



<b>Subtask ID</b>	GO1	<b>Start Month</b>	1	<b>End Month</b>	24
<b>Subtask Name</b>	Cluster Management and focusing collaboration				
LRI	<b>C. Germain</b>	2			
<b>Goals</b>					
<p>This subtask has in charge the coordination of the whole GO cluster.</p> <ul style="list-style-type: none"> <li>▪ Manage interactions with the EGEE/NA4 and with the EGEE Project Office, including administrative aspects which are not familiar to some participants.</li> <li>▪ Monitor and coordinate activities of all other subtasks and the progress for each of them.</li> <li>▪ Set up tools and services to foster the collaborative work within the cluster (mailing lists, wiki, documentation repositories, etc.). A specific issue is the relationship between publicly accessible collaborative tools associated with the web portal, and the tools internal to the project.</li> <li>▪ Organize cluster meetings, both face-to-face and remote (through collaboration tools).</li> <li>▪ Produce reports and other documentation requested by EGEE/NA4 SC and by the Project Office.</li> </ul> <p>The other goal is focusing collaboration with Computer Science:</p> <ul style="list-style-type: none"> <li>• Publicizing the GO portal on appropriate supports (newletters, web sites), Seeking for opportunities of presenting the GO in major conferences (eg SC)</li> <li>• Collaborate with related projects (eg answer to RENCI requests)</li> <li>• Catalyze new projects, especially from national or regional institutions, in the NGI perspective.</li> <li>• Interacting with the initiatives in the same area (eg Grid Workload Archive). .</li> </ul> <p>This activity would be the place to integrate standard related activities. The cluster cannot provide it on its own resources, and will interact with NA4 management on this issue.</p>					
<b>Benefits</b>					
Administration					
<p>Given the small size of the cluster (albeit distributed in 3 countries), and their broadly different cultures, this task is both relatively light (in PM) and critical to actually integrate the new participants from Computer Science into the EGEE culture.</p> <p>Focusing collaboration:</p> <p>This is an expected added value of the cluster. While it success depends on the other subtasks, a proactive management is critical to reap the full benefits of these activities.</p>					
<b>Milestones and Metrics to gauge Progress</b>					
	<b>Milestone Description</b>	<b>Metrics</b>			
	Reports each 6 months	<i>administration</i> Number of cluster participants attending the EGEE events (All hands, conference, UF, internal cluster meeting)			
		<i>dissemination</i> : associated research			

		projects, and contacts
--	--	------------------------

<b>Subtask ID</b>	GO2	<b>Start Month</b>	1	<b>End Month</b>	6
<b>Subtask Name</b>	Data partial collection				
IC	<b>D. Colling</b>	1	LRI	A. Cady	3
				Germain	1
				J; Perez	1
				X. Zhang	1
<b>Goals</b>					
<p>The goal is to be able to collect example but significant data soon after the creation of the GO, and to publish them. The collection will include:</p> <ul style="list-style-type: none"> <li>• Raw traces: L&amp;B; internal logs of WMS, SE and MAUI; differential log of the BDII – provided by the GRIF/LAL site</li> <li>▪ Summary data: the Real Time Monitor (GridPP), which are the IC contribution.</li> </ul> <p>The bulk of the work will be performed in the first 6 months, the remaining being only exploitation.</p>					
<b>Benefits</b>					
<p>The early release of data will first bootstrap internal and external collaborations, and second provide a testbed for some of the issues raised in the comprehensive collection subtask..</p>					
<b>Milestones and Metrics to gauge Progress</b>					
<b>M-ID</b>	<b>Milestone Description</b>			<b>Metrics</b>	
<b>M-GO2-1</b>	Prototype at month 3			Number of type of traces collected	
<b>M-GO2-2</b>	End-to end production quality system at month 6				

<b>Subtask ID</b>	GO3	<b>Start Month</b>	6	<b>End Month</b>	24	
<b>Subtask Name</b>	Comprehensive collection					
LRI	<b>Germain</b>	3P M		IC	D. Colling	2p m
	To be hired	18 P M				

### Goals

This activity is oriented toward long-term, systematic collection of data at the full egee scale.

The first 6 months will target the collection of *raw* data (raw meaning the output of monitoring tools in their native format). The acquisition process must be carefully designed so as not to interfere with production. Given the present level of stress on some middleware services, the acquisition process must be carefully validated..The following issues will be thus be addressed:

- Identification and evaluation of data sources, assessment of their availability.
- Lightweight sustainable acquisition protocols

We do not expect major difficulties on these two points. **The major issues is in the deployment process.**

- Enacting the deployment. While the acquisition protocols can be validated in collaboration with the LAL site, the actual deployment on other sites does not depend on the GO cluster, but on the willingness of other EGEE activities (SA1, JRA3).
- Legal issues. It is likely that not all raw data can be exported from national sources. In any case, anonymization is a major requirement.

We will build on the IC experience (with the RTM). However, the whole process must be discussed (eg through the TMB) from the beginning.

The second year will be devoted to the design of efficient conversion from the operational logs towards higher level concepts. These concepts should be defined from two sources, namely the analysis task, and the external scientific users. They will finally drive the organization of a database. The existing conversion tools, in particular the Job Provenance software, will be considered.

### Benefits

.

### Milestones and Metrics to gauge Progress

M-ID	Milestone Description	Metrics
<b>M-GO3-1</b>	Report on the planned deployment process, month 9	Number of participating sites or ROCs .
<b>M-GO3-2</b>	Report on technical tools, and proposed architecture month 12	
<b>M-GO3-3</b>	Prototype implementations month 18, and final report month 24	

<b>Subtask ID</b>	G04	<b>Start Month</b>	1	<b>End Month</b>	24
<b>Subtask Name</b>	Publication				
LRI	Germain 2 PM				
<b>Goals</b>					
<p>The goal of the portal is to provide <i>public access</i> to traces of EGEE activity. Public access means that access grant will be controlled only to avoid malicious clogging of the site. Its creation will be publicized (newsletters, RSS).</p> <p>The first version of the portal will publish the partial collections. The next versions will include more raw data (eg extension of the GRIF/LAL collections to other sites, better use of gridPP software and network), and possibly (depending on external funding) the creation of an analysis portal, proposing analysis tools on EGEE, with public access à la GPSA. As some code might be MATLAB, there are interesting connection with the ongoing MATLAB deployment on EGEE. .</p> <p>In relation with the goal of fostering collaborations, collaborative tools will be part of the portal since the beginning.</p> <p>The Web Site will be created and maintained by an external contractor (health grid). The datasets (trace files) will be stored at GRIF/LAL, inside the EGEE storage area.</p> <p>Access to full datasets must be controlled, but should not be certificate-based. It will be based on a simple registration form.</p> <p>A major change will integrate semantically organized data, which will require a significant re-design of the portal, including presentation and access technologies.</p>					
<b>Benefits</b>					
<p>The portal is both a unique scientific infrastructure, and a major communication medium. Offering professional quality for this highly visible activity is the reason for outsourcing.</p>					
<b>Milestones and Metrics to gauge Progress</b>					
<b>M-ID</b>	<b>Milestone Description</b>			<b>Metrics</b>	
<b>M-GO4-1</b>	First release of the portal (V 1) at EGEE 08 Conference			Number of access Number of registered users	
<b>M-GO4-2</b>	Improved portal at months 12 and 18				
<b>M-GO4-3</b>	Access to semantically organized data at month 24				

<b>Subtask ID</b>	GO5	<b>Start Month</b>	1	<b>End Month</b>	24
<b>Subtask Name</b>	Analysis				
Unipm	<b>L. Saitta</b>	6			
	A.Giordana	6			
	I. Lovro	6			
	To be hired	18			
IC	D. Colling	3			
LRI	C. Germain	4			
	M. Sebag	6			
	J. Perez	11			
	X. Zhang	11			

**Goals**

This task aims at modeling the dynamics of both the grid and the user community. It will participate in the operations activity and technological evolution of gLite and EGEE. The areas concerned are dimensioning, capacity planning, and a-posteriori evaluations of the services performance. The task should also contribute to the design of self-regulation and maintenance functionalities such as real-time fault diagnosis and anticipation, and possibly intrusion detection.

The currently envisioned areas are

- Fault detection and diagnosis
- Time-series models
- Structured hidden markov models
- Dynamic Bayesian Networks

**Benefits**

The general interaction with computer science

The task is willing to propose realistic methods to answer real operational issues. However, because the GO is the newest and smallest cluster, a significant question is how to become aware of these real operational issues. The risk is either to get none, or unrealistic ones. Another serious issue lies in the following fact: answers will be considered for integration only if fully implemented; this is a costly process, with no reward if the implementation finally does not make its way into the egee middleware stack.

To tackle these two issues, a formalized reporting process will be proposed to SA1 and JRA1, in order to bootstrap the interaction (issue 1), and to get a qualitative assessment of the answers (issue 2)..

The BDII scalability has been pointed to the GO as a major question. Given the limited manpower and the need for extensive simulation apparatus, this issue will be explored only if external support can be secured.

<b>Milestones and Metrics to gauge Progress</b>		
<b>M-ID</b>	<b>Milestone Description</b>	<b>Metrics</b>
<b>M-GO5-1</b>	Status on the report and evaluation process at month 3	Number of publications Evaluation of the interaction with SA1 and JRA1.
<b>M-GO5-2</b>	Report on models at month 6	
<b>M-GO5-3</b>	Report on data schema at month 12 Report on models at month 12	
<b>M-GO5-4</b>	Report on models at month 18	
<b>M-GO5-5</b>	Final report	