Working Group 4
Facilities and Technologies
Summary of May 27 discussions

Stephen Wolbers, lead

Workshop on Data Preservation and Long Term Analysis in HEP

SLAC, May 26-28, 2009

Outline of report

- There are costs to transition to an "archival phase". This includes the DB reconfiguration, documentation, data copying, operating system upgrades, etc.
- From the facilities viewpoint there are costs associated with the long term support of preservation services, including access.
- Funding should be provided to the facilities based on a well-defined model of data preservation and analysis (e.g. archival phase of BaBar).
- The size of the resources required for the archival phase should be estimated and agreed to by the experiment and the facility. Minimization of the data and complexity is very useful.
- Consistent agreements for support are recommended between.
  - Experiment and facilities.
  - Funding agencies and facilities
  - Funding agencies and experiments.

Page 2

- These agreements should be written to cover the short-term transition and long-term archival needs.
- Open access issues should be recognized as a key part of the process which may require different support and infrastructure than the archive phase.
- Expert assistance should be funded as part of long-term data preservation and analysis. This could be an archivist-type position filled by an physicist.
- Technology investigations should be part of this process. The focus should be on minimizing the complexity of the system and long-term support needs. However, technological change will require modifications in the future.
- Documentation is critical for data preservation and analysis and must begun from the earliest phases possible and include all parts of the system. This includes experiment, facility and other (e.g. GRID middleware) components.

Page 3

- **Validation is recognized as an important aspect of long-term data preservation and analysis.**

Definitions

- A computing facility is:
  - A computing facility that is part of an accelerator laboratory hosting an experiment.
  - A major host of data and CPU for HEP experiments.
- The archival phase of the experiment corresponds to the several years needed to fully exploit the data set. It follows the steady analysis phase. During this phase the collaboration structure becomes essentially virtual, with no central management at SLAC.