CVMFS Project Plan

# Statement of Need

Current IF experiments have demonstrated a wide range of file access patterns which affect delivery of files to jobs that run on Fermilab resources and on grid [OSG] and remote computing resources. These file access requirements are currently partially satisfied by the use of Fermilab central disk services which permit the experiments to perform large scale data processing on specific Fermilab resource. This reliance on central disk limit the different experiments abilities to perform large data processing outside of the FermiGrid environment and leverage remote computing resources.

A scalable, caching file delivery system has been identified as a method of satisfying and providing significant optimization to most of the core data handling needs of the experiments. The hosting of the experiment’s code repositories a central service repository that is easily accessible to sites around the world allows for easier management of the code bases and synchronization between sites. These two features are provided in an integrated manner in products like CVMFS .

The adaptation of CVMFS or a similar system to distribute files and code bases for IF and future experiments will greatly simplify the users ability to perform data processing at Fermilab and remote sites and will reduce the maintenance and management overhead currently associated with large scale data processing and Monte Carlo generation.

# Tasks

* Gather Experiment’s User Cases and Requirements
* Evaluate potential solutions
  + CVMFS
  + Others?
* Review OSG requirements
* Evaluate CVMFS compatibility with OSG
* Develop test system at FNAL for use with FermiGrid resources
* Develop test system with OSG for use with OSG resources
* Deploy solution to FNAL resources
* Deploy solution to OSG resources
* Test and evaluate deployed systems

# Risk

The following general risks have been identified with the project:

* Experiment requires exceed technical capabilities of CVMFS
  + Large files (>1GB) delivery
  + Frequent updates of files (incremental builds?)
* OSG Incompatibility
  + CVMFS can not be synchronized properly across OSG sites
  + Update cycles are not pushed correctly
  + Delivery of large files on remote sites
  + Other OSG specific implementation restrictions
* OSG Deployment schedules are to long to be of use to current experiments