

# A Distributed File Catalog based on Database Replication

*Wednesday, 15 February 2006 09:00 (20 minutes)*

The LHC experiments at CERN will collect data at a rate of several petabytes per year and produce several hundred files per second. Data has to be processed and transferred to many tier centres for distributed data analysis in different physics data formats increasing the amount of files to handle. All these files must be accounted for, reliably and securely tracked in a GRID environment, enabling users to analyze subsets of files in a transparent way. The talk describes a distributed file catalogue that gives consideration to the distributed nature of these requirements. In a GRID environment there is on one hand a need for a centralized view of all existing files for job scheduling. On the other hand each site should be able –for performance reasons - to have autonomy to access files without the need of centralized services. The proposed solution meets the need for a local and global operation mode of a file catalogue. Commands can be executed autonomously in a local catalogue branch or heterogeneously in all of them. The catalogue implements a file system like view of a logical name space, user-defined meta data with schema evolution, access control lists and common POSIX user/group file permissions. Architecture, interface functionalities, performance tests and very promising results in comparison to other existing GRID catalogues will be presented.

**Primary authors:** PETERS, Andreas Joachim (CERN); TISSOT-DAGUETTE, Pierre Elias (CERN); BIEGLUK, Slawomir (CERN); PINTO MORAIS, Vagner (CERN)

**Presenter:** PETERS, Andreas Joachim (CERN)

**Session Classification:** Poster

**Track Classification:** Distributed Event production and processing