

# PROPOSAL FOR A GRID OBSERVATORY SPECIALIZED SUPPORT CENTER – GOSSC

DRAFT 29/06/09

## 1 Introduction

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 modeling 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.

The goal of the Grid Observatory SSC is to contribute to an *experimental theory* of large grid systems by setting up a *data repository*, enabling the development of *behavioral grid models* and supporting the deployment of *software experimental facilities*. The Grid Observatory SSC includes the collection and publication of grid activity traces, the construction of an ontology of the domain knowledge, and the exploration of new grid models and control methodologies. 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.

Moreover, the GO SSC will act as a catalyst for developing synergies at the European level between scientific communities that have had so far limited opportunities to interact, with a special emphasis on the cross-fertilization of autonomic computing on one hand and grid research and engineering on the other hand.

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

An important asset of the GO is to provide the multidisciplinary skills and institutional positions required to make truly significant contributions towards this goal. The partnership includes 1) high-level scientific expertise in both Distributed Systems, Machine Learning, and Ontological Analysis 2) support from the EGI-related projects both in operations and in middleware development.

## 2 User Community

### 2.1 Scientific and engineering Scope

The scientific area covered by the SSC are the segments of Computer Science and engineering related to complex systems on one hand, and large scale distributed systems on the other hand. <2be developed>.

Industry is active in the area of Autonomic Computing. In order to ensure a production-quality result on the data collection and publication activity, as well as keeping pace with the developments in the cloud area, collaboration with industry and EU industrial projects has been sought. Discussion for partnership with IBM, Logica, SixSq, and the EU project RESERVOIR are ongoing.

### 2.2 Geographic Scope

Discussions for support on the current proposal are in progress with various NGIs. The topic has or will be discussed with the NGIs in France, Czech republic, Italy, and UK, which are expected to be actively involved in the SSC.

The SSC aims at fostering international collaboration by enabling the research groups from different countries and regions to collaborate with the same data, to build common reference sets (benchmarks), and to propose operational challenges derived from real-world datasets.

### 2.3 Aims

The aims of the SSC are

- to provide production-quality services for the Computer researchers and engineering in Europe and beyond, through data collection, publication and descriptive analysis
- to foster basic research collaboration through scientific networking.
- to 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 GOSSC are listed below, with the related tasks in the workplan indicated

- Develop of a full-fledged acquisition process integrated within gLite (T1).
- Provide and develop on-line analysis services running on the EGI grid (T3).
- Provide a network of expertise in the interpretation of production grid behavioral models (T3).
- 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) (T2).
- Foster the creation of a COST project as a support for the basic research networking (T3).

- Define and enact processes for the specification of interpretation and control challenges , and the evaluation of the proposed solutions (T3)

## 2.4 Legitimacy

The SSC must have support from, on one hand stakeholders involved in actual production, such as some NGIs and EGI.eu, 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 maintenance tasks. The other NGIs are expected to contribute to the acquisition and interpretation tasks, and thus to commit human resources in the SSC. The GO SSC requires interaction with the future EGI.eu, 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.

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 will be re-initiated soon with the RESERVOIR project. However, the research activity will be mainly funded by applying to relevant academic, regional, national and international calls, as has already been done.

## 3 Governance Model

To be defined

## 4 Work Plan and Budget

Proposed tasks are listed below.

### ***T1. Data collection and publication***

This task will expand and consolidate the process of acquisition, long-term conservation, and publication of traces of EGEE activity and applications, that has been successfully initiated by the Grid Observatory cluster in EGEE-III

Type of activity: Service Activity (SA)

Effort: 5.5 FTE

Funding source: Commission (3.5 FTE); participating NGIs (2 FTE)

### ***T2 Grid ontology***

- Concepts describing not only resources, as in the Glue Information Model, but also the Grid inputs (users, virtual organizations and applications), the Grid dynamics (lifecycles of individual jobs and the data transfers),
- Applications to data consolidation and semantic inference.

Type of activity: Joint Research Activity (JRA)

Effort: 2.5 FTE

Funding source: Commission (1 FTE) ; participating NGIs (1.5 FTE)

***T3 Foster scientific collaboration***

Proactively engage the relevant scientific communities and institutions, create and implement the interaction processes, contribute to on-line analysis tools.

Type of activity: Networking (NA)

Effort: 5.5 FTE + unfunded

Funding source: Commission (2.5 FTE), Participating NGIs 3 FTE

***T4. Management:***

Coordinate the efforts between national, regional and European level. Manage the SSC and be involved in EGI.eu boards.

Type of activity: Management (MGT)

Effort: 0.5 FTE

Funding source: Leading participating NGI

TOTAL EFFORT: 14 FTE (7 from the European Commission)