



Contribution ID: 414

Type: **oral presentation**

## Data Management in EGEE

*Monday, 27 September 2004 17:50 (20 minutes)*

Data management is one of the cornerstones in the distributed production computing environment that the EGEE project aims to provide for a European e-Science infrastructure. We have designed a set of services based on previous experience in other Grid projects, trying to address the requirements of our user communities.

In this paper we summarize the most fundamental requirements and constraints as well as the security, reliability, stability and robustness considerations that have driven the architecture and the particular choice for service decomposition in our service-oriented architecture. We discuss the interaction of our services with each other, their deployment models and how failures are being managed.

The three service groups for data management services are the Storage Element, the Data Scheduling and the Catalog services. The Storage Element exposes interfaces to Grid managed storage, with the appropriate semantics in the Grid distributed environment. The Catalog services contain all the metadata related to data: The File Catalog maintains a file-system-like view of the files in the Grid in a logical user namespace, the Replica Catalog keeps track of identical copies of the files distributed in different Storage Elements and the Metadata Catalog keeps application specific information about the files. The Data Scheduling services take care of controlled data transfer and keep the information in the Catalog services consistent with what is actually available in the Storage Elements, acting as the binding between the two.

We conclude with first experiences and examples of use-cases for High Energy Physics applications.

**Primary authors:** FROHNER, A. (CERN); MCCANCE, G. (CERN); NIENARTOWICZ, K. (CERN); BADINO, P. (CERN); KUNSZT, P. (CERN); DA ROCHA, R. (CERN)

**Presenter:** NIENARTOWICZ, K. (CERN)

**Session Classification:** Distributed Computing Services

**Track Classification:** Track 4 - Distributed Computing Services