Worldwide LHC Computing Grid Project
Project Status Report
Resource Review Board – 18th Oct. 2011

Ian Bird
Version 1.0
30th September 2011

This status report covers the period from April – September 2011. Further details on progress, planning and resources, including accounting and reliability data for CERN and the Tier 1 centres, and detailed quarterly progress reports, can be found in the documents linked to the LCG Planning Page on the web.

1. The WLCG Service

During this reporting period, the accelerator has continued to deliver increased luminosities, with ATLAS and CMS exceeding 4 fb^{-1} and LHCb acquiring over 1 fb^{-1}. From the perspective of the WLCG this period has seen on-going production operation for all experiments, without significant problems. To remark is the fact that the majority of the Tier 1 and Tier 2 resources are fully occupied now, and the experiments have invested significant efforts to improve the performance of their software to make the best use of available resources.

Tier 0 performance

The Tier 0 storage system has seen data rates at a rather steady level of close to 2 PB/month being written to tape, and significantly more read back from tape. These levels are shown in the Figures below: Figure 1 shows the rates per month, while Figure 2 illustrates the total amount of data read and written by each experiment so far in 2011.

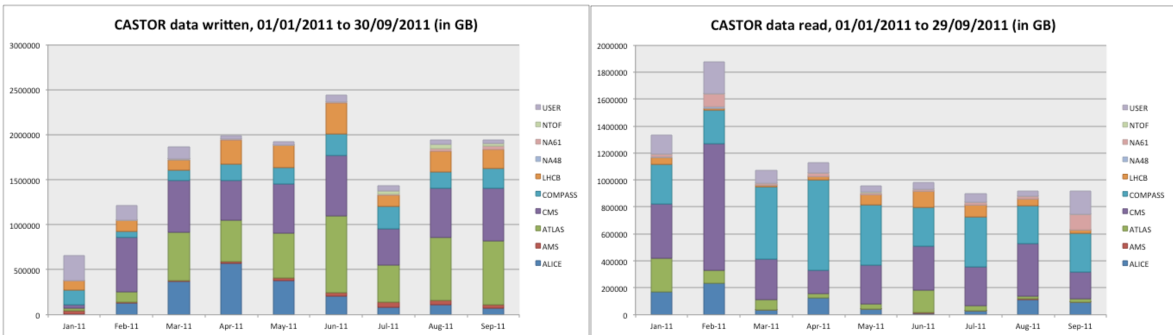

Figure 1: Data written into (left) and read from (right) Tier 0 tape

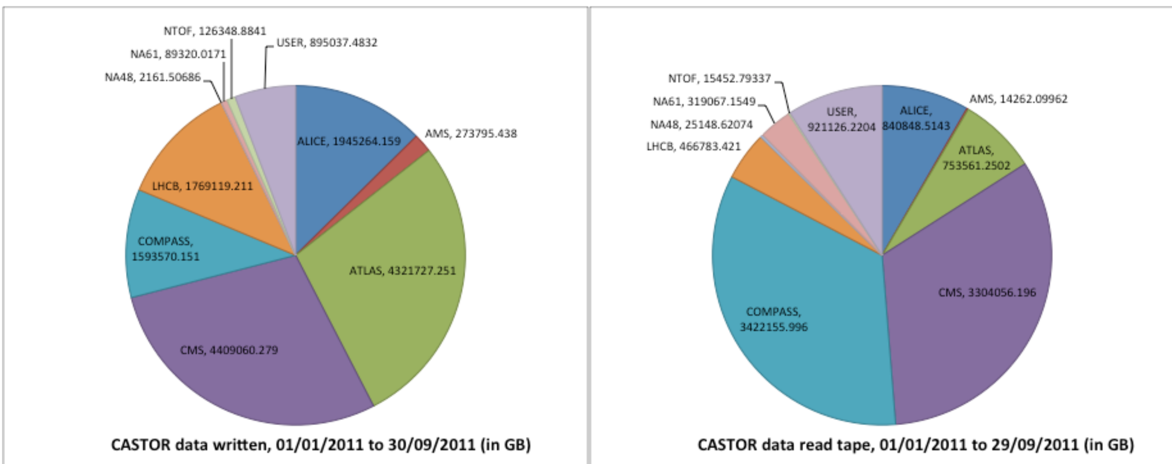

Figure 2: Total data written to tape in 2011 (left) and read (right) by experiment

The pie charts show the amount of data stored into Castor in 2011, divided between the various experiments. To note also is that by end of September some 15 PB has been stored, the expectation being that this will reach at least 20 PB by the end of the year – well in excess of the previous expectations of ~15 PB/year.

The different access patterns to tape show some of the different assumptions in the various experiments' computing models. ATLAS for example has organised itself to limit tape access to only production managers.

The total data rates through the Tier 0 mass storage system (tape and disk) continue to be at a rather high level – see Figure 3– but remain well below the levels reached during heavy ion running last year. These rates are well managed by the Tier 0 service. Data transfers out of CERN to the Tier 1 sites also continue at similar levels to those observed in 2010 and early 2011, and are also well understood and managed.

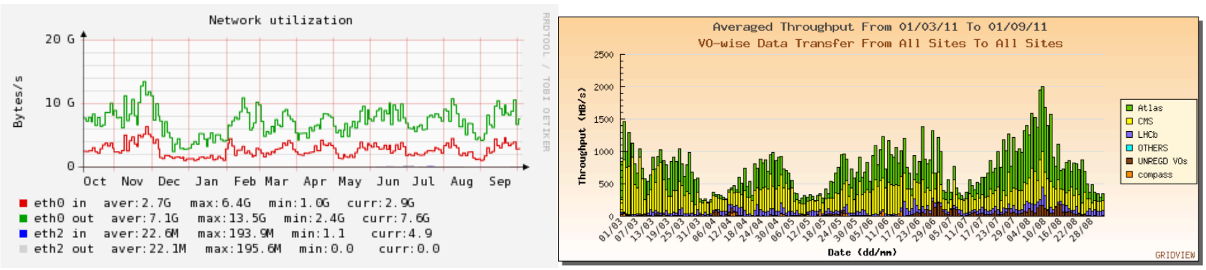

Figure 3: Total data traffic in the Tier 0 mass storage service (left), and exported from CERN (right)

During this period also, the Tier 0 has introduced a new disk pool manager – EOS – that removes some of the performance limitations of the Castor disk pools, particularly for high bandwidth data access. ATLAS and CMS have migrated significant quantities of data to the new system and report very good performance.

Use of resources and workloads

Figure 4 and Figure 5 show the continued high use of the infrastructure in terms of number of jobs and CPU usage. The CPU use remains at a high level, approximately equivalent to some 150,000 processors in continuous use. Most of the Tier 1 and Tier 2 sites are now regularly fully occupied with work from all experiments. The resource usage report presented to this meeting provides further details on this point. To note however, is that the use of resources in terms of CPU vs wall-clock time is now very good for 3 of the experiments – at a level of above 80% for Tier 1s (see Figure 6). This level is at or better than the assumed efficiency in the resource planning process. The exception is ALICE, where several different problems have contributed to a long-standing low efficiency. Most of these problems are now better understood and ALICE has made significant progress in improving this efficiency, especially for production work.

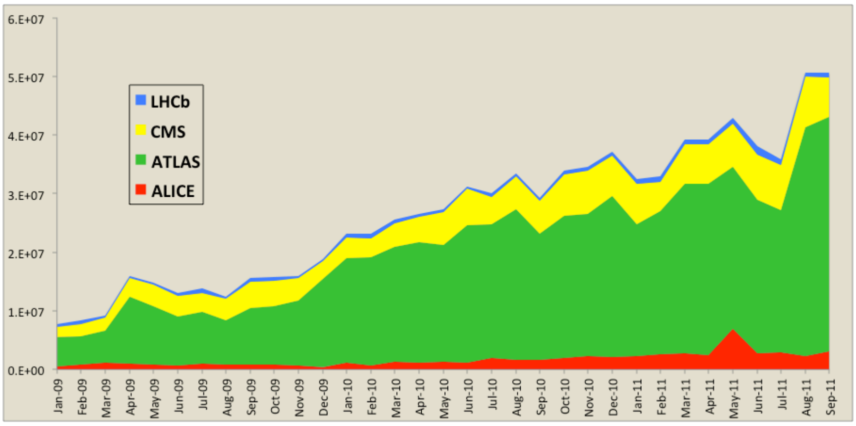

Figure 4: Continued evolution of jobs run per month; this is well in excess of 1.5M/day

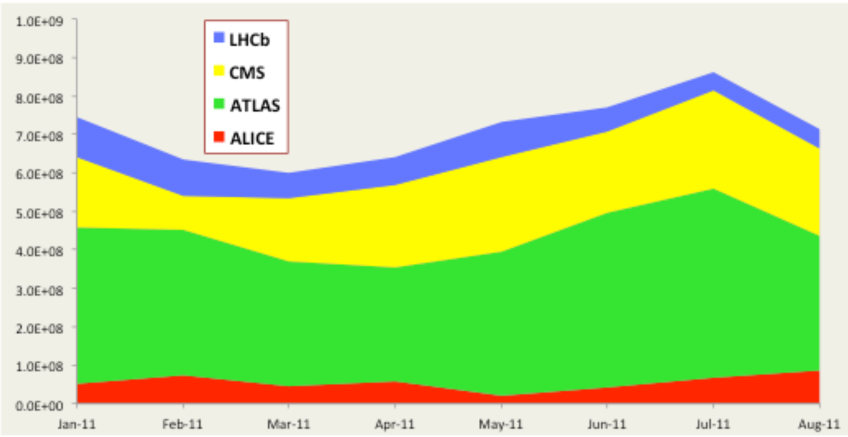

Figure 5: CPU use remains at a very high level; equivalent to ~140k CPU continuous use

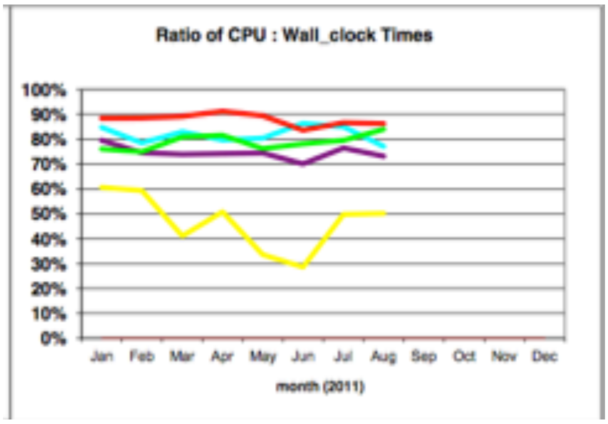

Figure 6: CPU efficiencies for Tier 0/1 sites

WLCG Service status

As agreed, significant service interruptions require a documented follow up (Service Incident Report – SIR). The full list for this period (summarised in the Table below) including the full incident reports can be seen as a summary in each Quarterly Report, or consulted on line at https://twiki.cern.ch/twiki/bin/view/LCG/WLCGServiceIncidents. These are followed by the Management Board, with the goal being that lessons are learned and disseminated to other sites.

Table 1: Service incidents requiring a review, Q2 2011 – Q3 2011

Site	Service Area	Date	Duration	Service	Impact
CERN	DB	Sep 27	1.10h	CMS Offline	CMS offline production database stuck
BNL	DB	Sep 6	4.25h	Streams for conditions	Discrepancy in an Oracle database table
IN2P3	Infrastructure	Aug 26	7.5h	CE	7500 job failures
IN2P3	Infrastructure	Aug 15	19h	CEs	CEs at 100%, others at 85% degradation
CERN	DB	Aug 09	17h	CASTOR	CASTOR nameserver database overload
CERN	DB	July 29	Scheduled+ 2h	CASTOR	Upgrade-related problems with stager DB (ATLAS and CMS)
KIT	infrastructure	July 22	5d	GGUS	GGUS alarm emails not working
IN2P3	Databases	July 19	3d	LFC, FTS, VOMS, 3D, AMI	services unavailable, some data loss
KIT	Storage	July 12	15d	ATLAS dCache	11k files lost
CERN	CASTOR	July 7th	4 h	CASTOR	Garbage Collector taking too long
CERN	DB	July 05	Scheduled+ 7h	CASTOR	Upgrade-related problems with stager DB (ATLAS and CMS)
CERN	DB	July 04	Scheduled+ 1h	CASTOR	Upgrade-related problems with stager DB (ATLAS and CMS)
Q2 2011					
CERN	CASTOR	June 26	8 h	CASTOR	CMS was unable to stage files back from tape
PIC	Computing/Storage	June 10	5h	dCache PNFS	dCache namespace overload
KIT	Storage	Jun 5	14 d	ALICE xrootd managed storage	3% of the files unreadable
CERN	Infrastructure	May 26	6 wks	KDC	high KDC load
CERN	CASTOR	May 24	6 h	DB overload on the CASTOR CMS instance	Progressive degradation gradually affecting 80% of the servie
CERN	VObox / Lxplus / SVN,CVS / Batch	May 24	3 h	XLDAP overload and nscd problem	Logins blocked, access to software version control blocked, batch jobs failed
PIC	Computing	May 25-26	12h	Batch System	BS instabilities, ~600 jobs lost
ASGC	Infrastructure	May 21 to May 23	36h	Whole Site	DC Power Cut
CERN	Batch / Lxplus /	May 10	8 h	Kerberos KDC	Logins blocked, batch jobs failed, some file access blocked

	Vobox / Lxadm / Castor				
RAL	DB	May 10	1h	LFC Outage After Database Update	>80%
ASGC	Network	May 01 to May 08	8 days	Storage service and CMS Squid	Slow transfer from/to ASGC Taiwan
CERN	DB	Apr 28	1.5h	CMS offline DB cluster	service was down for 1.5h
IN2P3	Infrastructure	Apr 8	5h	various, incl. batch system, LFC, VOBOX	job failures, various services unavailable

The number of incidents requiring a formal report has decreased compared to 2010, although there is still a clear correlation between a service change and problems, which probably implies the need for continued improvement of the change management process. Most problems are solved rapidly but several still require several days for resolution. Response times are well within the targets, and the majority of problems are resolved within 24 hours (more than in 2010). There are still a significant fraction of complex problems that require more than 96 hours for resolution.

The main sources of problems remain at the infrastructure level (site service failures typically related to power and cooling problems), storage systems, and database services. The incident rate at storage systems at Tier 1 sites has decreased significantly over the past year.

In summary the incident rate has decreased significantly despite the huge LHC performance increase, although we still see complex problems that take time to resolve. There is a strategy in place to ensure that interventions are not scheduled simultaneously at several Tier 1s.

Today the support level for sites in daily operation is at a level that is considered to be sustainable in the long term, which had not been true even 1 year ago. This is a notable success of improving the processes, although there are still areas where gains are to be made.

2. Site Reliability

The reliabilities for the last 6 months for CERN and the Tier 1 sites are shown in Table 2.

Table 2: WLCG Tier0/1 Site Reliability – last 6 months

Apr 2011-Sep 2011

Average of 8 best sites
(not always the same 8)

Month	Reliability
Apr 11	100
May 11	100
Jun 11	100
Jul 11	100
Aug 11	100
Sep 11	100

Average of ALL Tier-0 and Tier-1 sites

Month	Reliability
Apr 11	99
May 11	99
Jun 11	98
Jul 11	99
Aug 11	98
Sep 11	97

Detailed Monthly Site Reliability (OPS tests)

Site	Apr 11	May 11	Jun 11	Jul 11	Aug 11	Sep 11
CA-TRIUMF	100	100	100	100	100	100
CH-CERN	100	100	99	95	89	96
DE-KIT	99	100	100	100	100	100
ES-PIC	99	100	100	100	100	100
FR-CCIN2P3	95	99	100	100	100	100
IT-INFN-CNAF	100	100	100	99	98	100
NDGF	100	100	98	100	100	100
NL-T1	100	96	99	96	99	95
TW-ASGC	95	86	76	96	85	82
UK-T1-RAL	97	100	99	99	100	99
US-FNAL-CMS	100	100	100	100	100	100
US-T1-BNL	99	100	100	100	98	100
Target	97	97	97	97	97	97

Colors: Green > Target Orange > 90% Red < 90%

The figures below show the recent evolution of the reliabilities for the Tier 1 and Tier 2 sites. These figures are now quite stable for the Tier 1 sites and the majority of the Tier 2 sites (which provide most of the resources). It is interesting to note that good and stable levels of reliability for most of the large sites, and that of the Tier 2s, more that 75% of them regularly have reliabilities better than 95%, although some of the smaller Tier 2s could still improve their overall level of reliability quite significantly, and consequently improve their overall usage.

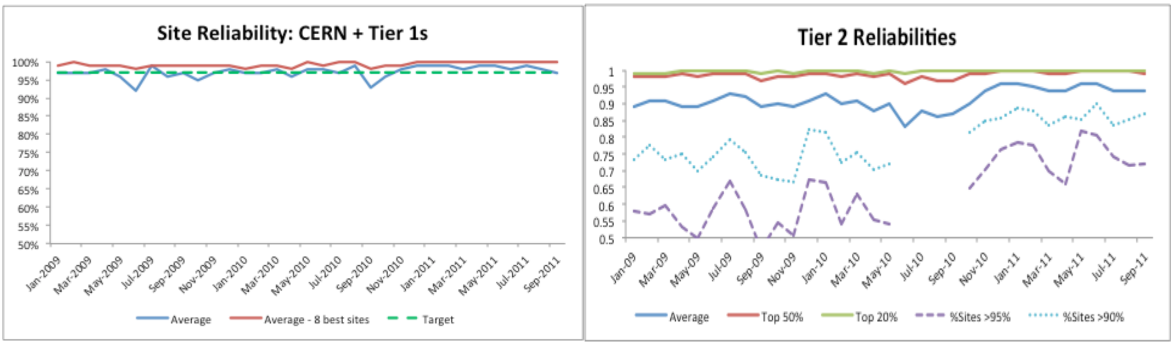

Figure 7: Site reliability evolution for Tier 0+1 (left) and Tier 2 sites (right). NB: More than ¾ of all sites now have reliabilities above 95%.

This is reflected in the typical site readiness measured independently by each experiment, with experiment-specific tests, while the reliability above is measured by the generic operational tests.

All of the availability and reliability reports for all sites can be consulted at: http://lcg.web.cern.ch/LCG/reliability.htm.

3. Applications Area

ROOT

The ROOT project released a new production version 5.30/00 on June 28. This version featured prominently, as default, a new "modern" look of the ROOT graphics as well as the introduction of CMake as an alternative build tool to generate native makefiles and workspaces (i.e. Xcode, Eclipse, Visual Studio). In addition there were many other changes in many parts of the system.

Persistency Framework

New releases of the PF projects have been prepared in Q2 2011 for the two new configurations LCG_60c for ATLAS and LCG_61 for LHCb. Both releases, motived by the upgrades to new versions of ROOT (5.28.00e and 5.30.00, respectively), include major changes in CORAL and several fixes and enhancements also in COOL and POOL. These releases also include a new Oracle client configuration to work around the redefinition in the Oracle client of some kerberos symbols, conflicting with those in the system libraries.

The new CORAL 2.3.16 code base, from which both LCG_60c and LCG_61 have been built, includes a major internal reimplementation of the OracleAccess plugin. This patch fixes potential crashes when manipulating queries created on sessions that have become invalid, in both single-threaded and multi-threaded use cases.

In collaboration with the ROOT team and several teams in IT, the PF team was also active in following up a service incident affecting the Kerberos KDC at CERN, that had initially been reported as possibly caused by the POOL software. The main cause of the problem was eventually identified as a bug in the xrootd client; ATLAS picked up a fix for this issue, already included in ROOT 5.28, in a back port to 5.26.00f.

Finally, the future of POOL was discussed with ATLAS and LHCb during Q2 2011, motivated by the recent decision by LHCb to drop POOL in favor of direct ROOT access, which will leave ATLAS as the only user of POOL. It is likely that the POOL code and the responsibility for its support will be taken over by ATLAS, on the timescale of the LHC shutdown in 2013 or possibly earlier.

Simulation

The new Beta release 9.5-beta of Geant4 has been released on schedule in June. Several contributions to the physics code have been included, to note the improvements and the new tuning applied to the Bertini Cascade, where trailing effect has been added (this was a contribution from colleagues at SLAC); improvements and extensions to the FTF hadronic model, now simulating baryon - anti-baryon annihilation, anti-baryon - nucleus, anti-nucleus-nucleus and nucleus-nucleus interactions; improved calculation of the excitation energy in the QGS model, now taking into account binding energy; several new developments in the hadronic neutron package, upgraded to use ENDF/B-VII.0 with also the possibility to make use of the complete data set from IAEA (CIEMAT team contribution). To also mention the implementation of a new alternative version of the Urban multiple-scattering model, providing improved sampling of angular distribution's tail and lateral displacement.

The new release introduces a new scheme for ordering physics processes, implementing a mechanism for assigning automatically the correct ordering to processes being registered; also introducing a new mechanism for checking Energy/momentum (E/p) conservation in all hadronic models, where limits can be set per model.

The implementation of materials has been revised and extended to allow for sharing of internal tables for dE/dx and cross-sections between similar materials, thus enabling the possibility of defining materials with common table entries and varying densities. 9.5-beta also includes an embedded module for the CLHEP classes (based on CLHEP- 2.1.0.1), allowing for the installation of the Geant4 libraries in a stand-alone mode without requiring an external dependency on CLHEP.

The web site of the Physics Validation project (http://sftweb.cern.ch/validation/) has been populated with first plots from experiments comparisons of test-beam data vs. Geant4 simulations. A second round of plots will come later in the year, with the idea to also include beam plots coming directly from the experiments.

4. Level-1 Milestones

A full report on milestones and progress can be found on the WLCG web at http://lcg.web.cern.ch/LCG/milestones.htm. Several of these have been mentioned in sections above.

Progress on specific milestones includes:

- Support for multi-user pilot jobs. Although the deployment of glexec, and associated tools has continued, with many sites now having the services running, it is nevertheless the case that making use of these by the experiments' frameworks has not been a priority in this time period, and that several real problems have still been recently uncovered. Later in this report we describe a technical evolution activity, and it has been agreed that the deployment of glexec etc should now wait until a full review of security risks and future strategies has taken place.

- CREAM CE deployment. The only remaining blocking factor for finally switching to the CREAM CE and away from the LCG-CE is the support for the SGE batch system by CREAM. This is the responsibility of the EMI project and is now promised before the end of 2011.

5. Planning and Evolution

Tier 0

Following the very large number of expressions of interest to the call for proposals for a remote Tier 0 facility, site visits and detailed discussions have been held with the majority of the respondents. A call for tender has now been published, with responses expected by the end of November. The anticipated timescale now foresees having a remote facility available for testing in late 2012 or in 2013 and entering into production in 2014. This timescale is consistent with the anticipated power evolution in the current CERN Computer Centre.

Network Evolution

There is now a fairly large-scale operational prototype of the LHCONE network, which has successfully demonstrated the principle functionalities and the concept of open exchanges. However, there have been expressed some doubts on the manageability and scalability of such a complex service if implemented with today's technology. Since there is no immediate bandwidth problem that will impact physics production, as Tier 2 networking has continued to improve, it was agreed in a recent WLCG Overview Board that it would be better to slow development of the prototype and to work on a long term solution, building on technologies that will reach production in the next year or so, rather than compromise on a short term solution.

Additional details can be seen at http://www.lhcone.net.

Technical Evolution of WLCG

Over the summer there has been a discussion on how to evolve the technical implementation of the WLCG infrastructure. Some considerations include consideration that:

- The experiment computing models have evolved
- There is today a far better understanding of requirements now than there was 10 years ago, and this has evolved even since the large scale challenges,
- The experiments have developed various workarounds to manage weaknesses in the middleware
- Pilot jobs and central task queues are almost ubiquitous, and potentially change the way workloads and trust relationships are managed,
- The operational effort is often too high; lots of services were not designed for redundancy, fail-over, etc.
- Recognition that technology evolves rapidly, and the rest of world also does (large scale) distributed computing – we should not need entirely home grown solutions,
- We must be concerned about long term support and where it will come from

However, it is important to recognise that we have a production operation that must not be disturbed. It is also important to remember that we have a grid infrastructure for good reasons, and simply replacing "grid" with "new technology" is not appropriate, although we should understand how to make best use of new technologies to implement our grid.

We should also recognise some of our considerable successes: the fact that we have an operating distributed system at an unprecedented scale, that we have the world's largest (only?) worldwide trust federation and a single sign-on scheme covering both authentication and authorization together with the very strong set of policies that are exemplars for all other communities trying to use distributed computing. In parallel we have developed the operational security teams that have brought real benefit to the HEP community. We have also developed the operational and support frameworks and tools that are able to manage this large infrastructure.

Taking these points into consideration we have agreed that WLCG must have an agreed, clear, and documented vision for the future. This will help express our community's needs to technology and infrastructure providers such as EMI/EGI, OSG, etc. A result should be to allow us to improve our middleware stack to address existing concerns. This is also an important opportunity to try and re-build common solutions where possible – both across experiments and also between infrastructures (e.g. between OSG and EGI). The work should take into account lessons learned (functional, operational, deployment, management…) and understand the long term support needs for the software. It is important that our community focus its efforts where we *must* (e.g. data management), and use off-the-shelf solutions where possible. It is also important that there be a balance between the needs of the experiments and the sites.

Thus it was agreed to set up 6 working groups to address the areas of data management, storage management, workload management, databases, security, and operations. The process will be managed by the WLCG Management Board and has a goal of initial reports on strategy for the next 2-5 years to be produced by the end of January.