



# LCG Services Report April – June 2009

13 July 2009

Jamie Shiers

This quarterly report covers the second three months of 2009, during which a major activity has been the preparation and execution of a Scale Test of the Experiment Programme (STEP'09). The main results of this exercise were covered in a Post-Mortem workshop held at CERN from July 9 – 10 2009, which was attended by 100 people. At a high-level, the results from this exercise are encouraging, with many sites meeting or exceeding the metrics set by the different experiments and with a much realistic balance of concurrent (intra- as well as inter-VO) activities than in the past. On the down side, the number of major service incidents continues at a more or less constant rate, although sites are generally much better at providing reports and follow-up than previously, as well as the equally if not more important controlled recovery. Although the main additional Use Cases stressed in STEP'09 – re-processing at Tier1s and analysis at Tier2s – were successful at many sites, a small number of Tier1s and a larger number of Tier2s showed problems in certain areas. Work is underway to understand the root causes and, once resolved, the sites in question will be re-tested. At least some element of multi-VO activity will be required to ensure that the problems have been fully addressed, although it now looks like a possibly smaller scale rerun – tentatively dubbed SEPT'09 – is unlikely to happen *per se*. Instead a set of actions and milestones, due for completion by end September 2009, will be followed up by the WLCG Management Board. In addition, there are some major machine room moves (RAL, NIKHEF) or potential service upgrades (possible change of nameserver for dCache from *pnfs* to Chimera which will require re-testing in addition to the above.

In addition to the technical results of STEP'09 it is important to record that the operational load on sites is now at a sustainable level, although further improvements to experiment operations – in particular for ATLAS – are desirable. Furthermore, STEP'09 was run during a period with a non-negligible number of both scheduled and unscheduled service interventions, ranging from [multiple fibre cuts](#) in the LHC OPN through major service upgrades, such as at IN2P3, to vendor software bugs that resulted in tape drives being marked offline, as experienced by CNAF. Thus the environment can be considered representative of what one might expect during data taking and processing and was by no means an artificially stable period.

In conclusion, in STEP'09 all aspects of the experiments' computing were exercised simultaneously for the first time: simulation, data processing and analysis. This gives us confidence that the WLCG service will be able to efficiently support these activities at the required scale for the first data from the LHC later this year.

## Experiment Metrics

Whilst many of the Tier1s and Tier2s met the experiments' metrics, it is encouraging to note that of the 5 sites considered particularly successful by ATLAS, the 3 that also serve CMS performed especially well for that experiment too. There are still concerns that at least one site is very late with its pledged hardware deployment, but there is no major concern that the relatively small number of issues seen at the Tier1 level cannot be addressed on the timescale of months. Notwithstanding the fact that we should have

been in full data taking mode now, this is probably acceptable – albeit undesirable – with the current schedule for the LHC restart.

At the Tier2 level there is a wide disparity in the results – the strongest correlation being between successful sites and those that are well staffed by motivated people (which in itself should be no major surprise). It is proposed that the most efficient way to address those Tier2s that have issues is via the physics community that they support – often involved in the setup and operation of the centres directly. Concerns over data throughput exist, which may even require redesign of the internal site setup and / or more powerful storage systems. Again it should be no big surprise that the last Use Case to be fully stressed appears as the one that is least ready – although there is relatively little time for addressing the remaining concerns.

Very thorough presentations on the results of STEP’09 as viewed by the experiments, the sites and the services can be found through the STEP’09 Post-Mortem workshop [agenda](#), which also has links to more detailed internal analyses.

### Site Metrics

A question that has been repeatedly posed by the sites is how can they tell whether they are offering a good service to the experiments? Building on existing monitoring and reporting, together with the metrics established both by the experiments and at the service level for STEP’09, an attempt has been made to formalize these. Although further automation and additional monitoring and tests can be expected, the site metrics can be summarized as in the table below. These metrics were proposed at the June Grid Deployment Board in the context of the Tier1 sites but are clearly applicable to the Tier0 and most likely Tier2s as well.

We note that items 2 and 4 are currently covered by the CMS site commissioning tests and are tested continuously. 1 & 3 need to be measured on rather different timescales but are also essential: a site cannot be considered to be meeting the overall metric if it is not providing the necessary resources. Equally, sites are required to follow agreed WLCG operations standards and procedures.

| # | Metric  |
|---|---|
| 1 | Site is providing (usable) resources that match those pledged & requested;                                      |
| 2 | The services are running smoothly, pass the tests and meet reliability and availability targets;                |
| 3 | “WLCG operations” metrics on handling scheduled and unscheduled service interruptions and degradations are met; |
| 4 | Site is meeting or exceeding metrics for “functional blocks”.   |

### Service Issues

The following service summaries were also presented at the STEP’09 Post-Mortem workshop. Other than the major incidents summarized below, the services ran relatively smoothly, albeit with the usual mix of glitches and operational problems. Whilst we continue to make slow but steady progress on these issues, the risk of a big increase in

support load as the number of users increases – most likely dramatically – remains a significant concern.

| <b>Service or Component</b>                  | <b>Summary</b>   |
|--|--|
| <b>Monitoring &amp; Dashboards</b>           | <p>Existing monitoring systems, although not perfect, did provide the necessary information to follow the STEP09 activities.</p> <p>The issues and problems seen during STEP09 will define the short term development plans in the monitoring area.</p>  |
| <b>Data Management</b>                       | <p>No real showstoppers, despite earlier concerns and some incidents of data loss. On the other hand, data management operations were far from smooth - significant effort required by all parties: 1) sites, 2) experiments 3) developers.</p> <p>Configuration errors keep popping up, hw failures happen, bugs are discovered.</p>  |
| <b>LHC Optical Private Network (LHC OPN)</b> | <p>Performed well despite more than one case of fibre cuts (one time affecting multiple Tier1s with traffic rerouted over backup circuits).</p> <p>Some concerns on resilience in general and on whether scope should be extended to at least major Tier2s.</p> <p>No specific concerns about load, even with STEP'09 traffic other than in the case of use of backup links (see above).</p> |
| <b>Databases</b>                             | <p>Distributed Database Services continued to show good service robustness, reliability, availability and performance.</p> <p>Smooth running during STEP'09 at CERN and Tier1s, with load well under control &amp; within current resource limits.</p>   |
| <b>Other Services</b>                        | <p>Some continued instabilities and operational problems, as in previous quarters. New releases in the pipeline which should address at least some of these problems, but past experience has not always been positive. More than slow progress cannot realistically be expected.</p>  |

### Summary of Main Service Incidents

The following table lists the main service incidents for which a "[Service Incident Report](#)" or "Post Mortem" was produced. These incidents are typically characterized by a serious degradation or total loss of service for several hours or longer. Reporting of such incidents is still not fully systematic (some are only reported at the daily operations meeting or via e-mail), although the use of a simple [template](#) has recently been proposed. Further details can be found in the weekly reports to the WLCG Management Board.

| Site  | When         | Issue  |
|-------|--------------|--|
| OPN   | 10/06        | Multiple fibre cuts affecting primary circuits to ASGC, CNAF, KIT, NDGF, TRIUMF (incl. backup) |
| FZK   | STEP         | SAN problems affecting tape access   |
| ATLAS | 27/06        | PVSS COOL – online reconstruction stopped for 2 days   |
| ATLAS | 24/06        | Degraded PanDA service for 8 hours, impact on other offline DB services on ATLR                |
| CERN  | 11/06        | LHCb conditions access, LFC scalability problems   |
| CERN  | 18/06        | Batch & CASTOR services down for 2 hours   |
| IN2P3 | 10/06        | GridFTP transfers down for 7 hours   |
| CERN  | 04/06        | Accidental garbage collection of tape0disk1 files  |
| CERN  | 03/06        | Accidental re-enabling of garbage collection: data loss  |
| CERN  | 01/06        | DB services unavailable for about 4 hours  |
| PIC   | 23-<br>26/05 | LFC instability for 3 days   |
| PIC   | 14/05        | Down for 5 hours due to cooling problems   |
| SARA  | 04/05        | MSS tape backend down for 36 hours   |
| IN2P3 | 03/05        | Cooling problem lasting 44 hours   |
| IN2P3 | 25/04        | Robotic library outage for 7.5 hours   |
| IN2P3 | 20/04        | Robotic library outage for 12 hours  |
| CERN  | 12/04        | VOMS (2 days) and SRM (1 hour) service degraded  |
| PIC   | 10/04        | Severe performance degradation on SRM for 8 hours  |
| IN2P3 | 02/04        | Tape service down for 24h due to hardware failure of the robotic library.                      |

### Outlook for the remainder of 2009

It is currently expected that the LHC will restart in October 2009. This gives very little time for retesting and / or of further service upgrades. The main Tier1-related concerns described above started to be addressed even prior to the end of STEP'09 and will be retested over the summer and possibly in some mini-“STEP rerun” early in September. (An alternative that is also being discussed is via a set of actions / milestones to be completed by the end of September, with regular reviews at the WLCG Management Board).

Relatively few middleware or other updates are required prior to the start of data taking. These include an upgrade to [FTS version 2.2](#), which addresses a number of issues seen during STEP'09, such as the way that timeouts are calculated and handled. An Oracle patch which addresses the long-standing “bigID” problem is in the process of being applied at the Tier0 and Tier1 sites, although a further Oracle bug has been uncovered

in a related area and for which a Service Request has been opened (a work-around has also been implemented in CASTOR version 2.1.8-8).

### **Summary and Conclusions**

The WLCG service continues to deliver at a reasonably reliable and responsive level, with continued improvement as seen on the timescale of months. Well established procedures for responding to exceptions exist and are largely but not always respected. Further improvement will clearly be iterative but is nevertheless required – the acceptance of the “site metric” as described above should simplify this. The technical results of STEP’09 are globally satisfactory: a few weak areas were identified and a programme is being developed to address these followed by re-testing well prior to the restart of the LHC.

In conclusion: globally the most successful period of WLCG service operation seen to date – with the best still yet to come.