection of CCMST programs geographically distributed on heterogeneous platforms to be used either on an individual or on a shared basis. Over the past years many scientific applications have been ported to the grid and can be accessed via different middleware. To help users we aim to gather together both documentation and scripts, tools accompanying ported software. If possible a unification of execution schemes under different middleware will be performed.
- A central (web based) interface for accessing and composing some computer programs and applications as web services. This is a point of interest especially for beginners not being familiar with command line tools. Such a service allowing easy use of grid infrastructure will not only speed up a process of grid adoption newcomers but also increase overall grid attractiveness
- A workflow manager finalized to the handling of grid empowered simulations and, when is the case, to compose out of them more complex multiscale simulators
- A prototype QoS and QoU evaluator of the services and the users of the CCMST VRC
- A prototype system of credit, award and redemption to support the users of CCMST VOs. Such a system is needed for example to reward the most active users donating and maintaining their own software codes for wider use.
- A repository for QC data complying with the model and standards enabling easy reuse of existing computational data
- An agreement with the owners of some widely CCMST programs for supporting shared usage and/or common formats

We expect that achievements of our goals will lead to vast exploitation of grid architecture of the level similar to exploitation of computing resources in HPC centres. The crucial part is to preserve sustainability of all the services and tools extended/developed during project duration. To achieve this all the partners started long time cooperation with NGIs they belong.

3.2.2.4. Computer Science & Engineering

Exploitation

- Provide production-quality services of *digital curation* of the internal production of the EGI grid, for computer science researchers and engineers in Europe and beyond, through data collection and access gateway.
- Build bridges between the operational requirements emerging from the new EGI model of operations and the computer science and engineering community, encouraging and facilitating the experimentation of new innovative ideas contributing to grid middleware and applications improvement in reliability, stability and performance.

Sustainability

The GO VRC is intended to be a stable entity whose primary goal is safeguarding and publishing datasets in the long run, and providing stable analysis tools. Consequently, GO VRC will evolve towards a permanent structure and define a sustainable financing model. The contribution of the ROSCOE project is decisive in funding the initial effort required to demonstrate production quality services.

The transition to sustainability has to take into account two facts.: 1) the activity is much younger than heavy-users scientific VRC, as the corresponding kernel in EGEE has been created only within EGEE-III,, 2) the Computer Science community has no international body comparable to CERN, ESA, or or flagship projects such as those recognized in the ESFRI roadmap. The closest approximation would be EGI itself, but it covers only the engineering segment. Overall, the VRC 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 CSE VRC must have support from stakeholders involved in actual production, such as some NGIs and EGI, and research institutes not yet involved in EGI, but prospective users of the GO data and services. The French NGI is committed to providing the bulk hardware resource and participating 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 enriching the knowledge base, and will be bootstrapped by the constitution of a network of experts in JRA2. Finally, in its short existence, the EGEE-III GO cluster has proven its efficiency at providing credibility for projects addressing national funding schemes. We are thus confident that the CSE VRC activity in the ROSCOE project will be able to set up the Grid Observatory as a sufficiently reputable product to attract sponsoring and be supported by the national research programmes.

3.2.2.1. Photon Science

The light source themselves will serve the user communities on a very long time scale. The continuous upgrade and expansion of the facilities and instruments warrants the continuous support and development. Apart from this intrinsic sustainability, the strategy for long-term sustainability is based on the participation of individual partners (e.g. DESY and STFC) in NGI's and related support projects. The participation in several ESFRIs and European projects guarantees long-term maintenance and operation of the proposed user portals, tools and standards. PS VRC will also greatly benefit from the scientific gateways and results proposed by the LS VRC like the working collaborations between INSTRUCT and EGI.

3.2.2.2. CS

The main focus of the Complexity Science VRC is placed upon the development and the production deployment of several services that will both lower the barrier of entrance to Grid and distributed computing technologies for new users and in parallel aid more advanced users to boost their research efforts. These tools and services will be developed within the duration of the ROSCOE Project. The subsequent maintenance and the continuation of user support services that will be provided in the context of the Complexity Science VRC will be managed by the partners of the consortium directly.

3.2.2.3. Humanities

H-VRC: Short-term effort will be focused on the exemplar implementations of the Grid data services that integrate smoothly with already existing individual repository/archiving systems and (2) Cooperating repository system networks that are or will be emerging within the H domain often pushed by other infrastructure projects. This integration will address but where possibly respect choices already made for solutions with respect to distributed datasets and metadata catalogues and interoperability, as well as persistent repository issues.

In the longer term, many of the issues surrounding the creation of sustainable research communities faced by the H-VRC reflect those encountered by digital resource creators and users in the broader Digital Humanities themselves. The H-VRC will identify, prioritize and articulate humanities projects of activities that produce high quality research that is publishable in prestigious journals, which are likely to sustain themselves through the normal academic publication and dissemination channels. Corresponding demand for e-Infrastructure services will be maintained; and we shall seek to discuss with NGIs how this can be managed in the longer term. While the H-VRC will target its technical resources towards facilitating use of EGI, existing user networks will be underpinned by (VRC-produced) publications and case studies. In the H-VRC's lifetime, these will presented at 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, which can be documented and referred to by established activities, and by nascent/new ones.

3.2.3. Intellectual Property Management

Please provide information about how intellectual property issues will affect your work plan or play a role in the "products"/"outcomes" from the ROSCOE project. I don't think that the project partners can impose any particular license or access restrictions for software/data to which ROSCOE contributes. However, anything that is created within ROSCOE should be done with conditions similar to the Apache2 license. There should also be a preference to working with open source products where the impacts are equal.

3.2.3.1. HEP

It is expected that any developments be made available under schemes such as GPL / LGPL, as has been the case in previous grid projects.

3.2.3.1. Life Science

This has to be discussed with each partner committing tools and services.

3.2.3.2. CCMST

The CCMST Partners have agreed to the following Memorandum of Understanding:

- Although CCMST VRC will deal with licensed software (both commercial and freely available) all the tools, plug-ins, documentation etc. will be created with help of the best suited and widely accepted open source software technology available, whenever possible

- If a partner decides to use, incorporate, or make the project's knowledge dependent on any pre-existing knowledge in the project, reasonable access rights shall be granted to other partners needing to use the knowledge during and after the end of the project in order to draw benefits from the project as a whole.
- All partners agree that they will continue to provide support, for a reasonable fee, after the end of the project, for the tools they have developed within or provide to the project, for a reasonable transition time after the end of the project.
- Knowledge shall be the property of the partner carrying out the work leading to that knowledge. Where several partners have jointly carried out work generating the knowledge and where their respective share of the work cannot be ascertained, they shall have joint ownership of such knowledge.

3.2.3.3. Computer Science & Engineering

The IPR issue for the GO has three aspects.

- All developments will be made available under public license schemes such as GPL / LGPL or Apache2, in conformity with the overall ROSCOE policy.
- It is expected that some analysis tools will be developed with the Matlab products (specialized toolboxes), for which no public domain equivalent exists. While releasing the codes does not suffer any legal limitation, discussion with MathWorks will be required to precise the legal aspects of actually using them trough the gateway.
- The datasets are publicly available for scientific usage. Redistribution of small portions of the data (or derived versions of them) is permitted if accompanied by a standard acknowledgement of the GO, EGI and ROSCOE, which will be specified on the gateway. The same acknowledgment is required for publications making significant use of the datasets.

3.2.3.1. Photon Science

The Photon Science communities use and depend on a large number of commercial products (e.g. IDL, Matlab, Mathematica and so) and closed source or copyrighted software, which often has to be licensed on a personal basis. Developments based on this products/software might therefore be subject to legal restrictions preventing release of the developments under a public domain license.
dCache developments will be subject to the dCache.ORG license agreement, i.e. remains free for academic and non-profit organizations.
IPRs of other developments will be subject to a formal cooperation agreement between the partners.

3.2.3.2. CS

Tools and services that will be developed and implemented in the context of the Complexity Science VRC will be licensed with conditions similar to the Apache license.
Several data sets and tables that will be included in the collaborative Complexity Science database may be proprietary or licensed under specific circumstances, and thus accessibility mechanisms (using VO roles and attributes) will be enforced and properly configured.

3.2.3.3. Humanities

The H-VRC's user community typically does not deal with industrial or commercially sensitive data; therefore IPR issues are not likely to affect its operation to any significant degree. Most humanities data resources are either publicly owned or publicly funded, and thus released under licences which permit re-use, at least in the academic sector.

3.3. Contribution to socio-economic impacts

Describe the socio-economic impacts of the project.

Briefly describe the social and economic impacts of the work done within ROSCOE. This should include the potential social or economic impacts of the scientific work supported by the VRC as well.

3.3.1. HEP

The past twenty years has shown significant advances in our understanding of the Universe, as described well in "The New Cosmic Onion", by Frank Close. Whilst the discoveries made during this period and refinement of the associated theoretical models can be justified in their own right, this work has a much broader impact. By continuing to attract young people to science and ensuring that Universities and research institutes have a broad and vibrant atmosphere, we continue to train a large number of young people whose subsequent careers – both of the relatively small fraction that continue in science as well as the majority who move into different fields – are of vital importance if European science and economies are to remain competitive over a period measured in decades. A small investment that encourages not only interdisciplinary and also multi-national collaboration – as is the case with the ROSCOE community – can have a major impact in this area.

On a somewhat more prosaic level, the expected results of this work are a marked increase in the number of grid users, as usage expands from the data processing activities that have dominated until now into the realm of data analysis, scientific discovery and publication. This will be accompanied by wider inter-disciplinary collaboration, both through science (i.e. related disciplines) and technology (e.g. grid tools). This can only be achieved by a significant simplification of user interaction with the grid, through further adoption of existing tools such as those described in detail below, and by a flexible and scalable end-user support model. This includes the establishment of community support, whereby the communities are encouraged and enabled to be largely self-supporting, with expert guidance to establish and optimize the support structures and associated tools. This is essential not only to deal with the large expansion in terms of number of users but also for long-term sustainability. These items will have a positive and measurable impact on the quality and effectiveness of the e-Infrastructure that in turn will lead to corresponding benefits to the research communities that use it. These activities will help to ensure Europe's leadership role in the areas of grid design, deployment and efficient exploitation.

3.3.2. Life Science

Healthcare is one of the most information-intensive sectors of European economies and can greatly profit from grid technology. Given that the health sector currently lags behind other sectors in the use of grids, there is great potential for rapid, sustained growth. The eHealth market is currently some 2% of total healthcare expenditure in Europe, but has the potential to more than double in size, almost reaching the volume of the market for medical devices or half the size of the pharmaceuticals market.

The LSVRC by fostering the wide adoption of the National production grids federated by the European Grid Initiative in the field of life sciences and healthcare will contribute to the expansion of eHealth in Europe.

3.3.3. CCMST

The structuring of CCMST prevented a representation of the whole community interests in the past. Actions and coordinating initiatives undertaken within VRC will ensure continuous support for community members, access to the e-infrastructure resources but also benefits from single representation of needs of majority of CCMST members. The key role of VRC is seen especially in case of community gathering. Ten supporting laboratories and over seventy cooperating ensure truly international character of the community.

The VRC is not meant to support CCMST community only. Vast majority of software packages provided within several VOs can be attractive to other branches of science enabling exhaustive explanation of phenomena and processes running in biological cells, living organisms, nature and universe. Also software vendors can benefit from VRC achievements, as for majority of frequently used software packages grid ports are available. Thanks to constant

maintenance new version of chemical codes find its way to grid infrastructure. Also provision of software packages as a service for the community is highly beneficial as new users adopt on the grid much faster and their scientific research is not distracted by new technology
To assure sustainability of community support, tools and services developed by VRC large-scale cooperation with NGIs has started. This forms a solid basis for long-term visibility of CCMST community on pan-European e-infrastructure.

3.3.4. Computer Science & Engineering

- The GO VRC is the only initiative in the area of grid digital assets curation. The size and the complexity of the grid data are a challenge in all the issues under the broader perspective of data curation. While this project had to limit its the partnership, industry (including SMEs) has expressed interest in the GO. The GO VRC will provide the core services, experience, and visibility required to build dedicated partnerships addressing integration of diverse data sources, management of ontologies, and empowering users via semantic web portals.
- The methodologies created in the analysis activity will be developed in relation to the issues of grid control and maintenance, application optimization, dimensioning and capacity planning, with potential contribution to EMI and EGI-operations.

3.3.5. Photon Science

Photon science communities produce huge amounts of data; the computational requirements are equally huge; the communities, the facilities and the instruments are very heterogeneous. Photon science can therefore greatly benefit from well-established grid technologies. However, only a very small fraction of the communities currently make use the e-infrastructure, so there is great potential for rapid, sustained growth. The establishment of a common framework can greatly enhance efficient use of the facilities, which can significantly enhance and accelerate the scientific progress. Since the PS communities are covering a wide spectrum of scientific areas. The PS VRC will support several tens of thousands of users every year and will have a strong impact and visibility in many fields. Since user communities, problems and methodologies are partially overlapping with other VRCs in ROSCOE, particularly LS and CCMST, synergies between research areas and grid initiatives will greatly enhance the outreach and impact of the PS VRC.

3.3.6. Complexity Science

The operation of the Complexity Science VRC within the ROSCOE Project will have two major impacts of socio-economic nature.
The formation of a large community of users from the field of Complexity Science functioning on top of the underlying Grid infrastructure will activate several interdisciplinary collaborations in the research of complex systems encompassing practically many fields of modern science.
The outcome of these large research collaborations, especially in the fields of climate and earthquake extreme events, social interaction studies and financial markets modelling, will directly influence the strategies for recovery efforts, after catastrophic or extreme events, that are currently in place.

3.3.7. Humanities

An important aspect of the H-VRC's community building work will be to identify areas where Grid resources can be used to increase the socio-economic impact of Grid resources. The importance of such impact of humanities research ('conventional' as well as digital) has been recognized recently (one NGI-level example can be found in the UK's Arts and Humanities Research Council's recent report, "Leading the world: the economic impact of UK arts and humanities research" [Ref. 18]). Yet how this impact can be measured and assessed is not currently well understood: indeed it is a matter of some considerable debate. However, there can be little doubt that linking dispersed and/or fuzzy resources in the humanities can improve and enhance access to them, including public access. The H-VRC will therefore strive to ensure that its work on interoperating collections and providing cross-metadata schema is, where possible, exploitable for public access.

4. Ethical Issues

Table 31: Ethical Issues Table

	YES	PAGE
Informed Consent		
• Does the proposal involve children?		
• Does the proposal involve patients or persons not able to give consent?		
• Does the proposal involve adult healthy volunteers?		
• Does the proposal involve Human Genetic Material?		
• Does the proposal involve Human biological samples?		
• Does the proposal involve Human data collection?		
Research on Human embryo/foetus		
• Does the proposal involve Human Embryos?		
• Does the proposal involve Human Foetal Tissue / Cells?		
• Does the proposal involve Human Embryonic Stem Cells?		
Privacy		
• Does the proposal involve processing of genetic information or personal data (e.g. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)		
• Does the proposal involve tracking the location or observation of people?		
Research on Animals		
• Does the proposal involve research on animals?		
• Are those animals transgenic small laboratory animals?		
• Are those animals transgenic farm animals?		
• Are those animals cloned farm animals?		
• Are those animals non-human primates?		
Research Involving Developing Countries		
• Use of local resources (genetic, animal, plant etc)		
• Impact on local community		
Dual Use		
• Research having direct military application		
• Research having the potential for terrorist abuse		
ICT Implants		
• Does the proposal involve clinical trials of ICT implants?		
I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL	YES	

5. Annex: Notes for Preparation

5.1. General

For the Gantt charts use "phases" of the activities rather than the subtasks. This requires that the descriptions be expanded slightly to describe these phases. The phases for each activity could be:

- NA1: tool/team setup, policies, templates
- NA2: tool/team setup, naming of contacts
- NA3: contact lists, meeting schedule, setup of UF, follow-up of UF, common booth for EGI conference, follow on specialized training events at conference, dedicated training events in scientific disciplines
- SA1: team setup, meeting schedule support, document review, production, organization
- SA2: initial operation, transition to standard implementation, maintenance, plug-in development
- SA3: selection of applications, porting, results, case studies, flagship applications

Community metrics:

- Number of registered VOs in the domain; numbers of users in the VOs
- Number sites supporting VOs in the domain; number passing QA tests
- Average reliability and availability of sites supporting a domain
- Number of registered applications in a domain, broken down by stage
- Amount of CPU used; number of jobs
- Amount of storage used; number of files
- Amount of bandwidth; number of transfers

Subjective progress of VRC:

- Scientific impact: papers, presentations, …
- Technology impact: integration of datasets, tools, …

5.2. HEP

As described above, a multi-disciplinary project such as ROSCOE offers an opportunity for results, both direct and indirect, that would not be possible in an environment that focussed on a single community. Experience has shown that a tool or technique that is able to support multiple communities is of greater value and often achieved at a lower total cost that the sum of those developed to address individual areas. HEP has a long tradition of working in such an environment and is strongly motivated to build on its past success.

6. Glossary

Term	Definition

7. Annex: References

. **Ref. 1**	http://journal.embnet.org/index.php/embnetnews/issue/archive
. **Ref. 2**	http://www.embnet.org/en/QuickGuides
. **Ref. 3**	http://education.embnet.org/
. **Ref. 4**	http://moodle.org/
. **Ref. 5**	https://gilda.ct.infn.it
. **Ref. 6**	http://www.bioinfogrid.eu
. **Ref. 7**	http://www.enmr.eu/
. **Ref. 8**	http://www.eu-eela.org/first-phase.php
. **Ref. 9**	http://www.eu-eela.eu
. **Ref. 10**	http://www.euasiagrid.eu
. **Ref. 11**	http://www.euchinagrid.eu
. **Ref. 12**	http://www.euindiagrid.eu/
. **Ref. 13**	http://www.eumedgrid.eu
. **Ref. 14**	http://www.adobe.com/products/acrobatconnectpro
. **Ref. 15**	http://www.adobe.com/uk/products/connect/productinfo/features/training/
. **Ref. 16**	http://www.iceage-eu.org/iwsgc09/index.cfm
. **Ref. 17**	ISO/IEC 38500 Corporate governance of information technology, (very closely based on AS8015-2005) provides a framework for effective governance of IT to assist those at the highest level of organizations to understand and fulfil their legal, regulatory, and ethical obligations in respect of their organizations' use of IT.

	ISO/IEC 38500 is applicable to organizations from all sizes, including public and private companies, government entities, and not-for-profit organizations. This standard provides guiding principles for directors of organizations on the effective, efficient, and acceptable use of Information Technology (IT) within their organizations.
. **Ref. 18**	UK Arts and Humanities Research Council, "Leading the world: The economic impact of UK arts and humanities research". http://www.ahrc.ac.uk/About/Policy/Documents/leadingtheworld.pdf
. **Ref. 19**	David-Alexandre Trégouët, et al, "A genome-wide haplotype association study identifies the slc22a3-lpal2-lpa gene cluster as a strong susceptibility locus for coronary artery disease", *Nature Genetics*, 41:3, 2009.
. **Ref. 20**	Tristan Glatard, Remi S. Soleman, Dick J. Veltman, Aart Nederveen, Silvia Olabarriaga. "Large scale functional MRI study on a production grid" in press in Future Generation Computer Systems, 2009.
. **Ref. 21**	http://www.computer.org/portal/web/computingnow/panel
. **Ref. 22**	"Livre Blanc sur les grilles de production : intérêt scientifique et utilisation", CNRS, 2009. http://www.idgrilles.fr/IMG/pdf/livreBlancecran.pdf
. **Ref. 23**	http://www.roadmaptgi.fr
. **Ref. 24**	http://www.mathworks.com/products/distriben/
. **Ref. 25**	http://www.scilab.org/

8. Annex: Letters of Support

Possible organizations from which to obtain letters of support:

- EGEE-III (requested)
- EGI (requested)
- ESFRI projects
- WLCG (requested)
- Center for Autonomic Computing (CAC)
- CoreGRID