User Analysis Workgroup Discussion at the WLCG workshop

- **Proposal:**
  - Focus on collecting performance relevant information
  - Work on experiment specific benchmarks
  - I/O for analysis
    - Understanding the different access methods
    - How can we help the T2s to optimize their systems?
    - Can the ALICE analysis train model help?
    - Role of xrootd as an access protocol?
  - Clarification of requirements concerning storage ACLs and quotas
    - What can be done now
      - Protect staging, handle ownership, etc.
    - What is really missing?
      - Most SRM implementations have VOMS ACLs.
    - Short and longterm solutions for quotas
    - SRM calls

What has happened

- To improve the involvement of OSG Ruth Pordes is now co-chairing the workgroup

- Experiments, sites and SRM providers worked at understanding the problems

- The first measurements of SRM command frequencies and timing have been presented at the last GDB and at the DESY workshop
    - Agreement by all the SRM providers on common metrics
    - Complete measurements can be expected soon

- Discussions on Xrootd as a common access protocol took place
    - No agreement to use as the only protocol

What has happened

- ATLAS run the HammerCloud tests against all T2s and worked on the tuning of access parameters
  - Different access strategy for different sites
  - Some of the root-tree caching issues have been fixed
    - Need to repeat the detailed analysis

      Application I/O   <-> Network load

- CMS (Frank W.) working on similar tests for CMS

- LHCb carried out a large scale analysis exercise
  - Very high failure rates (30% on the best sites) have been observed
    - Data access based errors
  - Ongoing follow up by data management people

What happened?

- **Experiments defined with T2s split of shares between analysis and production use**
  - Steve T. started working on a template configuration as a reference

- **Experiments documented the I/O requirements for their workflows**
  - However rated are from the application perspective
  - As the tests demonstrated the I/O that the fabric sees can be several times larger

- **STEP09 is the first large scale test where all activities run in parallel**
  - This will provide necessary data to understand the interference between analysis and other activities and provide data for site wide tuning
    - WAN access to storage, analysis access, reconstruction access

Next Steps

- The critical questions that are relevant for the user analysis are followed up by the experiments, sites and storage system developers.

- Results are communicated frequently at the relevant WLCG meetings
  - Operations, GDB, etc.

- Solutions for some open issues, such as support for proper ACLs and quotas will not be resolved before the run start
  - Focus is on tuning sites and services

- It is not clear what role the Analysis WG should play
  - Collect information and point out open issues?