

The Grid Observatory cluster of Enabling Grids for E-science (EGEE) aims to develop a scientific view of the dynamics of grid behaviour and usage by analysing the behaviour of the EGEE grid.

With extensive monitoring facilities already in place, EGEE grid offers an unprecedented opportunity to observe, and gain understanding of, new computing practices of e-Science. Considering it has tens of thousands of CPU's, petabytes of storage, an extensive coverage of scientific communities, and the perspective of sustainable development, the EGEE grid is one of the most exciting artificial complex systems around.

The Grid Observatory models the dynamics of the grid, using advanced statistical, learning, and signal processing methods. This can help computer science researchers and grid developers improve reliability, stability and performance.

Through its web portal, Grid Observatory offers the public a **repository of grid traces** to observe:

- The demands of e-Science users. EGEE provides a good approximation of the current and future needs.
- Grid status and middleware activity. These can be explored for a wide range of motivations, from operational usage (e.g. improving performance) to scientific usage (e.g. testing classification methods for fault detection).

The Grid Observatory seeks to give researchers a **better understanding of the grid** and through this, better optimisation.

- Application developers need synthetic characterisations of grid activity and the grid applications for predicting and optimising application performance.
- Grid models are required for dimensioning, capacity planning, or evaluating the impact of evolutions in grid configuration and middleware.
- Self-regulation and self-maintenance are desired functionalities in many areas, ranging from resource allocation to real-time fault diagnosis, including green computing as an increasingly urgent constraint.

Opportunities for collaboration

The Grid Observatory is an open project, keen to collaborate with various areas of computer science.

- The database of traces will contribute to grid research and engineering. The availability of reference datasets about usage of the grid, including job treatment, data traffic, and traces of middleware services, with some level of explanatory tagging, will be a step towards a quantitative approach of grid design.
- Organizing the grid data into a comprehensible structure includes provenance issues at the operational level and ontology issues at the fundamental level. Grid research input is needed to design an appropriate organization and indexing of the datasets.
- Interoperability with other trace repositories is a major goal, and traces from other sources are welcome.
- Autonomic computing is highly relevant at a time where production grids are moving to sustainable infrastructures, are experiencing increased usage, and reducing the manpower dedicated to daily operations.
- Machine Learning has proved successful in solving large and noisy problems that defeat the human expert.

Group contacts

The Grid Observatory portal: www.grid-observatory.org

Cecile Germain (LRI/LAL), email: cecile.germain@lri.fr
For further information write: contact@grid-observatory.org

Application webpages

EGEE is keen to consider other applications. For further information on how to participate see <http://technical.eu-egee.org/index.php?id=392>.
More information about the applications running on EGEE be found on the EGEE website at <http://technical.eu-egee.org/index.php?id=148>.