ROSCOE

Robust Scientific Communities for EGI

Draft Vxx
19 October 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:

Part. no.	Participant organisation name	Part. short name	Country
1 (coordinator)	Centre National de la Recherche Scientifique	CNRS	FR
	CREATIS	CNRS (CREATIS)	FR
	University of Toulouse, Dept. of Chemistry	CNRS (IRSAMC)	FR
	Laboratoire de L'accélérateur Linéaire (U. Paris-Sud and CNRS)	CNRS (LAL)	FR
	Laboratoire de Physique Corpusculaire de Clermont-Ferrand (U. Blaise Pascal and CNRS)	CNRS (LPC)	FR
	Laboratoire de Recherche en Informatique (U. Paris-Sud and CNRS)	CNRS (LRI)	FR
	Modélisation, Information & Systèmes (U. de Picardie Jules Verne)	MIS	FR
	BT INFRASTRUCTURES CRITIQUES SAS	BT	FR
	INSTALLATION EUROPEENNE DE RAYONNEMENT SYNCHROTRON	ESRF	FR
	HealthGrid Association	HealthGrid	FR
	European Organization for Nuclear Research	CERN	CH
	Hôpitaux Universitaires de Genève	HES-SO	CH
	Swiss Institute of Bioinformatics	SIB	CH
	University of Zurich, Grid computing competence centre	UZH	CH

	CESNET, ZAJMOVE SDRUZENI PRAVNICKYCH OSOB	CESNET	CZ
	STIFTUNG DEUTSCHES ELEKTRONEN-SYNCHROTRON DESY	DESY	DE
	EUROPEAN MOLECULAR BIOLOGY LABORA-TORY	EMBL	DE
	GSI Helmholtzzentrum für Schwerionenfor-schung GmbH	GSI	DE
	JUSTUS-LIEBIG-UNIVERSITAET GIESSEN	JLUG	DE
	State and University Library Goettingen	UGOE	DE
	European Xfel GmbH	XFEL	DE
	University of Barcelona, Dept of Phys. Chemistry	GDRQ	ES
	University Politecnica of Valencia	UPV	ES
	IT Center for Science ltd	CSC	FI
	ARISTOTELIO PANEPISTIMIO THESSALONIKIS	AUTH	GR
	Greek Research Network	GRNET	GR
	Foundation for Research and Technology Hellas, Inst. Electronic structure and lasers	FORTH	GR
	Institute of Biological Research & Biotechnology	IBRB/NHRF	GR
	BAR ILAN UNIVERSITY	BIU	IL
	CONSORZIO INTERUNIVERSITARIO RISONANZE MAGNETICHE DI METALLOPROTEINE PARAMAGNETICHE	CIRMMP	IT
	Consiglio Nazionale delle Ricerche	CNR	IT
	Democritos, ICTP, Trieste	DEMOCRITOS	IT
	SINCROTRONE TRIESTE SCPA	ELETTRA	IT
	Ente nazionale energie alternative	ENEA	IT
	ISTITUTO NAZIONALE DI FISICA NUCLEARE	INFN	IT
	UNIVERSITA DEGLI STUDI DI PALERMO	UNIPA	IT
	University of Perugia, Faculty of Chemistry	UNIPG	IT
	UNIVERSITA DEGLI STUDI DEL PIEMONTE ORIENTALE AMEDEO AVOGADRO	UNIPMN	IT
	EMBnet Stichting	EMBnet	NL
	Logica	Logica	NL
	University of Amsterdam	AMC	NL
	UNIVERSITETET I OSLO	UIO	NO
	Academic Computer Centre Cyfronet AGH	CYFRONET	PL
	INSTITUTO DE TECNOLOGIA QUIMICA E BIOLOGICA - UNIVERSIDADE NOVA DE LISBOA	ITQB	PT
	University of Aveiro	UA	PT
	Stockholm University	SU	SE
	THE UNIVERSITY OF BIRMINGHAM	BHAM	UK
	Cardiff University	Cardiff	UK
	European Bioinformatics Institute	EBI	UK
	UK Computing for Particle Physics	GRIDPP	UK
	IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE	IC	UK
	King's College London (Arts and Humanities e-Science Support Centre)	KCL	UK

	University of Manchester	MAN	UK
	SCIENCE AND TECHNOLOGY FACILITIES COUNCIL	STFC	UK
	Boston University	BU	US
	Open Science Grid	OSG	US

Table of Contents

List of Tables

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.

NOTE: The following text is an abstract and not the final content of this section.

Context

Over the last decade, grid infrastructures have become indispensible scientific platforms, facilitating European scientific research from a diverse set of disciplines. The move from project-based organizations to sustainable ones will reassure researchers of their long-term viability and thereby increase their use and impact. The European Grid Infrastructure (EGI) will federate current cluster-based grids into a sustainable European grid while ensuring continuity from the existing grid infrastructures.

Aim of the Project

Sustainable funding levels will not permit EGI alone to fully support all of the grid's users and communities. The user communities themselves must fill the gap. This project aims to create robust scientific communities that are self-supporting and self-sustaining. These Specialized Support Centers (SSCs) in EGI jargon will be long-lived Virtual Research Communities that will coordinate their grid activities, safeguard the community's expertise, and provide specialized support.
ROSCOE targets seven scientific areas: High Energy Physics (HEP), Life Sciences (LS), Computational Chemistry & Material Science Technology (CCMST), Grid Observatory (GO), Complexity Science (CS), Photon Science (PS), and Humanities (H), including both large, mature communities and new, strategic scientific disciplines.

Work Program

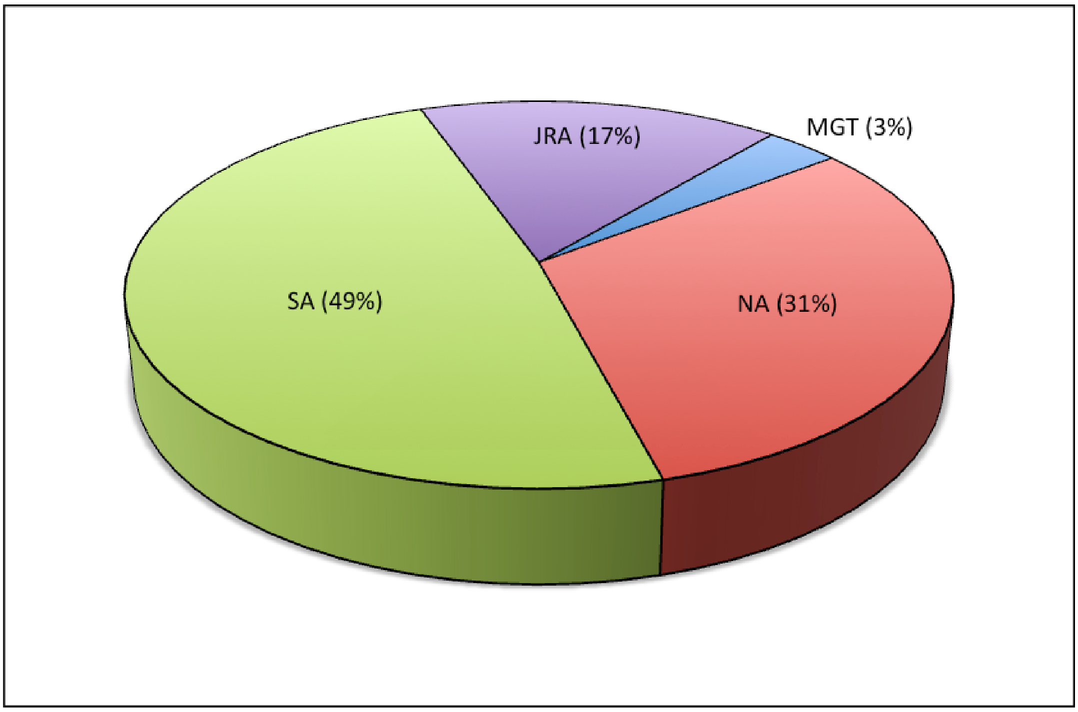
As required for I3 projects, the work program is divided into Networking Activities, Service Activities and Joint Research Activities. The Networking Activities include the project management, SSC coordination, dissemination, and training. These activities aim to expand the user community of each SSC, to ensure efficient communication within the community, and to make users aware of grid functionality and best practices.
The Service Activities include user support, scientific gateways, and targeted application porting. The user support will augment the standard support services with domain-specific knowledge. The scientific gateways will form the hub of the community providing information pertinent to the community and in many cases providing direct access to grid services and community data sets. The porting activity will target analysis frameworks, common libraries, or high-profile applications that will have a large impact on the use and visibility of the grid for the community.
The Joint Research Activities will revolve around topics of interest to several different communities such as tools (HEP, LS, CS, PS), molecular dynamics code and studies (CCMST, LS, PS), data mining of complex systems (CS, GO, HEP, H), large-scale data management (HEP, PS), and next generation data access and management (all areas).

Expected Impacts
The most difficult aspects of creating SSCs for EGI will be the social and political ones; the SSC coordination activities will directly tackle these problems and facilitate the transition to robust scientific communities on the grid. At the same time, the project will expand the reach of the grid by increasing the number of users, improving their grasp of grid technologies, and easing their use of the grid through its comprehensive work plan. The project will positively impact the multidisciplinary research already carried out in the seven targeted communities and similarly impact collaboration and cooperation between them.

1.1.1. Approximate Budget

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 (SSC). The project is following a "matching effort" policy such that each SSC 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	EU Contribution	Fraction
Management	500 k€	6%
High Energy Physics	2700 k€	32%
Life Sciences	1000 k€	12%
Comp. Chemistry	1000 k€	12%
Grid Observatory	1000 k€	12%
Complexity Science	750 k€	8%
Photon Science	1000 k€	12%
Humanities	500 k€	6%
TOTAL	**8450 k€**	100%



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.

Information about how/if people inside your community are/will be using super-computing resources (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 Center in Stuttgart – while the bulk of the simulations was subsequently run on the WLCG/EGEE Grid. To further increase the precision of the LatticeQCD 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.

1.3. Methodology

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

1.4. 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).

The Network Activities defined below are intended to foster a co-operation infrastructure between the participants of the ROSCOE project such that each participant can obtain a maximal benefit of the technical and administrative infrastructures provided by the project.
Two major actors, project coordination body and the specific SSCs, should have specific tasks in the context of the network activities being the project coordinator the link between the different participants.
The specific tasks for each major actor are defined as follows:

- Coordination infrastructure: Project coordination layer
    - This layer should also ensure the relation and communication of the participants with the EU
    - Fundamental bodies: PMB and TMB
        - PMB: Responsible of the daily operation of the project
        - TMB: consists of representatives of each SSC and will the technical forum for all participants
            - It will ensure effective communication among the different participants
            - Forum to share and propose common solutions for all participants; it will trigger proposals towards the integration of new solutions
            - It will define best practices in both security and operations terms for the participants and the use of the resources provided by the NGI
    - Additional facilities to be ensured by the coordination layer: agenda facilities, data bases, document access, event coordination, communities operation facilities, communication with other EGI-projects, etc
- In addition this layer should be the responsible of:
    - Dissemination of knowledge, external communication, coordination of inter-discipline meetings (also with other SSCs beyond ROSCOE) and specific documentation
    - Definition of common standards, protocols and interoperability usable by all participants
    - Common databases for both control of users and common tools and infrastructures. This body should ensure the spread of this information among all participants to avoid waste of manpower resources and to optimize and dissemination of commons tools also by other SSCs inside the EGI project
- Coordination infrastructure: the SSC coordination layer
    - The coordination of each SSC will ensure the establishment of internal training structures, regular meetings inside their own communities
    - Participation at the TMB, creation of regular reports and specific documentation and spread of this information through the standard commu-

nication forums established and fostered by the project coordination body
- In addition this layer will be responsible of:
  - Dissemination of knowledge among their users, internal communication at the SSC and at the general project layer
  - Trigger new studies and methods (management through the TMB) which can be re-used also by other communities.
  - Promotion of initiatives to national and international level (NGI) (fortering new communities and users entering the Grid through the corresponding SSC)
  - Spread of good practices among their users following the conclusions of the TMB

1.4.1. Overall strategy of the work plan

(Maximum length — one page)

1.4.2. Timing of the different WPs and their components

(Gantt chart or similar)

1.4.3. Detailed work description broken down into work packages

Table 1: 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	COORD		CNRS		M01	M36
NA2	SSC Coordination	COORD		CNRS		M01	M36
NA3	Dissemination & Training	COORD		EGI.eu		M01	M36
	TOTAL						

[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).

Table 2: Deliverables list

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

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).

Table 3: 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	CNRS	BT	GRNET	EGI.eu		
Person-months per participant:	72	18	18	15		

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.

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.

Managing and Monitoring Progress Towards Goals
ROSCOE is centred on having our Virtual Research Communities play a strong, visible role as SSCs within the European Grid Infrastructure. This task will monitor each community's progress towards becoming or effectiveness as an accepted EGI SSC. 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 reviews of the project, and to provide input on European programmes affecting our user communities.

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

Interactions with the European Commission with 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 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.
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. The coordinating partner will provide the effort for the financial reporting for the project at no cost. However, the Project Office will coordinate the gathering of the necessary information and following each partner's contributions.

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 control 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)

Probably quarterly or semi-annual reports focused on the progress of the project as well as highlighting any scientific or administrative issues.

Table 4: SSC Coordination (NA2)

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

Objectives

- Coordinate SSC activities
- Liaise with other EGI activities and EGI-related projects
- Facilitate sharing of resources
- Develop and implement SSC 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 Specialized Support Centres (SSCs). The scientific SSCs represent a given community within EGI, acting as a liaison between the community and the various activities within EGI. Additionally, the SSCs 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 SSC 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 SSC will cover one or more Virtual Organizations. As one of the primary reasons for using the grid is the sharing of computing resources, the SSCs will encourage sharing of resources within the covered VOs and between those VOs.
To understand its effectiveness, each SSC will have to evaluate its technical and scientific impact. This analysis will be done periodically, in collaboration with EGI and other SSCs.

For each area provide: the short name of partners involved and the associated effort (in PM) for each partner.

CERN will act as a liaison to WLCG and other HEP-wide coordination bodies as HEPIx (2 FTEs, unfunded)
OSLO will act as a liaison to EMI (0.5 FTEs, co-funded)

NA.HEP.1 Task 1: Service coordination and liaison with other projects
For the High Energy Physics SSC, activities are widely distributed both in content and geography. Many are continuations of previous WLCG efforts, in addition to new undertakings. Coordination of these activities will maintained through quarterly meetings of

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

the partners involved in the SSC. The coordination effort will be led by CERN, through general SSC coordination, daily WLCG operations meetings, weekly reports to the WLCG management board, preparation of quarterly service reports and regular topical and "collaboration" workshops. CERN will also play an active role in HEP-wide coordination bodies such as HEPiX, and coordination of the EGEE NA4 HEP cluster. Other HEP-wide coordination efforts include 18-monthly Computing in High Energy Physics conferences, the ACATconference series and so forth. (CERN)

NA.HEP.1 Task 2: Middleware coordination
As middleware stacks are a core component for the SSC's activities, liaison and coordination with EGI's Middleware Coordination Board and other relevant bodies is essential. Operations and middleware teams need to work closely together, and the middleware experts should provide the operations team with recommendations and procedures to take the maximal benefit of the middleware. Active communication between these parties will ensure both sustainability of the efforts and quality of services. The primary requirements of the SSC are related to the release and updates of middleware versions and Grid services. (OSLO).

NA.HEP.1 Task 3: Investigation and implementation of sustainability plan
A sustainability plan for the SSC efforts will be developed and put in to effect by the involved partners, through the quarterly SSC coordination meetings. JAMIE, WORDS HERE, SINCE YOU HAD INPUT FROM CERN? (CERN)
Partners and total level of activity:

- CERN will provide general coordination of the High Energy Physics SSC. Effort: 2.5 FTE (Fully funded by CERN)
- OSLO will provide middleware liason and communication with EGI's Middleware Coordination Board. Effort 0.5 FTE

Life Sciences
Coordination goals

- Integrate the LSSSC activities with NGI activities
- Develop synergies between NGIs
- Develop complementarities
- VO coordination, resource allocation and monitoring of use

Coordination with EGI dissemination SSC, NGI and Regional grid infrastructure dissemination structures (IBRB, INFN)
Coordination and management of the services provided by the Biomed VO

- Update list of VO members
- Resource monitoring
- Activity monitoring
- Integration of resources operated by ARC in the biomed VO

Computational Chemistry and Material Science Technology
The main management challenges are likely to be related to the dynamic nature of the SSC. Ongoing deployment and integration work, associated with the need to integrate and support our scientific partners in exploiting the SSC infrastructure, requires substantial agility of decision procedures. The management is structured in a way to allow a good communication and flow of information within SSC as well as between SSC and EGI or NGIs. The entities composing management structure will involve

- SSC chair elected by SSC Management Board, representing CCMST SSC in EGI User Forum

- Local coordinators (one from each institute or research group)

Front Desk responsible for a consulting Grid Observatory

The GO SSC 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 SSC will evolve towards a permanent structure and define a sustainable financing model.

Although we note that, 1) the activity is much younger than all other scientific SSC (the corresponding EGEE cluster has been created only within EGEE-III), and 2) the Computer Science community has no international body comparable to CERN, ESA, or even the large biomedical collaborations. Overall, the SSC is still in it ramp-up phase; thus it requires initial development funding, and has to invent a permanent structure and a governance model in the course of its existence.

The GO SSC must have support from, on one hand stakeholders involved in actual production, such as some NGIs and EGI, and on the other hand research institutes not presently involved in EGI, but prospective users of the GO data and services. The French NGI will provide the bulk hardware resource and participate in the operation tasks. The other NGIs are expected to contribute to the acquisition task (SA.GO.1) under the general operation scheme, thus will not have to provide dedicated human resources. Interpretation from operation experts is a requirement for JRA.GO.1 and NA.GO.1, and will be bootstrapped by the constitution of a network of experts. The GO SSC 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.

This section concerns the coordination of activities that are distributed amongst the work packages and will be described in the corresponding sections. The GO does not raise major management issues concerning the exploitation of the infrastructure: it is expected that the computing resource usage due to online analysis will not be significant, especially with respect to other SSCs. Storage usage is expected to be more significant, but remains modest with respect to other SSCs. Thus, the GO SSC will feature a light management structure, with one coordinator per partner. Partner coordinators form the SSC Management Board (MB) that is responsible for the effective and timely achievement of the SSC goals as described in the WPs.

The evolution of the MB by involvement of participants from other communities (Autonomic Computing, Computer Science, Grid engineering) is required. It is expected that a representation of these communities will be included in the course of the project. The strategic and organisational aspects of this integration are part of this task, while the selection of scientific themes is part of the GO SCC tasks Engaging the Autonomic Computing community and Engaging the Distributed systems community of NA3.

The MB will nominate an SSC Chair who will coordinate its activities and will represent the SSC in all instances where the GO SSC should be represented (including the EGI User Forum). The Chair will interact with the EGI administration coordinators, the EMI management, and all other SSC coordinators.

Complexity Science

Project coordination and liaison with other SSCs and EGI is an essential task for the SSC. The coordination of the Complexity Science SSC will be performed by the CS SSC Steering Committee which will be include representatives from each country that is member of the SSC. The Steering Committee will define 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 more long term technical planning and who will represent the SSC in the Middleware Coordination Board.

To ensure that the required progress is being made at all times during the course of the project the SSC 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 SSC. AUTH will be responsible for organizing the User support and CS SSC services and operations weekly meeting as well as the monthly conference call on Work Package Progress (18PM)

BIU, UNIPA, JLUG, UA and SU **will (in rotation) organize and host a CS SSC Face to face meeting (3PM each)**

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
NA.HUM.1 Community Engagement

Deliverables (brief description and month of delivery)

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

Table 5: Dissemination & Training (NA3)

Work package number	NA3	**Start date or starting event:**		M01	
Work package title	Dissemination & Training				
Activity Type[12]	COORD				
Participant number					
Participant short name					
Person-months per participant:					

Objectives

- Coordinate dissemination activities
- Disseminate SSC activities within the targeted SSC communities
- Coordinate training activities
- Provide focused training events for SSC communities
- Coordinate SSC 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 SSC'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 SSC, 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 SSC with cross-SSC 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 SSCs 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 SSCs at the annual EGI User Conference and at strategic non-EGI conferences identified by the SSCs 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 SSC dissemination contacts provide relevant materials.

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

For each area provide: the short name of partners involved and the associated effort (in PM) for each partner.

Coordination of support providers, namely experts from the VO user communities.

Significant communities using the grid require a dependable service. This is provided by distributed instances of general and VO-specific services. Each VO depends in a similar way from each of these services for their successful operation. Equally important the effort required at the supporting sites (deployment, operations, security audits) is again similar (if not larger) for the VO-specific services. VO-specific development (best practices including security) and deployment (coordination, sharing of information, sharing of existing tools and procedures) need equivalent quality standards in the VO-specific area.

Coordination of general and VO-specific training for end-users and support providers.Activity 3.1 and 3.2 need a solid foundation in terms of tutorials, documentation and trouble-shooting and developer guidelines. This activity provides a framework for the different communities to organise the documentation for the different components (e.g. tutorials need to be structured as a set of loosely coupled modules to address the needs of the different sub-communities. Each VO uses general services and specific ones. The overall schema and the general services parts should be the same).

CERN will coordinate the ROSCOE contribution to relevant HEP Computing conferences and schools (1 FTE, unfunded)

CERN **is actively involved in EGEE User Fora (and others), EGEE conferences and other HEP events, as well as Grid Schools and other "regional" events. CERN also provides regular input into e-zines such as iSGTW. Total effort: 0.5 FTE, fully funded by CERNLife Sciences**

Dissemination goals

- Integrate the LSSSC in the research communities targeted
- Participation to existing events
- Promote the services offered by the LSSSC to the potential users
- Brochures
- Participation to events
- Collaboration with other dissemination efforts
- Promote the activities of the NGIs and regional grid initiatives in the field of life sciences
- Contribute to EGI User Forums

Training goals

- Provide training to application developers, to the use of grid services and to the use of high level tools on top of grid services (TAVERNA, WISDOM, Vle, MOTEUR, etc.)
- Provide training to access and use Life Sciences SSC services
- Provide training to the use of Biomed VO specific services
- Integrate grid in existing international training programmes in life sciences
- Address the digital divide and foster the use of grids for medical development
- Pursue the existing yearly biomed grid school

Dissemination

- Organization of yearly LSSSC events in conjunction with other events
- Dissemination to the research communities involved in ESFRIs (BBMRI, INSTRUCT, ELIXIR, LIFEWATCH)
- Graphical design of LSSSC web portal and scientific gateways
- Contribution to EGI User Forums
- Promotion to life sciences user community including ESFRIs
- Joint demos and booths with regional grid infrastructures

Training:
- Coordination with EGI training SSC
- 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

Computational Chemistry and Material Science Technology
Dissemination is meant to focus on scientific communities promoting the SSC 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 escience development and grid enabling procedure in the area of its competence including:
- Simulation software (quantum-espresso and related projects see qe-forge.org portal for a full list) a
- 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 SSC focus will be both on promoting project results in general and on bringing these results to community members.
A key prerequisite for the successful dissemination of project results is a project website. This website will provide a public section with general information about the project and the consortium, dissemination of public material and delivery of various training material, be it general documentation, training videos, presentations and the like. Furthermore, a private section will serve the consortium members with internal documentation and the possibility of information exchange.

Grid Observatory
Engaging the Autonomic Computing 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. Experience gained from use of autonomic techniques via the Batch Queue Prediction (BQP) algorithm within the TeraGrid will also be made use of. The goal is to demonstrate the effectiveness of autonomic concepts for end users who utilize grid computing 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. There are multiple considering necessary: (i) demonstrating a reduction in cost when using the concepts; (ii) demonstrating that there was added value above existing tools; (iii) demonstrating improvements in QoS or fault tolerance, with a particular focus on how such QoS leads to improvements in the scientific findings.
It must be noted that the interaction with AC concepts covers the whole range of grid exploitation, from infrastructures to applications and including data quality. One of the key challenges to improve take up of autonomics was the need to empower end users and developers to address uncertainty. On the other hand, due to the very nature of problem solving approaches in autonomic computing, it is necessary to attempt solutions of specific problems, before attempting to generalize.
To be practical in addressing this issue, we propose to constitute a knowledge base of Open Issues in Grid Operations and Applications. Many of them fall in the scope of Autonomic Computing. The problem will be specified informally, but with formal performance criteria to be met. To facilitate better operational support using autonomic concepts, the corresponding entry in the knowledge base will include relevant and significant datasets, thereby relating to SA.GO.1 and JRA.GO.1. To facilitate better application use, publicly accessible Grid applications will be exploited to assess issues such as scalability and performance which could be improve through the use of autonomic computing algorithms. The outcome of these two activities will become a basis for specifying scientific challenges or "autonomic benchmarks". To make sense, the process has to be two-ways, and iterative, meaning that it has to include a process for specifying, and guarantying the significance of the datasets, as well as evaluating the potential solutions or hints provided by the AC community.

Engaging the Distributed systems community: The distributed systems community will be engaged primarily through interactions with the CoreGrid/ERCIM WG (follow-up of the CoreGrid NoE). As the overlap between this community and the autonomic computing is increasing., the activity of this task will be to create a process for gathering requirements from the Distributed Systems community toward the GO, both concerning the nature and presentation of the traces.
Another extremely interesting avenue is creating an experimental testbed for gLite. This is in relation with possible collaboration with the G5K infrastructure in France. An experimental facility would be a major consumer for the GO data. It must be stressed that the development of such an apparatus cannot be the responsibility of the GO SSC: specific projects addressing national funding schemes would have to be created. However, the responsibility of the GO is to engage in initiating this process, and participate in evaluating the feasibility of the facility, and the contribution of the GO datasets.

Strategic Conferences:
IEEE International Conference on Autonomic Computing (ICAC) – keep the associated workshop Grids Meet Autonomic Computing (GMAC) created in 2009 running.
IEEE CCGRID be present and in the PC.

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 community and strengthen it. Complexity Science SSC partners will constantly seek to publish the status and progress of the Project results through their participation in User Fora (expected to be organized in the context of EGI) and Complexity Science conferences.

AUTH, BIU, UNIPA, JLUG, UA, SU, GRNET and BU will contribute to the Dissemination Plan of the Complexity Science SSC (6PM each)
GRNET and BU will also contribute to the liaison of the CS SSC with external Projects (6PM each)

Two 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 SSC 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 CS SSC Knowledge Base (see SA1.CS.3 for further information).
JLUG and SU will organize and deliver the CS SSC Training Events (12 PM each)

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 a 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 SSC web-portal will be created.

Liaison with pan-European initiatives and projects: Dissemination to other PS facilities, beyond the partners of the SSC, 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
SA.HUM.1 Dissemination

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 6: Summary of staff effort

Participant no./short name	NA1	NA2	NA3	person months
CNRS				
BT				
GRNET				
EGI.eu				
Total				

Table 7: List of milestones

Milestone number	Milestone name	Work package(s) involved	Expected date[13]	Means of verification[14]
CS.1.1	CS SSC Face to face meeting	NA2	M08	The CS SSC Face to face meeting will bring together a group of 10-15 people involved in the Project in order to discuss the progress of the Work Packages.
CS.1.2	CS SSC Face to face meeting	NA2	M14	The CS SSC Face to face meeting will bring together a group of 10-15 people involved in the Project in order to discuss the progress of the Work Packages.
CS.1.3	CS SSC Training Event	NA3	M18	The Training Event will bring together a group of 30-50 people from industry and/or science in order to discuss the tools and services the CS SSC has to offer and

[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.

				the common usage scenarios of the Grid infrastructure
CS.1.4	CS SSC Face to face meeting	NA2	M20	The CS SSC Face to face meeting will bring together a group of 10-15 people involved in the Project in order to discuss the progress of the Work Packages.
CS.1.5	CS SSC Face to face meeting	NA2	M26	The CS SSC Face to face meeting will bring together a group of 10-15 people involved in the Project in order to discuss the progress of the Work Packages.
CS.1.6	CS SSC Training Event	NA3	M30	The Training Event will bring together a group of 30-50 people from industry and/or science in order to discuss the tools and services the CS SSC has to offer and the common usage scenarios of the Grid infrastructure
CS.1.7	CS SSC Face to face meeting	NA2	M32	The CS SSC Face to face meeting will bring together a group of 10-15 people involved in the Project in order to discuss the progress of the Work Packages.
M1.NA3.PS.1	Creating PS SSC web-portal	NA3	M3	Service released

M2.NA3.PS.1	Create training material	NA3	M9	Documents published
M3.NA3.PS.1	Provide training in user community specific workshops	NA3	At regular intervals until M36	Documents published
M4.NA3.PS.1	Disseminate to user communities at annual user meetings	NA3	At regular intervals until M36	Proceedings
M5.NA3.PS.2	Cooperation with ESFRI projects	NA3	M6	Support documents
M6.NA3.PS.3	Cooperation with-PanData	NA3	M6	Support document
M7.NA3.PS.3	Participation in semi-annual PanData and EuroFEL events	NA3	Bi-annual until M36	Meeting Documentation

1.4.4. Graphical presentation of component interdependencies

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

1.4.5. Significant risks and associated contingency plans

Table 8: Risks for Project Management (NA1)

Risk WP1	**Impact**	**Occurrence Probability**	**Mitigation**

Table 9: Risks for SSC Coordination (NA2)

Risk WP1	**Impact**	**Occurrence Probability**	**Mitigation**

Table 10: Risks for Dissemination & Training (NA3)

Risk WP1	**Impact**	**Occurrence Probability**	**Mitigation**
Low participation in the CS SSC Training Events	Enlargement of CS community will be hindered by low interest	Medium	In order to make CS SSC Training Events attractive to the wider Complexity Science community we will have to focus and disseminate upon success stories.

1.5. 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.

1.5.1. Overall strategy of the work plan

(Maximum length — one page)

1.5.2. Timing of the different WPs and their components

(Gantt chart or similar)

1.5.3. Detailed work description broken down into work packages

Table 11: 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						

[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).

Table 12: Deliverables list

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

[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 13: 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

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 SSC will maintain documentation targeted to its user community, concentrating on domain-specific applications, techniques, and data repositories. Often documentation by itself it 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 SSCs 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.
For each area provide: the short name of partners involved and the associated effort (in PM) for each partner.

High Energy Physics

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

Operational support for all of the above tasks 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 (2 FTE co-funded)

CESNET will investigate generic and sustainable implementation of LHC data analysis (Tier3) support (1 FTE, co-funded)

This work explicitly covers support for the WLCG, FAIR and ILC applications communities. Other HEP experiments will intrinsically benefit from the activities given the general nature of many of the tasks described below and the fact that the HEP community is highly cohesive – many people being involved in more than one experiment, often at different host laboratories.

Task 1: Testing of new middleware features and functionality in pre-production environments, as well as stress testing of key components following experiment requirements. This includes negotiation of service setups with various NGIs, computer centers and middleware providers, definition of the test environment, scenarios and metrics, development of the test framework, test execution and follow up.

Task 2: Offer general grid expertise for identification and solution of middleware issues as well as site configuration and setup problems. This includes a possible risk analysis and definition of action plans to prevent escalation to criticality.

Task 3: Development of experiment specific operational tools. Such tools include intelligent mining of grid monitoring data (for both workload and data management), automation of workflows and procedures, enforcement of data consistency across various services (storage and catalogues). This aspect is particularly important for the running experiments.

Task 4: Support for the integration of experiment specific critical services into the grid infrastructure. This includes service deployment, definition of escalation procedures and support models.

Task 5: Development and operation of frameworks to facilitate end-to-end testing of data management, production and analysis workflows. This includes functional testing integrated with SAM and VO specific monitoring and stress testing of real scenarios to investigate site and VO specific bottlenecks. Generalization of well established tools for service and site readiness validation to provide coherent information to each participating grid site.

Task 6: Investigation and deployment of solutions to enable an effective user-to-user and user-to-expert support model.

Life Sciences
Users in the biomedical community range from technology experts developing applications 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 web services in the research community
This task will focus on collecting and structuring information and providing first line user support for the access to VO core services in collaboration with EGI user support teams

Users developing and deploying Grid applications for health will surely meet doubts, problems and unsolved needs that might be solved by other experts in the field. There are different ways of providing support:

- Knowledge base. This will contain references to other general purpose documentation sources about Grid programming and deploying, but it will also 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 SSCs for driving developers on new generation components.
- Ticket-based support request. Tickets on unexpected behaviour, failures or usage doubts are a very powerful tool to help particular users and to contribute to the knowledge base. However, this approach has been unefficient 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).

Computational Chemistry and Material Science Technology
Front Desk (FD) is the technical unit responsible for several activities concerning consulting (for example direct interaction with application developers to get their application(s) running on the grid infrastructure), integration of CMST community resources with the grid infrastructure or assistance for application porting including integration of grid services necessarily to utilize the application in grid environment. Front Desk can also be used to spread information about the SSC among its members, to offer information about the membership to NGIs or consortia taking care of aspects relevant to the SSC.
User’s Support (US) is the technical unit responsible for User support. We can distinguish two main areas that User’s Support will have to deal with, namely Direct Users Support and Technical Support – responsible for all the services and tools needed to keep the infrastructure ready for utilization by users. The first one will be responsible for direct interactions with users including disseminations, trainings and first line support for users. SSC members involved in this task will also provide new and review existing documentation. Technical Support role is Operations related. Duties of Technical support will include VO registration, site validation tests, provision of core services as well as services specific to software needed by CMST community. Coordinators oft these tasks will closely cooperate with EGI User Technical Support Group as well as with middleware developers.

Grid Observatory

User support for the grid observatory has two faces. The first one is related to the usage of the gateway, and will be assured by HG. Given the fact that the target 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 LRI. Documenting both the data organization and the analysis facilities is essential for facilitating expert usage of the gateway.

Complexity Science
The Complexity Science SSC will set up a specialized support team that will provide user support services to the wider Complex Science community. The support team will take advantage of the helpdesk infrastructure provided by the EGI and will create a support unit specific to this SSC. The main task of the support team will be handle trouble tickets coming from the complex 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 (problem or query is too generic to be considered CS SSC specific or is out of the scope of the CS SSC Support Team). The Support Team will also provide answers to applications related queries so that best practices in the porting of applications are met. For advanced user questions related specifically to the porting of Applications to the Grid infrastructure the Application Support SSC will be contacted and assistance will be asked for.
In addition to providing answers and fixes to CS SSC specific user problems the Support Team will also maintain the Projects Knowledge Base making sure that related material is up-to-date and that new CS SSC services are properly documented. Answers to application related queries that will be specific to the CS SSC community will also be archived in the project knowledge base.
AUTH will supervise the Support Team Operation (12 PM)
BIU, JLUG, UA ans SU will participate in the CS SSC Support team (3PM each)

The Complexity Science SSC Support Team will consider producing documentation related to the CS SSC services and tools in novel forms of content such as podcasts and screencasts. The produced streams of audio and video content will be available online through the Knowledge Base (see SA2 for further information).
AUTH will organize and deliver the CS SSC Novel documentation sub task. (6 PM)

On top of the CS SSC Scientific Gateway we plan to implement a plug-in that will allow users to directly communicate with the GGUS helpdesk and through it with the CS SSC Support Team. This interface will ease the CS users in that they will not have to go through the GGUS helpdesk directly in order to submit a trouble ticket but instead use a more attractive and much simpler interface to ask for support.
AUTH will be focused on the implementation of the Support Team plug-in on top of the CS SSC Scientific Gateway and its further operation (3 PM)

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 a 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.

Humanities

Deliverables (brief description and month of delivery)

Table 14: 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 SSC portal
- Ensure that the SSC portal functions effectively for the target community
- Extend functionality of SSC portal to meet needs of the community

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

Each SSC 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 SSC/VO
- Monitoring of services
- Access to and management of data
- Access to grid services

Although the scope for each SSC 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.
SSCs with existing portals will continue to operate them until they can be migrated to the common accepted implementation(s). All SSCs will maintain documentation, information, and news that reside on the portal. The SSC will also ensure that the gateway functions well for the community and meets its needs. When used to access grid resources for an SSC, the SSC 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.

For each area provide: the short name of partners involved and the associated effort (in PM) for each partner. ***Split this effort into two categories: maintenance and development.***

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

High Energy Physics

CERN (2 FTEs co-funded]

The four main LHC experiments - ALICE, ATLAS ,CMS and LHCb – have individually developed specific workload and data management systems. In order to provide a global view of the status of LHC distributed computing, the experiment-specific systems and the generic monitoring frameworks need to be integrated.
The goal of this integration is to create a global monitoring infrastructure with dedicated portals providing a complete and reliable picture of the status of the LHC computing activities on the Grid and of the status of the distributed sites and services.

Tasks1: Development of experiment-specific plug-ins for existing generic frameworks such as Service Availability Monitoring (SAM), Service Level Status (SLS) and Nagios;

Task2: Enabling a common way of communication between various components of the monitoring infrastructure via the Messaging System for the Grid based on the Apache ActiveMQ message broker;

Task3: Development of monitoring portals serving the needs of various information consumers including the LHC user communities, WLCG support teams, WLCG management and support teams at the distributed sites.

Life Sciences
One of the major goals of the LSSSC is to enlarge the community of users of e-infrastructures in life sciences. It is crucial that users are able to access grid resources through different mechanisms. Through dedicated grid portals, users are able to deploy their application in a user-friendly way. But it must be understood that grid portals are only one entry point. Indeed, the user communities in the field of life sciences are being structured around ESFRIs: each of these ESFRIs is going to have its own distributed computing and data infrastructure to which the LSSSC must be able to propose the right services. The LSSSC partners are directly involved in ELIXIR (distributed infrastructure for molecular biology), LIFEWATCH (infrastructure for biodiversity), INSTRUCT (infrastructure) and BBMRI (Biobanking and Biomolecular Resources Research Infrastructure). The objectives of this activity are to provide to users of these three ESFRIs as well as to rest of the community services for scientific data analysis on EGI.
The services to be provided by this work package are the following

- scientific gateways under the form of grid portals for on demand access to grid resources
- tools and services to create grid-aware or grid enabled bioinformatics and medical informatics web services for execution on e-infrastructures and integration into pipelined analysis

Provision of these services requires a number of tasks:

- for the scientific gateways
    - definition of the list of requirements for each of the communities and ESFRIs targeted
    - Selection of the SG implementation technology. This will be done in col-

laboration with SG dedicated projects
  - integration of existing tools and services into the scientific gateways: design and conception of plug-ins
  - Customization and maintenance of the scientific gateways
  - Access to the BioCatalogue of Web Services and extension to Grid Services
- for the provision of grid-enabled bioinformatics and medical informatics web services
  - definition of the list of requirements
  - design and implementation of tools for wrapping bioinformatics and medical informatics tools into web services for execution on the grid

Definition of requirements and liaison with the project providing the portal implementation(s)
Customization and maintenance of the scientific gateways for the molecular biology community
Customization and maintenance of the scientific gateway for the structural oilogy community (CERM & INFN)
Customization and maintenance of the scientific gateway for the biodiversity community (HealthGrid ?)
Customization and maintenance of the scientific gateway for the Healthgrid community (HealthGrid -HESSO)
Customization and maintenance of the scientific gateway for the medical imaging community (AMC)
Customization and maintenance of the scientific gateways fo the genetics population study community (CNR)
Customization and maintenance of the scientific gateway for the microarray analysis community (IBRB)

Description of the services to be integrated in the scientific gateway
SOMA2
SOMA2 is a web browser based workflow environment for computational drug design and general molecular modeling (http://www.csc.fi/soma). Purpose of the SOMA2 environment is to provide users an easy access to computational tools. SOMA2 hides all technicalities related to execution of scientific applications in complex computing facilities. This allows users to focus in their actual scientific tasks. SOMA2 provides user friendly and intuitive WWW interface to applications that user can seamlessly connect into application workflows where several applications are automatically processed one after another. The SOMA2 environment also processes data provided by applications and offers direct result analysis view within the system. The SOMA2 environment is designed so that integration of new applications and tools into the system is easy. SOMA2 includes interfaces for, e.g., protein docking software (GOLD)
Vbrowser
The Virtual Resource Browser (Vbrowser, http://www.vl-e.nl/vbrowser) is an interactive graphical front-end and framework to access grid resources (both data and web services). It is developed by the Informatics Institute of the University of Amsterdam.
MOTEUR
MOTEUR is a workflow management system developed at the CNRS Sophia Antipolis (http://modalis.polytech.unice.fr/softwares/moteur/start). Workflows can be started from a Vbrowser plugin and enacted on grids using various middleware (gLite, ARC). The platform MOTEUR+Vbrowser is currently successfully adopted by various biomedical researchers in France and The Netherlands, and it could be extended to a larger community.
e-NMR platform

The e-NMR platform is a comprehensive ensemble of integrated web services that are aimed at structural biologists, particularly those making use of NMR spectroscopy as a toll to investigate protein structure. It also includes applications for the investigation of macromolecular adducts that potentially exploit a wide variety of different experimental data. This platform will be embedded into the scientific gateway for the structural biology community and supplemented with services relevant also to other applications in structural biology such as x-ray crystallography. The options for the development of plug-ins within the gateway as well as the requirements by the structural biology community regarding middleware will be explored by CERM and INFN in collaboration.
Luciano: description of LINKAGE
BioCatalogue (http://www.biocatalogue.org) is a expert and community curated catalogue of web services that are relevant and useful to the Life Sciences, which takes over from the EMBRACE registry. It is developed jointly between the University of Manchester UK and EMBL-EBI; the latter host the catalogue. The catalogue is a REST-based service itself with read and write APIs. We propose to (a) register and promote Grid Services and (b) incorporate the BioCatalogue into the gateways.
myExperiment (http://www.myExperiment.org) is a community-sourced repository and social networking environment for scientific workflows of any kind of workflow system. It is developed jointly by the University of Manchester and the University of Southampton. It already holds workflows developed by the Healthgrid community. The catalogue is a REST-based service itself with read and write APIs. We propose to (a) register and promote workflows such as those developed for MOTEUR, Taverna and other workflow systems and (b) incorporate myExperiment into the gateways.
Taverna (http://www.taverna.org.uk) is an open source scientific workflow management system designed to link together service based resources and enact dataflows. It has been widely adopted throughout Europe the USA, South America and SE Asia. The development is primarily at the University of Manchester. It has plugins for gLite (developed by CNRS), ARC (developed by KnowARC) and Globus Toolkit 4 (developed by Argonne Labs/Manchester). We propose to consolidate the ARC/gLite execution from Taverna.
WISDOM
The deployment of large scale data challenges for in silico drug discovery since 2005 within WISDOM initiative has led to the development of a dedicated framework with specific features:
interoperability: the data challenges have involved many grid infrastructures around the world so the framework was designed to allow easy deployment on multiple infrastructures
scalability: up to several thousands CPU must be simultaneously loaded and monitored
distributed and secured data management:: input and output data must be securely stored according to a complex data model
The WISDOM production environment has been developed as the result of a collaboration between EGEE-III and EMBRACE and is used for large scale docking and bioinformatics analysis.
GRISSOM
The GRISSOM portal(GRids for In Silico Systems BiOlogy and Medicine) (www.grissom.gr) enables exploitation of GRID resources for DNA microarray distributed processing. It provides experts with a complete web-based solution for generic management, search and dissemination of biological knowledge in the context of gene expression patterns on a genomic scale. The platform is developed and deployed using open source software components. GRISSOM supports versatile analysis for both cDNA and oligonucleotide (Affymetrix/ Illumina) microarray data, encompassing among others data import, filtering, normalization, statistical selection, annotation, clustering, gene ontology based pathway analysis. The underlying algorithms are parallelized through the use of either MPI computing or Direct Acyclic Graph (DAG) Scheduler for optimal

performance and flexible grid deployment. Through the use of web service technologies (WSDL language) GRISSOM can be encapsulated in other biomedical processing workflows through Taverna workbench, providing transparent access to its algorithms. The GRISSOM portal integrates a local repository of microarray data, complying to both MIAME and miniML (MIAME Notation in Markup Language) annotation systems. GRISSOM foresees advanced security mechanisms regarding access control and data encryption, in order to ensure proper usage of grid computational resources and entrust data security.

Computational Chemistry and Material Science Technology

This package provides services that support researchers in their daily work. In this activity, a robust, easy to use, web portal will be adjusted to community needs and maintained. Together with the portal a set of plug-ins to CCMST packages and suit of codes will be provided to promote 'software as a service' model of computing.

Particularly, CSC develops and maintains SOMA2 - a web browser based workflow environment for computational drug design and general molecular modeling (http://www.csc.fi/soma). Purpose of the SOMA2 environment is to provide users an easy access to computational tools. SOMA2 hides all technicalities related to execution of scientific applications in complex computing facilities. This allows users to focus in their actual scientific tasks. SOMA2 provides user friendly and intuitive WWW interface to applications that user can seamlessly connect into application workflows where several applications are automatically processed one after another. The SOMA2 environment also processes data provided by applications and offers direct result analysis view within the system. The SOMA2 environment is designed so that integration of new applications and tools into the system is easy. SOMA2 uses the Chemical Markup Language (CML) as the internal data format. QC5 and CML share common features and as such can be made to work together.

UNIPGCHIM and ENEA will also develop tools for providing molecular science codes as web services.

Grid Observatory

Data acquisition: The primary role of the GO SSC is acquisition and long-term conservation of the monitoring data produced by the EGEE grid about its own behavior. The SSC will continue its approach of 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. The GO SSC will thus limit its activity to exploiting their results, with one notable exception. Exploiting the results will take three paths:

- Enabling the general deployment of the acquisition tools prototyped in the GO cluster of EGEE-III, as a certified component of the gLite middleware. Another data source of particular importance is the Real Time Monitor acquisition system, developed and operated by IC, which provides a summary of the gLite-monitored grid activity.
- Long term conservation of the monitoring data collected by HEP experiments, currently gathered at CERN, which are so far discarded after their immediate operational use has passed.
- Other SSCs, and specifically HEP and Life Science, have built and exploited specific monitoring services (e.g. DASHBOARD), or services equipped with monitoring facilities (e.g. DIANE, GANGA, etc). These traces may give access to alternative exploitation models of the grid as well as additional semantic information, especially in the area of diagnosis.

The first two activities involve active collaborations with the operations of EGI , and the first one involves collaboration with EMI.

The notable exception quoted above is the acquisition of data related to power consumption, where acquisition tolls will be developed. Due to limited access to such information, the research in optimization is often limited, with researchers focused on a small-scale sub-problem that could be simulated. This might be a point of particular interest for interaction with the Cloud computing community.

The GO gateway: The GO gateway is the visible part of the project. In EGEE-III, the GO portal has been built as a trace repository. The goals of the gateway are as follows:

- Scale access to much larger communities and provide more comprehensive datasets;
- Present data utilizing additional semantic information;
- Provide analysis facilities.

Scaling access both quantitatively and qualitatively is a major challenge for a sub-community of computer science. The first step is to re-structure the datasets according to standards, either event-oriented or resource-oriented, for which standards exist or are in progress (e.g. the Common Base Event for event-oriented data). This corresponds to "lossless" compression. The final goal would be to provide full-fledged database facilities, allowing for dynamic presentation of data along the needs of specific users. This is an extremely difficult issue, because it combines 1) high-performance requirements (on-the-fly operations over massive datasets) and 2) the need to make the database schema evolve without waiting for the finalization of the grid ontology to structure the data description. A more realistic goal will be to build the technical specifications and requirements associated with the deployment of the GO database, in order to engage the process of getting the required support from the French NGI on a sound technical basis.
Analysis facilities will initially share codes developed by the community, either within or outside the GO. The effort will be put onto structuring and documenting the code produced inside the GO. A more ambitious scheme is to propose on-line facilities, ranging from basic statistics to the exploitation of stabilized analysis methods. The implementation of the Matlab distributed engine on EGEE will be exploited.

Complexity Science
The Knowledge Base 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 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 authenticated users to contribute with their thoughts and ideas as time progresses. Eventually the Knowledge Base will become the documentation repository of the CS SSC containing the necessary information for both new and advanced users of the Grid infrastructure stemming from the CS community.
Use cases and success stories related to the Complexity Science SSC will also be provided through the Knowledge Base. Documentation in the form of web content (such as screencasts, podcasts and recorded webinars) will in addition be available through the Knowledge Base.
The Knowledge Base will be a part of the CS SSC Scientific Gateway.
UNIPA will develop and maintain the Knowledge Base (12 PM)

The CS SSC will be responsible for managing and maintaining the information stored under the VOMS interface(s) supporting the Complex Systems VO(s). Thus the CS SSC 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 SSC 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 (3 PM)

We plan to design and deploy a web portal that will serve as a point of entry for new users. Through this portal we plan to provide registration forms for all the steps a user has to complete prior to using the underlying Grid resources.

Thus depending on the country the researcher is based in we plan to provide well documented guides on how to acquire an IGTF approved personal digital certificate signed and issued by the corresponding CA. Printable template forms required for the identity vetting procedure will also be provided.

The subsequent 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 SSC 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 SSC 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 access a User Interface in just a few steps.

AUTH will develop the Registration Portal (6 PM)

UNIPA will maintain the Registration Portal (6 PM)

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 SSC 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 plan to design and develop a UI front-end, which will be available through the CS SSC 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 will develop the UI module (6PM)

BIU will maintain the module operation (6PM)

Within this sub task we will develop 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.

BIU will develop the Database module (6PM)

UNIPA will maintain the service (3 PM)

On top of the CS SSC Scientific gateway we will implement a Resources monitoring mechanism so that users are notified at close to real time of CS SSC 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 (6 PM)

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:

- 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.
- 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 to explore local and distributed computing infrastructure likewise.

Security and fine grained access control: Experiments at light sources are often highly competitive, data are exclusively owned by the individual research collaboration (at least until publication) and data as well as metadata need to remain fully protected for a time which 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.

It remains however rather unclear if such schemes can be deployed in the PS environment, dealing with extremely large data volumes, analyzed by a huge number of individual, international research collaboration. Particularly in case of the European XFEL, where the available bandwidth it barely sufficient to export data to users and/or national data repositories, potential bottlenecks introduced by data protection mechanisms, for example based on encryption, need to be avoided.

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 favored 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. For example, federated ID's are commonly valid nation-wide but not outside a country. Trust mechanisms between facilities located in different countries seem to be lacking. Short living certificates permit to operate on the grid, but it's unclear if such mechanisms conform to the requirements of fine-grained authorization. Grid-proxies can too easy be hijacked poking severe security holes into federal authentication; OpenID completely lacks mechanisms to retract authorizing cookies along an authentication chain.

Envisaged action include therefore

- Evaluation of existing data protection solutions in a PS environment.
- Support and integration of suitable solutions into existing middleware.
- Development of data security solutions tailored for a PS environment.

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
JRA.HUM.1 Task 1: Develop a community portal for Humanities for EGI
JRA.HUM.1 Task 2: Develop an open repository infrastructure for EGI

Deliverables (brief description and month of delivery)

Table 15: Targeted Application Porting (SA3)

Work package number	SA3	**Start date or starting event:**		M01	
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 SSC 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.

For each area provide: the short name of partners involved and the associated effort (in PM) for each partner.

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

High Energy Physics

The contribution to the GANGA evolution is from CERN (1 FTE co-funded), IC (1 FTE co-funded), BAHM (1 FTE, co-funded), UIO (1 FTE, co-funded)

The contribution to the integration of LHC experiments frameworks is INFN (2 FTEs, co-funded), CERN (1 FTE, co-funded)

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. The required effort will focus on enabling Ganga to support a sharp increase in the number of user communities and of users within the existing communities. The following tasks are foreseen (involved 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.

Life Sciences

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

Actors in the LS SSC 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.

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 SSC 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
Computational

Chemistry and Material Science Technology
The Work Package will focus on porting applications to the Grid, providing support on MPI environments on clusters and grid-enabled supercomputers, and giving technical user support on the general usage of the infrastructure.
Molecular and materials science applications often demand high amount of computing time, thus making parallel computing crucial in achieving results within feasible time. Parallel computing is an aspect, which has so far not been in focus in distributed computing, although development of multicore processors will make this inevitable. More and more supercomputers are also available through Grid middleware and thus MPI applications are very relevant also for Grid user communities. Thus one task of task WP is to give support on the MPI environments in Grid infrastructure as well as contributing to support of applications parallelized using MPI.
The applications ported to the Grid will be selected in such a way that they are of particular use for material scientists and that they require large computing resources.
Test runs for novel applications and with applications requiring large amounts of resources will be run.
The WP will also aim to parallelize and optimize grid-enabled codes within materials science. The WP will select codes that either are already ported to grid or would benefit from grid usage given possibility to parallel runs through the grid (utilizing, e.g., OpenMPI). Execute test runs to find out best parallelizing methods and to demonstrate the speed-up achieved through parallelization and optimization.
Provide support on MPI environments on clusters and grid-enabled supercomputers.
Give support for the VOs who have been granted access to the resources governed by the SSC. Support service is given in the general usage of the infrastructures, such as job submission and obtaining certificates, as well as in using specified applications within the field of materials science.
A close cooperation link with Application porting SSC will be established to utilize their resources in order to jointly provide stable versions of chemical codes on all middleware platforms supported by UMD.

Grid Observatory

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 in a few easy steps. 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 numerical algorithms on top of that dataset. Due to statistical deviations that arise in these sorts of systems the re-application of the algorithms on top of the same or similar datasets is required so in order to fully evaluate the value of a needed parameter even in this simple case study would be largely facilitated by having the ability to design the workflow in advance. The subsequent post processing of the results and visualization thereof could and should also be considered as the final part of such a workflow.
Such workflow design scenarios that will optimize the usage of the underlying Grid resources will next be added as a tool developed and implemented in the context of the CS.SA.1 Work Package to the Scientific Gateway.
UA will lead and manage this sub task (6PM)
BIU will identify workflow patterns in Complexity Science studies (6 PM)
AUTH will participate in the implementation of the workflow design (3 PM)

We plan to develop and deploy the “application as a service” concept on top of commonly used Complexity Science applications. In this context users will be able to focus on their study and spend less time on setting up or porting their applications to the Grid infrastructure. In this course we will have to identify heavily used applications by the Complexity Science community and provide them on top of computational resources provided to the community in the form of services. Users will then easily perform parameter studies and engage specific applications easily into more complex workflows. A close collaboration with the Application Porting SSC will be asked for in the context of this sub task.
BIU will lead and manage this sub task (6PM)
AUTH will be involved in the identification of commonly used applications in the field of CS and participate in the implementation of applications as services (6 PM)

A thorough search will be carried out in order to identify parts of the most used Complexity Science algorithms such as the Network Analysis, the Detrended Fluctuations Analysis, the Wavelet and the multifractal DFA algorithms etc that consume a large amount of computational time and parallel counterparts of these algorithms (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 libraries so that Applications making use of these algorithms may benefit directly proper library calls.
AUTH will lead this sub task (6 PM)
UA will identify commonly used algorithms and participate in the optimization sub task (6 PM)

Photon Science
Operational Support: The PS SSC members are involved in several grid activities, e.g. serving as Tier 2/3 centers,. 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 SSC 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 SSC 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 are however a limited number of defined, de facto standards, 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

- integrate dcap/gsidcap IO into HDF5
- integrate dcap/gsidcap IO into CIF

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 deployment.

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 show case for a number of different aspects of the PS SSC.

Adaptation and maintenance: The PS SSC 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 RedHats 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

JRA.CS.1 Task 3: Scope shared text and geo-mining services

Deliverables (brief description and month of delivery)

SA3.CS.1.1: Results on CS SSC specific workflows development and implementation (M24)

Identification of workflows that are commonly used in the context of Complexity Science and technical documentation of the related implementations on top of the CS SSC Scientific Gateway.

SA3.CS.3.1: Report on parallel versions of common Complexity Science algorithms (M36)

Identification of commonly used algorithms in the study of Complex Systems and documentation on the optimization and parallelization techniques implemented. Documentation on the usage of the libraries build will also be part of the Report.

Table 16: Summary of staff effort

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

Table 17: List of milestones

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

1.5.4. Graphical presentation of component interdependencies

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

1.5.5. Significant risks and associated contingency plans

Table 18: Risks for User Support (SA1)

Risk WP1	**Impact**	**Occurrence Probability**	**Mitigation**
Low impact of novel documenta-	Only a few users benefiting from	Low to medium	Disseminate upon the novel documentation

[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.

tion content to CS SSC community	novel documentation		material within the community

Table 19: Risks for Scientific Gateways (SA2)

Risk WP1	Impact	Occurrence Probability	Mitigation
Scientific Gateway provided by collaborating projects will not meet with the CS SSC 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.

Table 20: Risks for Targeted Application Porting (SA3)

Risk WP1	Impact	Occurrence Probability	Mitigation
Low scaling or no benefit from parallel versions of CS SSC 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 fulfill users expectations	high	Coordinate with external parties (regional programmes) for training of additional manpower

1.6. 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.

1.6.1. Overall strategy of the work plan

(Maximum length — one page)

1.6.2. Timing of the different WPs and their components

(Gantt chart or similar)

1.6.3. Detailed work description broken down into work packages

Table 21: 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	Contributions to sustainable operations model; Grid Ontology; Quality and Credit System	RTD				M01	M36
	TOTAL						

[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).

Table 22: Deliverables list

Del. no.[34]	Deliverable name	WP no.	Nature[35]	Dissemination level[36]	Delivery date[37] (proj. month)

[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).

Table 23: Data Management Evolution (JRA1)

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

Objectives

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

High Energy Physics
JRA.HEP.1: Investigation of future data management technologies

Life Sciences

Computational Chemistry and Material Science Technology

Grid Observatory

Complexity Science
JRA1.CS.1: Scientific Database
We plan to design and implement a Scientific Database on top of Grid services and tools. The Database be available to all users of the Complex Systems SSC 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 (18 PM)

AUTH, UNIPA, JLUG and SU **will provide collaborative databases and participate in**

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

the design of the database scheme (6 PM each)

Photon Science

Investigation of future management technologies: Data and storage management solutions in Photon Science 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 behavior will not be working at all any longer. Due to the development of the detectors and the brilliance of the light sources the amount of data is increasing drastically. For the 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, in contrast to the LHC these facilities are running in a burst mode fashion. It is than expected and following to a certain extend the operational model of the LHC, that part of the data will be copied to appropriate large sites, which will act as something like a Tier 1 or Tier 2. The whole management of the data in terms of catalogs, security etc (and data formats, which is not a direct part of this proposal) have carefully to be investigated. This has to include the European Geant networking infrastructure as well.

Due to the diversity of the Photon Science Community, 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.

So, the key elements of this tasks are:

- investigation of the industrial and open source technology development in storage standards and solutions
- Cooperate with industry and open source in the development of standards like NFS, AFS, Lustre etc.
- Cooperation with other communities, facing similar problems
- Develop of end-to-end data datamodels for photon science including a distributed data management structure
- Develop data management solutions for large scale data in photon science

Humanities

Deliverables (brief description and month of delivery)

Table 24: Contributions to Sustainable Operations Model (JRA2)

Work package number	JRA2	**Start date or starting event:**		M01	
Work package title	Contributions to sustainable operations model; Grid Ontology; Quality and Credit System				
Activity Type[39]					
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 methods for analyzing, representing, and correlating the data, based on Machine Learning, Complex Systems, and Data Mining.
- Explore operational on-line usage of these methods, in order to evolve towards an information provisioning service organization.
- Build and exploit a grid ontology, accounting for both structural and dynamical aspects of grids at different levels of abstraction. It includes applications to data consolidation and semantic inference. This will contribute in the longer term to promoting a sheer data curation model for the overall EGI.
- Provision Quality and Credit System to judge and reward users/research groups highly contributing their resources (software, hardware, knowledge) for other members of CCMST community

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

High Energy Physics

Life Sciences

Computational Chemistry and Material Science Technology

The development and consolidation of the SSC is meant to be structured as a true cooperative endeavour through the creation of instances and mechanism inducing effective collaboration. This will be achieved by designing and implementing some procedures exploiting those features of the SOA approaches that allow a structuring of the services offered in a way that permits the evaluation of the parameters needed to quantify their quality (QoS). The parameters monitored to evaluate QoS will be of both objective and subjective type. Subjective types of evaluation parameters will be based also on procedures quantifying the quality of user (QoU). Both quality indices will be employed to drive the activities of the SSC towards its objectives and in particular to enhance collaborative efforts. To this end they will be connected to a system of credits awarding and redeeming that will assign selectively the resources of the SSC. In particular, a first pro-

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

totype implementation of the system will be applied to the provision of computational codes to the SSC for shared usage and for the composition of more complex computational procedures.

Grid Observatory

Models of the grid dynamics: Based on the acquisition of grid traces (SA.GO.1) and the representation of the grid domain (JRA.GO.2), the objective of this task is to model the dynamics of the grid. Grid, seen as a complex structure, has its own emergent behavior that we examine using techniques from the areas of Complex Network Analysis, Machine Learning and Data Mining. This task should give some new insights into grids and provide a better global picture of the system and its behavior. The found correlations and distributions should help grid scientists and managers to obtain a better understanding of the relationships that emerge in such a complex system, and provide the basis for their modeling.

The challenges of modeling the grid dynamics are threefold.

Firstly, the complexity and heterogeneity of the grid requires, in order to accurately modeling its behavior, i) to consider massive amounts of traces; ii) to use scalable algorithms and /or to exploit the grid itself to avail the computational resources needed. Secondly, as mentioned earlier on, the model accuracy depends on the quality of the representation and on the representativity of the data. Thirdly, the final goal is to provide an understandable model of the grid, allowing the system administrators and end users to exploit this model; therefore the model should be able to "explain" its output, or provide some insights into the typical uses of the system (e.g., clusters of users).

This task is mostly basic research-oriented. The motivation for such an activity in the SSC is to keep SA.GO.1 and JRA.GO.2 in-line with the users needs, which are an enthusiastic but yet fragile community. As building models of grid network using Machine Learning techniques is still in its infancy, gaining first-hand experience on the three above-mentioned challenges is required.

Considering the internal goals of the GO SSC, the implicit topology structure defined on the space of grid events through grid models should be reflected in the ontology built in JRA.GO.2. Similarly, the navigation tools constructed in SA.GO.1 must allow for acquiring information that is sufficiently precise to ground models, and parsimonious enough to allow for a wide range of experimentations in the end-user community.

Considering the user community, models can be used to generate test data for simulating grid behavior in future research (the target is here the computer science community), 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 behaviors in the system (the target here is the grid engineering community). Finally, demonstrating scientific advances is critical in building bridges between the grid engineering community and the Autonomic Computing community, 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.

It is thus necessary to actually tackle selected research issues. Two axes will be developed: complex networks as a general model, and statistical inference applied to fault detection, diagnosis and explanation.

An information service organization: Information Technology (IT) Governance focuses on performance of IT systems and risk management. Industrial governance standards are captured in ISO 38500[40], which strongly focuses on managing the IT resources on behalf of stakeholders who expect a return from their investment.

[40] 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 fulfill their legal, regu-

The specific focus of this sub-task is *governance support by intelligent monitoring and learning agents*. Given the complexity of grid infrastructures, automated support to processes of information retrieval and analysis has become necessary. As explained previously, an extensive monitoring infrastructure does exist: gLite logs, and user-level software (e.g. HEP experiments), as well as the generic monitoring environment/plug-in for Nagios. We thus focus on the exploitation of the output of these monitoring tools, in an operation-oriented perspective.
Whereas the acquisition and interpretation of monitoring within individual domains is done superfluously, correlation between domains is not commonly done yet. In collaboration with Subtask JRA.GO.1.1, we develop methods to correlate event-information between sites. We research how 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 interests are automatic feedback loops that enable near-real time failure identification and remediation.
From the technical point of view, we intend to develop Nagios plugins that implement such functionality. A challenge thereby is that we have to take into account that Nagios plugins are not statefull.
From the organizational point of view, we want to focus on both, administration as well as user perspectives. Having generic EGI goals in mind, we therefore foster relationships with other SSCs, such as for example the Life Science SSC.
The interaction with EGI-operations and UMD with ensure that the tools may be deployed on the live infrastructure.

Grid Ontology: For the construction of the ontology, many resources are used as data: (i) existing termino-ontological resources on grids (GLUE, which is the basis for interoperability between the EGEE grid infrastructure and other grid infrastructures e.g. the Open Science Grid project in the US) will be considered as a main reference resource); and (ii) native traces and results from the modeling of grid dynamics.

Ontology building: The focus of this task is to transform and enrich the GLUE schema, which is expressed as a UML model, into an ontology based on logical descriptions of concepts in order to carry out inferences. The ontology will cover concepts already present in the GLUE (physical resources, components and services) but it will also include concepts about logical resources, jobs and their lifecycle and, generally speaking, the dynamics of EGEE, all kinds of concepts needed to reason on the traces. Considering these last resources, which are not engaged in a standardization process, this activity will contribute to avoid the risks associated with storage format evolution and obsolescence.
The developed ontology will be based on the foundational ontology DOLCE, the core ontology of programs and software COPS, and integrate, when appropriate, existing grid ontologies (covering mainly structural aspects).
The informal descriptions associated to the entities and relationships structuring the conceptual model GLUE v. 2.0 are modeled in order to get a more formal and semantically richer model than the actual class model in UML. In parallel, the model is expanded with temporal entities to account for the dynamics of EGEE. Using DOLCE and the concepts coming from other termino-ontological resources will enable restructuring the concepts coming from GLUE v. 2.0. Throughout the building, two manifestations of the ontology are maintained: one (acquisition oriented) specified in the semi-formal language associated to OntoSpec (the methodology defined by MIS); the second (oriented towards inferences) specified in a dialect of OWL (presumably OWL-DL).

latory, 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.

Inferences of trace analysis and publication: This task will develop and/or adapt a tool for semantic analysis of the traces. This tool will exploit the ontology to carry out inferences on grid traces (especially to detect inconsistencies), and also to improve the information retrieval tools of the GO gateway (SA.GO.1).
As a result of hardware and software failures, and also of conceptual ambiguities, monitoring outputs (traces) may be erroneous. This may seriously limit the potential of trace usage, as scientists not experts in gLite (or other EGI –deployed middleware) would not be able to properly manage the unavoidable inconsistencies, missing data, or outliers, in the traces. A considerable expertise in this area has already been acquired in the EGEE project. However, this expertise is encapsulated in scripts or programs (such as those used in gStat), which make it inaccessible to the scientific community. Moreover, the lack of automatic inference hampers the error discovery process.
A set of tools is thus needed to manage and efficiently access the ontology, and to carry out inferences on traces. Inferences will be carried out on semantic representations of traces, so a tool is required to built such semantic representations from log files. Several semantic engines exist which are currently used in numerous projects. The choice of tools will be undertaken at the beginning of the second year of the project.
Scalability is the major challenge for this activity. Assessing the volume of data depends a) on the efficiency of the "lossless compression" achieved by SA.GO.1 and b) the number of concepts to be taken into account, and also of the database technologies that will be chosen. Large tests of chosen tools in trace analysis will allow to validate the ontology and to improve it. The tools will be extended to link them to the publication tool of the GO gateway (SA.GO.1).
The expected result is mainly automating the process of discovering plain errors, or suspicious data. If expert knowledge can be secured from EMI and EGI-proper, adequate remediation (i.e. correcting the data, or at least tagging the erroneaous data with a probable explanation, warning etc.) would be proposed and integrated in the semantic engines. One important application area for this activity is to provide explicit and exploitable foundations for reliable operation- or user- oriented metrics.

Partner contributions: LRI leads the WP. UNIPM leads task JRA.GO.1. Task JRA.GO.1.1 will be performed by LRI (fault diagnosis), UNIPM (Complex Networks) and CU (Autonomics). Logica, is in charge of JRA.GO.1.2.
MIS leads task JRA.GO.2, with the participation of LPC

Complexity Science

Photon Science

Humanities

Deliverables (brief description and month of delivery)

Table 25: Summary of staff effort

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

Table 26: List of milestones

Milestone number	Milestone name	Work package(s) involved	Expected date[41]	Means of verification[42]

1.6.4. Graphical presentation of component interdependencies

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

1.6.5. Significant risks and associated contingency plans

Table 27: Risks for JRA1

Risk WP1	Impact	Occurrence Probability	Mitigation
Poor collection of	CS SSC Community	Low	Stretch out to the CS

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

[42] 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.

data sets to be incoorporated into the CS SSC Scientific Database	of users		Community to provide us with a concrete set of data ranging from all aspects of Complexity Science fields of study. The community will be benefited the most with a large collection of data to be studied.

Table 28: Risks for JRA2

Risk WP1	**Impact**	**Occurrence Probability**	**Mitigation**

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 1 provides an overview of the ROSCOE management structure.

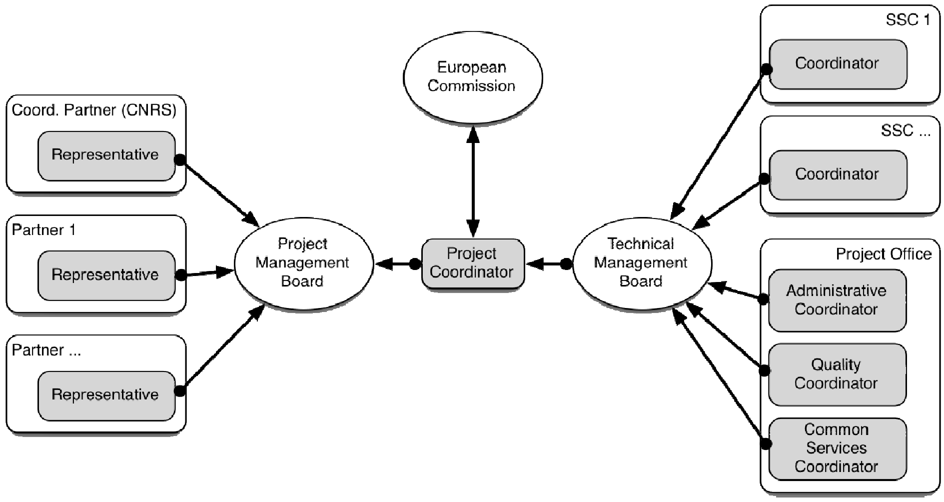


Figure 1: Management Overview Diagram

At the highest level, the Project Management Board (PMB), made up of one representative from each signatory, 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 regularly reports to the PMB and is responsible for the daily operation of the project. The Project Coordinator 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 common service activity, of each SSC, and the Administrative Coordinator. 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 Administrative Coordinator heads the Project Office. The Project Office 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 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, that the participants execute the defined work plans, and that the project ultimately achieves its goals. The Project Coordinator interacts with the following bodies:

- 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.
- European Commission: The Project Coordinator will be the sole liaison with the European Commission for the project.

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.

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.
- 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. Project Management Board

The Project Management Board will consist of one representative from each Signatory to the Grant Agreement. Each representative will have a vote weighted by the number of participating institutes he/she represents. The Chair of the PMB will be chosen from among the members of the PMB. The Project Coordinator, Administrative Coordinator, or member of the Project Office may not serve as a representative. The Collaboration 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 six months. 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.

The PMB must oversee and make decisions related to issues such as:

- 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.

2.1.1.6. Technical Management Board

The Technical Management Board (TMB) will consist of the Project Coordinator, Administrative Coordinator, representatives from Common Service Activities, and representa-

tives from the SSCs. 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 every two weeks at a fixed time and day of the week. Meetings can be rescheduled with adequate notice to the participants. 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 either the Project Coordinator will decide if it does not involve a topic under the auspices of the PMB or will be forwarded to the PMB if it is.
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.

2.1.1.7. SSC Coordinator

Each SSC 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. If he represents his activity in an outside project or organization, he is responsible for reporting information to members of his activity and significant advances or issues to the TMB. The coordinator is responsible for coordinating interactions with other activities within the project and contributing to common services.
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 SSCs as independent, stable, and sustainable entities, the SSCs will in many cases interact directly with structures within EGI. The SSCs 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.

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
- SAFE: SSCs for Astronomy, Astrophysics, Earth Science, and Fusion

- "APS": Application Porting Support project
- EMI: European Middleware Initiative
- StratusLab: Combining grid and cloud technologies

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

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)

Provide a description of your consortium (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.

2.4. 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)

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.

Deployment of e-Infrastructures in research communities in order to enable multi-disciplinary collaboration and address their specific needs.

HEP: Although the primary goal of this SSC is to support the High Energy Physics experiments at CERN and elsewhere, a number of the tools developed have already been adopted by a range of disciplines – including others in this "SSC cluster" as well as those beyond (UN initiatives and EU-funded projects such as EnviroGRIDS and PARTNER – a hadron-therapy project). Such inter-disciplinary collaboration is considered of great importance both to all partners and to the community as a whole and ways of expanding this through the Heavy User Community of EGI and beyond will be explored. This is true both "vertically" (i.e. within a given SSC) as well as "horizontally" – i.e. across distinct SSCs. (e.g. collaboration with Fusion (Ganga), Life Science (Ganga + GEANT4).

LS: The SSC will leverage the work of several European projects (EGEE, EMBRACE, E-NMR, Health-e-Child,…) to provide services for accessing the resources of the National Grid Initiatives federated in the European Grid Initiative to the research communities in molecular biology, medical imaging, drug discovery and next generation sequencing. Through the involvement of key European institutes or associations, it will foster the adoption of grids and the use of EGI resources by the Research Infrastructures that will map the field of life sciences in the coming years.

CCMST: Chemistry never was and cannot be considered as an isolated branch of science. It influences a wide range of scientific disciplines such as physics, astrophysics, biology, medicine, pharmacy, climate or earth science, giving a logic and exhaustive explanation of phenomena and processes running in biological cells, living organisms, nature and universe. Keeping this in mind CCMST SSC is highly interested in e-infrastructure supporting many scientific disciplines and enabling cooperation with them. A majority of software packages deployed to grid architecture can be of use by any scientific discipline needing it.

GO: Scientific communities worldwide have set up massive grids that manage several tens of thousands of CPU's and several PetaBytes of storage space. The control, and maintenance of these complex systems remain a significant operational challenge. Application developers need synthetic characterizations of the grid activity and the grid applications for predicting and optimizing application performance. Grid models are required for dimensioning, capacity planning, or predicting the improvements consecutive to changes in grid configuration or middleware.
The grid infrastructure consists of a variety of hardware and software components, which are, in their own right, complex systems. Experimental data on the grid activity in

real working conditions and advances in modelling method are necessary to discover adequate empirical models of the grid. On the other hand, fundamental grid research needs the experimental data created by the collective behaviour of the first grid users communities, as an input.

CS: Interdisciplinary science has drawn an increasing amount of attention over the last two decades. Particularly with the changes in the rate and nature of data production of economic, social, climate, seismological, physiological and biological sciences, new kinds of systems emerged. Studies on such systems evolved from early on around their common properties, their complex structure and the underlying dynamics. From such perspectives, in different research contexts complex systems are defined on the basis of their different attributes.
The aim of this proposal is the establishment of a Specialized Support Centre focusing on the ICT needs of the European research community dealing with Complexity Science, an emerging field of interdisciplinary scientific interest.

PS: Experiments in photon sciences are used by a whole variety of different communities coming from material science, chemistry, biology, medicine etc. The primary goal of this SSC is to support this heterogeneous set of user groups at the next generation of light sources in Europe, which are part of the ESFRI roadmap, in particular the European XFEL, the EuroFEL and the ESRF-upgrade. Therefore in particular the middleware and the tools developed in EGEE and elsewhere will be adopted and integrated. This effort highly relies on inter-disciplinary collaboration, which is considered as of great importance both to all partners and to the community as a whole and ways of expanding this through the Heavy User Community of EGI and beyond will be explored. This is true both "vertically" (i.e. within a given SSC) as well as "horizontally" – i.e. across distinct SSCs, in particular the HEP SSC and will extend commonalities across facilities as well as scientific communities.

H: The potential user community for an SSC 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 from 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 an e-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.

The Humanities SSC goals are therefore the following:

- To offer support services to the humanities community for accessing EGI resources
- To work in particular with the two Humanities ESFRI projects CLARIN and DARIAH on providing them with EGI services and

- To develop user-friendly ways of accessing EGI services. In the first instance we would like to concentrate on access to storage resources.
- Standard work in the context of EGI and OGF
- To work with National Grid Initiatives to make sure Humanities specific issues are addressed
- To consult with the social sciences communities to work together on common objectives regarding a common national infrastructure
- To work together with various international partners to establish a programme of collaboration for a Humanities e-Infrastructure
- To work towards establishing data-driven humanities needs

Deployment of end-to-end e-infrastructure services and tools, including associated interfaces and software components, in support of virtual organisations in order to integrate and increase their research capacities.

HEP: This is essentially the *raison d'être* of the proposed support centre. In particular, one of its main goals is to support the High Energy Physics and related communities at this critical phase of LHC startup and exploitation. This involves approximately 10,000 researchers worldwide who need to access and analyze data 24x7 using worldwide federated grid resources. The service and user support to this community – enabling them to maximize the scientific and discovery potential of the LHC machine and the detectors that will take data at it – is a fundamental goal.

LS: The Life Sciences SSC will operate and improve the services of the Biomed Virtual Organization which was the EGEE catch-all VO for life sciences, which accounted for more than 90% of the scientific production in life sciences on this infrastructure and already provided access to more than 20.000 CPUs in Europe and beyond in 2009. The goal of the SSC is to improve the services offered by the VO in the following ways:

- Extend resources accessible to the VO users beyond those operated by gLite to resources operated by KnowARC and UNICORE..
- Develop tools allowing users of the scientific gateways to access resources operated by the different middlewares supported by UMD (gLite, KnowArc, UNICORE)
- Improve the monitoring of the resources using existing tools like the Dashboard
- Maintain, promote and enrich a catalogue of grid services for life sciences and healthcare. The catalogue will be developed within the framework of the EGI_PROPER proposal.

CCMST: Encouraging user to run their software in grid environment I actually the basic motivation for instituting a SSC in the Field of CCMST. Looking at the statistics concerning EGEE Grid usage only about 8 percent of the CPU time is related to CCMST discipline whilst typical HPC centre utilisation oscillates around 70 percent. Taking in to account the rapid evolution of distributed computer technologies (which remain hard to master for researchers) CCMST will work on lowering the barriers by implementing scientific software as a web services. A standardisation of CCMST data and software codes is another topic of high importance for the community. Adopting and disseminating the use of shared models for Quantum Chemistry (QC) is crucial to support collaborative work and interoperability among the community.

GO: The Grid Observatory SSC includes the collection and publication of grid activity traces, the construction of ontology of the domain knowledge, and the exploration of new grid models and control methodologies. EGI will serve both as the primary data

source, and as the service infrastructure. As a data source, it is assumed that the extensive monitoring facilities already deployed in gLite and by the scientific communities with be continued and expanded in the future EGI and UMD developments. As the service level, EGI will provide the computational infrastructure required for research and development targeting intelligent data management and analysis.
The GO SSC will build on the previous activity of the GO cluster in EGEE-III, which has already successfully realized a grid trace portal and fostered national and international collaborations with the Computer Science community.
The GO SSC will empower users through its gateway, including advances towards a semantic portal, with the facility to access and retrieve data that would be otherwise inaccessible, and
The added value of the GO SSC is in the integration of its production goal – make available comprehensive and usable grid traces – and the long-term scientific goal of acquiring better knowledge and control of the grid as a complex system.

CS: The overall scope of the CS SSC is to strengthen the multi disciplinary collaboration of the European research community of Complexity Science through the creation and deployment of services and tools which will be build mainly upon the EGI Infrastructure with the aim of both facilitating new research groups joining the community and increasing the research capacities and capabilities.
Specifically we plan to,

- Further increase the usage of Grid technology by porting more applications and by introducing more users through seminars, workshops and personal contact.
- Develop and deploy a Web portal for the registration of new users and the support of existing ones.
- Provide a common toolset containing frequently used algorithms such as the Network Analysis, the Detrended Fluctuations Analysis, the Wavelet and the multifractal DFA algorithms.
- Build parallel (based on MPI) and hybrid (based on MPI and OpenMP) versions of basic algorithms that will help us optimize the usage of the EGI DCI.
- Build a data repository containing climate, physiological and stock market exchange data specific to the SSC needs deploying the AMGA metadata service

PS: Research today at the light sources is an international effort, both in terms of collaboration as well as in terms of geographical distribution of the researchers and the experiments. A light source like DORIS or ESRF supports typically about 5,000 – 10,000 visiting researchers per year. It can be expected to up to half of the users will use the facilities remotely in the near future.
The European XFEL is an international association with member states distributed world wide. The estimated yearly 10PByte of raw scientific data need to be made available to individual research groups and possibly also national data centre hosting replica of the relevant data.
The European e-Infrastructure both, in terms of connectivity and in terms of available software components through EGEE is therefore essential for future collaborative work of the scientists and of the labs itself.

H: This SSC will have a particular focus on working towards data-driven and data rich Humanities, in particular the provision of large and geographically dispersed data sets to the research community. It has been established in the various national Digital Humanities e-Infrastructure initiatives that 'processing' is at the moment a secondary need

in the digital humanities. Most important to our community is the development of a pan-European data infrastructure that will be able to accommodate the large amounts of humanities research data that is currently. The SSC will therefore work on the integration of repository technologies into the European infrastructure towards an open repository network.
Next to these repository services, the SSC will work towards integration services with the EGI (such as portals and service bridges) and services that will allow the consumption of large data sets by the research community, especially data and text mining services. We wish to scope a common agenda here, which would not only include the Humanities SSC.
The SSC will identify 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 and the management and storage of very large collections of image and multimedia objects. Support and outreach activities will be developed around these themes.

Building user-configured virtual research facilities/test-beds by coalition of existing resources (e.g. sensors, instruments, networks, and computers) from diverse facilities, in order to augment the capacities of research communities for real world observation and experimentation.

HEP: In the context of WLCG, this is performed via the Memorandum of Understanding (MoU) that brings together CERN, the experiments and the resource providers around the world with day-to-day supervision and decision making performed by a WLCG Management Board consisting of members from all the WLCG project actors. This involves several distinct grids – currently EGEE, Open Science Grid (OSG) and NorduGrid – and numerous funding agencies (the signatories of the MoU). Several bodies exist to monitor that pledges and commitments are met and to review requests for additional resources and the schedule for acquiring and deploying them (Computing Resource Review board and Computing Resource Scrutiny group).

LS: The Life Sciences SSC will provide support to the Research Infrastructures in the field of life sciences wishing to develop testbed and virtual research facilities. Its role is not to develop its own facilities but rather to help ESFRIs to develop their own Distributed Computing Infrastructures and fully take advantage of the services offered by EGI.

CCMST: Libraries of programs and suites of computer codes are probably one of the most valuable assets of the CCMST community. Their stable versions are used frequently in a black box fashion and in this case either the author(s) or the user(s) use to spend significant efforts to improve their user friendliness by providing appropriate Application Program Interfaces (API). Implementation of scientific software as web services not only provides simplified access to them for users but also providers/vendors who by saving large amount of time can implement more advanced futures of given code.

GO: The aims of the SSC are as follows.

- Provide production-quality services for the Computer researchers and engineering in Europe and beyond, through data collection, publication and descriptive analysis.
- Foster basic research collaboration through scientific networking.

- Build bridges between the operational requirements emerging from the new EGI model of operations and the computer science community, encouraging and facilitating the experimentation of new innovative ideas contributing to grid middleware improvement in reliability, stability and performance.

More specific aims of the proposed GO SSC are listed below, with the related tasks in the work plan indicated

- Develop of a full-fledged acquisition process integrated within gLite (SA.GO.1).
- Provide and develop on-line analysis services running on the EGI grid, as contributions to on one hand scientific research (SA.GO.1), and to the other hand a sustainable operation model (JRA.GO.1).
- Provide a network of expertise in the interpretation of production grid behavioural models (JRA.GO.1).
- Contribute to the definition of a grid ontology, which will be the basis for interoperability with other data repositories (GWA, …) and interaction with other computational production models (clouds, grid overlay networks, desktop grids) (JRA.GO.2).
- Foster the creation of a COST project as a support for the basic research networking (NA.GO.1).

Define and enact processes for the specification of interpretation and control challenges, and the evaluation of the proposed solutions (NA.GO.1)

CS: Initially we plan to integrate part of a 100 core cluster with 3TB of storage, that will provide an initial infrastructure upon which the CS community will be able to build, test, deploy and run CS specific algorithms.

PS: The aforementioned PanData project aims to define and extend commonalities among the neutron and light sources throughout Europe, for example by defining policy frameworks, repositories and standards for Photon and Neutron Sciences. Taking up these efforts in the grid context will the basis for a seamless integration of distributed data management and analysis. Together with further development of grid based remote operation of real as well as virtual instruments, this would assemble a wide spectrum of scientific experiments and instruments into a virtual research centre.

H: N/A

Addressing human, social and economic factors influencing the creation of sustainable virtual research communities as well as the take up/maintenance of e-Infrastructure services by communities.

HEP: This SSC and WLCG will be key drivers behind the interoperation of the gLite, ARC (NorduGrid) and OSG middleware stacks and related services. WLCG also has partners in Latin America and Asia Pacific.
Furthermore, one of the key challenges that faces fundamental research, such as High Energy Physics, is to allow researchers from around the world to fully participate in their experiments – which may be physically located on the other side of the world – whilst still playing a key role in the scientific and cultural life of the University or Research Institute for which they work. Realising that education is key to the long-term success of economies and societies as a whole, ways whereby this ambitious goal can be achieved are of great importance. One of the significant advantages of grid computing as compared to previous less integrated types of remote working is the realisation of worldwide virtual research communities that can consist of thousands of researchers at hundreds of institutes where researchers are not impeded by distance and can play equal roles regardless of location. This ability has enabled LHC experiment member

countries to invest in local and regional computing infrastructures at national laboratories and universities, with ten first level and over two hundred second level sites, confident that this infrastructure can be used. Success of this e-infrastructure project will reinforce this confidence leading to increased investment. Socio-economic benefits will include local employment and the continued development of local and regional centres of excellence. It has also strongly contributed to the success of worldwide distributed collaboration on grid services, whereby a highly functional data processing and analysis system can be run despite the challenges of multiple management domains, time zones, local priorities and other such challenges.

LS: the Life Sciences SSC will address the fractioning of the virtual research communities in Europe. It seems today that each middleware stack has its own user community in life sciences. The Life Science SSC will take the following initiatives to break the walls between these communities:

- it will promote the exchanges between the users of the different middleware stacks (gLite, KnowARC, UNICORE) through its support to joint events like the HealthGrid conference
- it will develop tools for transparent access to resources operated by different middleware stacks

CCMST: N/A

GO: The usage of the SSC data and services is, by definition of the SSC, open to the whole scientific community. Beyond that, the GO SSC will act as a catalyst for developing synergies at the European level and beyond, between scientific communities that have had so far limited opportunities to interact. A special emphasis will be put on the cross-fertilization of autonomic computing on one hand and grid research and engineering on the other hand. It must be noted that the Autonomic Computing community is mostly US based, with the NSF Centre for Autonomic Computing, and a strong involvement of industry (notably IBM).
The GO SSC 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, it is extremely important that the GO SSC can evolve towards a permanent structure and define a sustainable financing model.
On the other hand, 1) the activity is much younger than all other scientific SSC (the corresponding EGEE cluster has been created only within EGEE-III), and 2) the Computer Science community has no international body comparable to CERN, ESA, or even the large biomedical collaborations. Overall, the SSC is still in it ramp-up phase; thus it requires initial development funding, and has to invent a permanent structure and a governance model in the course of its existence.
The GO SSC must have support from, on one hand stakeholders involved in actual production, such as some NGIs and EGI, and on the other hand research institutes not presently involved in EGI, but prospective users of the GO data and services. The French NGI will provide the bulk hardware resource and participate in the operation tasks. The other NGIs are expected to contribute to the acquisition task (SA.GO.1) under the general operation scheme, thus will not have to provide dedicated human resources. Interpretation from operation experts is a requirement for JRA.GO.1 and NA.GO.1, and will be bootstrapped by the constitution of a network of experts. The GO SSC requires interaction with EGI and UMD, 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 practical plans are as follows.

- The French NGI ensures the infrastructure requirements, including non-dedicated operation costs, together with EGI, (e.g. providing the volatile and long-term storage). The national production grid has been recognized as a TGIR (Très Grande Infrastructure de Recherche) by the French ministry of research.
- A mix of permanent and project-based funding will provide human resources.
    - Permanent scientific personnel will provide scientific and operational steering.
    - The major source of project-based funding for this period will be, on one hand the EC (through this call, and possibly a COST programme), on the other hand the national or regional funding schemes (e.g. the French ANR, the UK e-science programme).
    - Industry partners are now involved. This will facilitate applying for national or international funding scheme targeting industry R&D; in a more ambitious scheme (such as the evolution of the TOP500), the GO gateway would become sufficiently popular to attract company sponsoring.

It is of major importance to define the perimeter of scientific consolidation for an SSC. For instance, in France the combination of the administrative system, and the FP7 administrative rules, preclude any possibility to consolidate 1) FP7 funding, and 2) national project-based funding, at the accounting level. In practice, eligible effort can come only from permanent staff, or (with extreme difficulty) temporary staff hired on recurrent funding. Thus, the EC should precise the guidelines for evaluating the evolution towards sustainability, which cannot be based solely on accounting reports.

CS: The SSC user community will be the primary driving force in the course of the SSC as most of the work invested will be focused on serving its needs and on the creation and implementation of a sustainability plan. Thus, the CS SSC will pursue to integrate computational and storage resources with the wider European Grid Infrastructure. Through the CS SSC a series of services and tools that will help users benefit from this distributed infrastructure will be designed and implemented. The sustainability of the virtual research community will heavily depend upon the usability of such resources and services and thus their thorough documentation as well as their implementation should meet the corresponding needs.

PS: A more efficient use of available large scale facilities like synchrotrons and FEL's is certainly a key economic factor, largely dependent on the definition and creation of commonalities among a very heterogeneous assembly of scientific communities. Sharing resources and know how across facilities and a huge number of user communities and a more efficient use of the facilities trough remote access and operation of the experiments will enhance efficacy significantly and might help accelerate multi-disciplinary research in areas, which have seemingly nothing in common, thereby utilizing the natural advances of grid computing in contrast to a poorly integrated experimental framework.

H: (N/A) It is too early too offer this kind of commitment, however we will undertake scoping work to ensure that we have a full understanding of the humanities community's needs in this area when such facilities do become applicable.
Addressing human, social and economic factors influencing the creation of sustainable virtual research communities has been a focus in the work of the various national e-Humanities initiatives. The key relevant finding of this work is that supporting the digi-

tal humanities in using e-infrastructure extends far beyond the kind of generic helpdesk support that they are typically used to dealing with. It requires a mixture of technical, domain and infrastructure expertise. In the UK, for instance, the Humanities have been involved in various JISC funded studies to investigate sustained uptake of e-Infrastructures. We are currently working closely with colleagues from the Oxford e-Research Centre on a study regarding Virtual Research Environment/Collaboratory use. This is particularly an area where we would like to seek collaboration with the social sciences communities such as UK NCeSS who have a long-standing expertise in this domain. For example, the UK National e-Infrastructure for Social Simulation (NeISS) is a coordinated initiative to develop and apply simulative methods in the areas of social and political science: the humanities SSC will be well placed to collaborate with NeISS to ensure mutual benefit to Humanities disciplines (such as history and archaeology) where emergent simulation and modelling applications are proving their capacity to add value.
The Digital Humanities themselves have developed the idea of a research life cycle and research primitives, i.e. those research functions that are repeated in most research processes in the Humanities. In the context of the SSC, we would like to intensify this work on scholarly primitives, which is also one of the main research topics for DARIAH. Digital humanities define themselves with respect to "methodological commons of techniques derived largely from and applicable across the other disciplines". The SSC would like to demonstrate how these would correspond to Grid infrastructures.

Integrating regional e-Infrastructures and linking them to provide access to resources on a European or global scale.

HEP: The Worldwide LHC Computing Grid (WLCG) is very much a federated grid and builds on today's EGEE infrastructure, together with grid resources provided through OSG in the US, NorduGrid in the Nordic countries as well as partners in other regions of the Americas and throughout the Asia-Pacific region. This is essential given the fully global nature of High Energy Physics and will be an important component of the proposed work.

LS: the Life Sciences SSC will build upon the success of the WISDOM initiative which received support from several regional e-infrastructures (AuverGrid, COMETA, EELA, EUMEDGrid, EUChinaGrid, SeeGrid, TWGrid) including e-infrastructures in the United States (Digital Ribbon, OSG). Several partners of the SSC (CERM, CNRS, HealthGrid, IBRB, INFN, UPV) will play a coordination role with projects such as AuverGrid, COMETA, EUAsiaGrid, EUIndiaGrid, EUMED-p, SEEGrid. The SSC will be able to provide support to the biomedical activities of the regional e-infrastructures while they will be able to contribute resources to global initiatives.

CCMST: N/A

GO: The objective is not directly applicable to the GO SSC. However, it is important to coordinate the GO action with the national Computer Science research infrastructures such as G5K in France, or DAS in the Netherlands, with due respect to their differentiated goals. Considering computer science users, the proposed SSC expects to receive Letters of Support from various high-level research groups, institutes and projects. As an example, discussions have started with the NSF Centre for Autonomic Computing, the French experimental infrastructure for Computer Science Aladdin/G5K, the Core Grid ERCIM Working group, and the RESERVOIR project.

CS: The Project partners and the NGIs supporting them will provide the computational and storage resources required for the scientific missions of the SSC. Additionally, the Project partners will provide the scientific data and the collaborative resources needed in the context of the SSC Project. In addition to that the Project partners will provide mailing list services and development collaborative tools to the wider Complexity Science community.

PS: Light sources on this scale are used as international research infrastructures and are key elements on the ESFRI roadmap. As a sort of spin off the developed user interfaces and technologies will be offered to the national resources and will be brought into the regions.

H: There are currently at least 3 functioning national Humanities infrastructures in Europe. In the UK, the centralised Arts and Humanities Data Service work is continued in the more distributed Network of Centres, which takes better into account the need of institutional involvement in the area. In Germany, TextGrid is working on a national DGrid level to provide services to German e-Humanities researchers. In the Netherlands, the DANS provide access to Humanities resources for all researchers. Next to these, there is commitment by the French government to set up an e-Infrastructure for Humanities in the context of the ADONIS project. Work on the actual implementation has just begun. In Greece, too, there are definite steps towards brining together humanities resources and researchers. All these countries and project are part of the ESFRI DARIAH consortium. The SSC will be the point of contact for these initiatives for EGI.
Active steps are already under way to connect these e-Infrastructure using lightweight service-oriented architectures. This will be part of DARIAH technical work. The SSC will ensure that these attempts are done according to the standards for EGI.

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

Information on the dissemination plan (HEP).

A key activity foreseen for this SSC 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 which 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 op-

eration. 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 SSC 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.

3.2.2. Exploitation Plan

Information on the exploitation/sustainability plan (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 SSC 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.

3.2.3. Intellectual Property Management

Information on any issues related to the intellectual property management (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.

3.3. Contribution to socio-economic impacts

Describe the socio-economic impacts of the project.

Socio-economic impacts expected from ROSCOE for your SSC (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 multinational 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 which 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.

Please also provide any comments on common tasks within the defined activities (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.

4. Ethical Issues

Table 29: 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: Letters of Support

Possible organizations from which to obtain letters of support:

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