Table of Contents

Week of 091005

WLCG Service Incidents, Interventions and Availability

VO Summaries of Site Availability				SIRs & Broadcasts	
ALICE⧉	ATLAS⧉	CMS⧉	LHCb⧉	WLCG Service Incident Reports	Broadcast archive⧉

GGUS section

- GGUS Escalation reports every Monday⧉ (used for WLCG Service Report to MB)
    - ♦ Full search: https://gus.fzk.de/ws/ticket_search.php⧉ and select "VO" - this will give all tickets, including team & alarm
- **LHCOPN** Tickets in GGUS: see https://gus.fzk.de/pages/all_lhcopn.php⧉ and change your selection criteria. Future actions are also listed.
    - ♦ If network group participation is necessary, please invite them in time.
- OSG items selected from the GGUS escalation reports⧉. (MariaD) No overdue tickets this week.
- Procedure to become a LHC Experiment VO TEAM member⧉
- Other recent GGUS FAQs⧉

Daily WLCG Operations Call details

To join the call, at 15.00 CE(S)T Monday to Friday inclusive (in CERN 513 R-068) do one of the following:

1. Dial +41227676000 (Main) and enter access code 0119168, or
2. To have the system call you, click here⧉

General Information			
CERN IT status board⧉	M/W PPSCoordinationWorkLog	WLCG Baseline Versions	Weekly joint operations meeting minutes

Additional Material:

STEP09	ATLAS	ATLAS logbook	CMS WLCG Blogs⧉

Monday:

Attendance: local(Lola, Harry(chair), Andrea, Danielle, Andrew, Gang, Gavin, Roberto, MariaG, MariaD, Jamie, Ignacio, Jean-Philippe, Olof, Jan, Alessandro, Nick, Steve, Dirk);remote(Gonzalo/PIC, Kyle/OSG, Michael/BNL, Ron/SARA, Gareth/RAL, Tiziana/CNAF).

Experiments round table:

- ATLAS - The throughput test started this morning with today mainly testing the behaviour of FTS 2.2 by exporting large numbers of small functional test files. Up to now problems with RAL in unscheduled downtime and Naples down since a few hours. Will continue with small files tomorrow then increase the file size on Wednesday. Michael queried that he could not see any exports going to BNL and Alessandro agreed saying that they are looking into why this is.

- CMS reports - Started the analysis focussed October Exercise today. Software for this was released on Saturday with 83% of the target sites installed within 2 hours. Commissioning the remaining T2-T2 links for this exercise has made good progress and there are only 4 T2 production sites with

problematic transfers. Meanwhile one CMS tape was lost at PIC over the weekend and there are source site transfer errors from FNAL as source to Perugia. CMS started the ggus alarm tests last Friday but there are conflicting dates for this. It showed up a problem for CNAF where the site shown in a scroll down list for CNAF is the wrong one. MariaD responded that she had written to remind VOs to run these tests from the 28th in order to have a summary for the gdb on the following Wednesday but in fact the gdb is not until 14 October so the tests should be this week. Jan Iven pointed out that their CERN test last Friday did not exercise the full chain as it was rejected by the operator as not concerning one of the piquet supported services of CASTOR, SRM, FTS and LFC. CMS will repeat the test this week. MariaG added that tests were sent by CMS and LHCb, but not ATLAS who would do so this week, and which showed problems at RAL where their GOCDB entry had disappeared and CNAF with the wrong site name. Jan Iven also said that not all sites received the test mail and gocdb was investigating what was sent.

- ALICE -

- LHCb reports - MC productions and stripping of minimum bias events has been stopped due to internal LHCb issues. Will resume when these are fixed and backlog has been drained. CASTOR at RAL outage since Sunday. Issue with stripping jobs at CNAF on Friday was found to be an acl problem in the storm file system of LHCb.
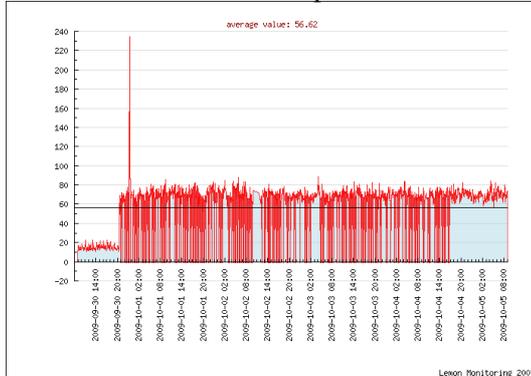
Sites / Services round table:

- BNL: There will be a major outage of the tape system till Thursday for an HPSS upgrade.

- SARA: Due to recent activities the dpm at NIKHEF created a large log file on their SE filling the /var file system. This has now been extended. Two downtimes coming - the FTS and LFC of ATLAS will be moved to new hardware and the worker nodes will be upgraded to SL5.

- RAL: In unscheduled downtime of CASTOR since yesterday. ORACLE DB runs on a SAN with two disk systems containing the DB and a mirrored copy. Mirrored disks had a fault on Sep 10 and have not always been used since then yesterday the primary system developed a similar fault. CASTOR is down and there is some evidence of DB inconsistency so they may have to roll back to Saturday and lose some transactions. Vendor is looking at problem and outage is currently scheduled until 14.00 Tuesday. Also had batch problems on Saturday when the pbs logfile reached 2GB. The LHCb 3D db has been migrated to 64-bit hardware and was reported (by Roberto) to be running better.

- CNAF: 1) Testing 64-bit SL5 worker nodes with ALICE and ATLAS. 2) This week will test storage with CMS using TSM as the tape backend instead of CASTOR. 3) Had issues with the CMS WMS due to a huge peak of submitted jobs at 09.00 this morning and want to know what is to be expected from the October exercise. At the moment they have moved some underused WMS to support CMS. Daniele reported the peak as due to the exercise starting and users will now be digesting results at different rates so the peaks should now be less though difficult to predict. CMS are in the phase of tuning the exercise. 4) Friday's alarm tickets came with the wrong site name coming out of ggus though these used to work. This exposed side effects that they could neither update the local or ggus tickets. As a work-around they have created a second site name. Apparently ggus switched to a name approved by the MB and this is not the same as the one used in Gocdb. MariaD will open a ggus ticket on this as such a hack should not be necessary and follow up on the naming.

- ASGC: 1) New tape drives (report from Jason said 18 to add to the current 5 for LHC) should be delivered at the end of this week. 2) Local DB operations team are asking for help to synchronise a DB table space. MariaG had not seen any request which may have been sent to the ATLAS DB team. She will mail Jason at ASGC for clarification.

- CERN services: 1) LHCb online DB cluster have a scheduled intervention today ending at 16.00 to

replace a faulty disk. 2) LFC instabilities yesterday. 3) CE128 (one of only two lcg-ce submitting to SL5) filled up /tmp.

CERN LHC Voms services

The VOMS issues reported on September 31st are still present. While the service is no longer continuously restarting the time for **voms-proxy-init** has increased by a multiple of three. The reason for this increase is not understood. Plot shows for voms115 the sum of time taken for voms-proxy-init across all 12 VOs. Plot Sep 30th to Oct 5th.:

AOB:

- dCache golden release:

```
Dear dCache community,

The dCache Golden Release 1.9.5 is available for download and testing. This release is meant to b

The system has passed our functional tests.
However, doing intensive stress testing, NDGF found a dead-lock in the SRM front end on high load

Nevertheless we would encourage sites to install this release on their test system(s), simulating

   We keep you updated on any progress
           your dCache team
```

- GGUS issues (MariaD):

1) We asked 5 times for T1/VO data to design the new workflow for ticket https://savannah.cern.ch/support/?108277 . We received nothing. Should it be closed? Note that the general policy here is to close tickets where no reply to a request for feedback has been made after one month as described in https://savannah.cern.ch/support/?108277#comment12 2). The multiple names issue for BNL_ATLAS appearing in the OIM view that GGUS uses is still not solved. Please see https://savannah.cern.ch/support/?109779#comment8 and more recent comments to see where we are.

Tuesday:

Attendance: local(Daniele, Harry(chair), Jean-Philippe, Olof, Gang, Jamie, MariaG, Flavia, Andrea, Alessandro, Roberto, MariaG,Patricia);remote(Xavier/FZK, Michael/BNL, Ronald/NL-T1, Gareth/RAL, Fabio/IN2P3).

Experiments round table:

- ATLAS - Throughput test with small files continued today. The transfers problem to BNL reported yesterday was due to a misconfigured FTS channel now fixed - data was going via the CERN-* channel rather than CERN-BNL. The state of the small files test is being validated with the plan to

switch to larger files tomorrow. Only the RAL Tier-1 is down. CNAF are publishing strange numbers for the MCDISK and DATADISK space tokens - going up and down by 200TB.

- CMS reports - Focussing on the two week October analysis exercise. 1) Improving the gLite WMS jobs going into failure/abort states with good support from the developers. 2) WMS at CERN showing proxy delegation failures was fixed by a full reconfiguration on Monday morning. 3) Migration backlog at CNAF still being investigated. 4) Files from the tape lost at PIC have been invalidated so this is under control. 5) Some Tier-2 sites having over 50% of jobs failing (eg London-IC).

- ALICE - No large scale production at the moment. Beginning to debug what they call hybrid sites where their VO box is in SL4 but worker nodes are a mix of SL4 and SL5. Sites are asking to install an SL5 VO box but release has been delayed following a security bug introduced during the build process. A rerelease will be made in a week or two.

- LHCb reports - 1) New DIRAC release in production solves their bookeeping registration problem and MC and stripping production has restarted. 2) Last weeks FEST test ran only till Thursday and was stopped by a major issue in the online. 3) The new CERN afs UI put in production yesterday prevents DIRAC services to load the lcg-utils libraries. Being investigated. No problem reports from the other experiments. 4) LHCb reported a slow deletion rate of CASTOR files of about 0.3Hz some weeks ago and now request to change the hard coded idle connection timeout from 60 to 10 seconds which should give a 2 Hz deletion rate. Olof reported the problem was in fact due to a race condition between multiple stager processes and proposed a different workaround in that they will stop the redundant processes.

Sites / Services round table:

- BNL: Had an incident yesterday at 14.00 EST when a dcache configuration sent data into a wrong space token. As a result space was reserved for files in the BNL dcache but transfers failed and the FTS at CERN kept retrying though it had been agreed not to retry. Olof will follow up.

- NL-T1: Connectivity problems in their grid services this morning when an internal network switch reset and took time to reconnect.

- RAL: Preparing some old hardware to be reused to restart CASTOR services then recertify the failed new hardware. This will recover the nameserver to some 4 hours before the failure. However at about midday had similar disk failures on racks hosting LFC and FTS services. May be temperature related as it happened while floor tiles were being removed and the earlier failure occurred after an air conditioning unit had failed changing the temperature distribution in the machine room. Currently checking consistency of the databases especially that of LFC.

- ASGC: Getting some information from other Tier-1 to help resynchronise their tables. MariaG repeated CERN thinks the ASGC DB is still corrupt and they have mailed Jason with specific questions.

- CERN Databases: Vendor intervention yesterday on the LHCb online cluster. Queried the status of dba's at RAL this week as one primary contact will be away - reply was cover will be assured.

AOB: 1) (MariaD) wrote summary of the BNL_ATLAS >1 name inconvenience in https://twiki.cern.ch/twiki/bin/view/EGEE/SA1_USAG#BNL_ATLAS_names, sent it to Arvind and added it in https://savannah.cern.ch/support/?109779 If Michael (BNL) could help with OSG GOC etc it would be great. Michael confirmed they are working with OSG to solve this issue. 2) ggus have now confirmed that the test ticket of last week was only sent to LHCb and not the whole string of the alarm list. The tests will be repeated this week. Olof reported CERN has received test alarm tickets from ATLAS and CMS.

Tuesday: 4

Security Issues: Last August a serious vulnerability in Scientific Linux affecting many WLCG sites was discovered. Patches were prepared and widely publicised and site and regional managers and security contacts have been circulated several times on the need to deploy these patches. The EGEE Project Management Board has now authorised the suspension from the WLCG of those sites which have not deployed the necessary patches of which there are over 50. A final 48-hours warning will be sent tomorrow with suspension (at the global bdii level) being activated on Friday 9 October. Sites with any doubts should contact their ROC security contact directly or via the mailing list of project-egee-security-support@cernNOSPAMPLEASE.ch.

Wednesday

Attendance: local(Somone, Harry(chair), Jean-Philippe, Roberto, Romain, Antonio, Gang, MariaG, Lola, Alessandro, Ricardo, MariaD, Olof);remote(Daniele/CMS, Fabio/IN2P3, Gareth/RAL, Michael/BNL, Ronald/NL-T1).

Experiments round table:

- ATLAS - 1) Throughput test continues till Friday. Still in phase of many small files but throughput will increase today when file sizes are increased. In preparation data had been distributed according to MoU shares but this results in insufficient subscriptions so more data has been distributed. The plan is to run at 3 GB/sec out of CERN equally to the ten Tier-1 sites. The rate will be limited to 3 GB/s since the Tier-0 is still taking data. A problem of excessive retries in FTS 2.2 was cured yesterday. Some pathological transfers were being retried 100 times due to a bad parameter passing. A patch has been prepared but not yet deployed. Fabio asked the goals of this exercise. Reply was to make a final stress test of Tier-0 data distribution and to validate FTS 2.2 including its use of checksums (due to be switched on on Thursday).

- CMS reports - Have been focussing on the proposed site suspension action - see below under security. Following up on Tier-2 issues and suffering here and there from stageout problems but very few site issues.

- ALICE - NTR.

- LHCb reports - 1) MC production, stripping and user jobs now running at about 11000 concurrent jobs. 2) Reran GGUS alarm ticket test with 5 sent from a different submittor than last week. 3) CERN UI issue reported yesterday fixed by removing all python path references from the UI. 4) Another storm acl problem at CNAF - quickly fixed. 5) Olof discussed the timeout change request made yesterday (to increase the CASTOR delete rate) and the developers will prepare a clent fix, deployed as a client library change, as a proper solution.

Sites / Services round table:

- RAL: Working to get all three services of CASTOR, LFC and FTS back on alternate hardware so three new separate clusters.FTS and LFC should be back this afternoon though LFC may have lost a few transactions. CASTOR outage is scheduled till midday tomorrow. Decided also to move the 3D service which will be a bit later. In parallel investigating into the disk storage failures. Still considering temperature but now also other environmental factors such as the power supply - all are on the same phase. When the services are back they will investigate the failing hardware more aggressively. MariaG offered specialised help in trying to minimise data loss when the point-in-time restore is made - currently likely to be about 4 hours before the failure.

- BNL: HPSS upgrade moving ahead on target notably with the database conversion.

- Security (Romain): Sites that will be suspended on Friday have all been contacted (the list was sent to the LHC experiment representatives at the daily meeting yesterday). Sites that start to upgrade will be

allowed to finish but those asking for days more of time will be suspended. Alessandro pointed out that ATLAS use a local cache of site bdii information and also have some functionality directly addressing sites that would not be suspended. Romain agreed that the security team should extend their metrics and, if possible, mechanisms to cover these cases with Alessandro suggesting some sort of TAG. Romain reported all ROC's, site managers and site security managers had been warned and broadcasts sent repeatedly since 24 August. Alessandro pointed out that some sites (eg Beijing) seem to only respond to ggus tickets but this, of course, would expose the vulnerable sites. For CMS Daniele said they had found 6 of their Tier-2 and 2 Tier-3 sites on the list and their CMS contacts had been mailed. One Tier-2 has upgraded and another is so-doing (details to be sent to Romain) while only one of the remaining 6 is participating in the October exercise. It would be good to understand if there are sites where only the experiment contact reacts.

- FNAL: Did not receive the first round of ggus ticket tests. MariaD will follow up at a developers meeting she has tomorrow.

- CERN Databases: 1) The ASGC database is definitely corrupted and information has been exchanged with Jason. Answers are awaited from the ASGC operations team. 2) The LHCb online now has a network switch failure which prevents their cluster interconnect from working so they are running in degraded mode.

Release report: deployment status wiki page : New update set 56 was released yesterday containing important fixes to the CREAM-CE. All CREAM sites should deploy this as soon as possible. Next week there will be an important re-release of many services correcting the security flaws made in a recent build. Some ten different services will need to be restarted to deploy this release.

AOB:

Thursday

Attendance: local(Jan, Olof, Roberto, Harry(chair), Eva, MariaG, Jean-Philippe, Lola, Gang, Andrew, Daniele, Alessandro, Ricardo, MariaD, Simone, Dirk);remote(Felice/INFN-T1, Gareth/RAL, Onno/NL-T1, Michael/BNL, Fabio).

Experiments round table:

- ATLAS - 1) File sizes for the throughput test were increased yesterday but lead to site services problems on an overloaded virtual machine. File sizes were further increased and the frequency decreased and exports from CERN are now running at the target 3 GB/sec. 2) A maintenance cron script has started hanging leading to a build up of connections to a DB server. To be protected against e.g. via a lock file. 3) Michael observed that that after the RAL LFC came back yesterday DQ2 caused sites to try and pull data from RAL when it should have been down with resulting transfer failures causing srm problems at many sites including BNL. Alessandro said there was a blacklisting system and asked to see the BNL configuration file. It was concluded that some sort of atlas-wide information system was needed so that source and target sites could be centrally blacklisted.

- CMS reports - October exercise continues. No updates on the gLite WMS proxy problems - the ticket left is probably a remnant. No open Tier-1 tickets nor major transfer errors and closed 8 Tier-2 tickets in the last 24 hours. In the analysis exercise most support requests are coming from physicists using internal channels such as hypernews. Most Tier-2 commissioning has finished giving a wide participation in the October exercise.

- ALICE - Not yet resumed full production.During today's ALICE task force meeting they will discuss the progess of the gLite VO-box middleware (delayed by security issues coming from a recent build).

- LHCb reports - Impressive number of 21000 concurrent jobs. No new tickets opened but IN2P3 had a pin manager outage this morning fixed locally. A new round of ggus alarm tickets has been completed with all 7 sites succeeding.

Sites / Services round table:

- RAL: LFC and FTS services resumed yesterday. CASTOR databases being restored at this moment expected to finish at the end of today. There are two separate databases with the common nameserver, CMS and ALICE on one and ATLAS and LHCb on the other and the point in time restores are different. For ALICE and CMS the recovery should be complete while for ATLAS and LHCb it will be to a point a few hours before the failure with an estimate of 10000 files lost. Two approaches for ATLAS and LHCb - either prepare a list of the lost files and then restart - estimated at several hours work - or attempt to recover lost files by resynchronising - estimated at several days work since tools would have to be developed. Simone asked if RAL could join todays ATLAS Distributed Computing operations meeting to discuss this and expressed a preference to just have a list of lost files as they were probably not unique. LHCb thought the same and CMS had no objections. Decision to be taken when ATLAS and LHCb have looked at the lists of lost files. Final point from RAL was that alternate hardware for the 3D service should be ready by the end of today. As regards the cause of these errors suspicion is now with power problems. A vendor engineer is on site.

- BNL: 1) Since yesterday many CERN to BNL transfers are failing with connection refused to a variety of CERN hosts (lxfs disk servers) and this has also been observed at ASGC and SARA. This happens in the ftp phase so is unlikely to be related to the new FTS. Michael will open a ticket containing the suspect hosts. 2) They will complete their HPSS upgrade later today. 3) There is an all-day scheduled dcache upgrade to version 1.9.4 next Tuesday. Apparently the higher level dcache golden release is not yet regarded as production ready.

- IN2P3: SRM service went down at midnight being restored at about 08.30. Suspected to be an OS problem. There was then no data from ATLAS till 11.00 so was human intervention needed ? Simone explained the gap was the usual weekly ATLAS rotation of the functional tests so just a coincidence.

- CERN: 1) Several operator issues with the alarm tickets may lead to some procedures being revised. 2) The LHCb online database is now back to 3 out of 4 nodes working properly. 3) Downtimes are planned next week to migrate the CMS, ATLAS and LHCb online databases to new hardware using the dataguard software.

AOB: MariaD reported from a meeting with the ggus ticket system developers. The alarm submission form will now contain both LCG and GOCDB site names and behind the scenes (for Italy at least where the local ticketing system uses this for direct communication to sites) use the GOCDB name (solving the CNAF problem reported yesterday).

Security: CMS would like a list of any OSG sites affected by the security vulnerability leading to suspension of EGEE nodes. They noticed that an OSG warning was sent to Nebraska some time after CMS had themselves informed the site so query how coordination with OSG is being done.

Friday

Attendance: local(Lola, Harry(chair), Jamie, Simone, Roberto, Jean-Philippe, Gavin, Jan, Alessandro, Eva, Olof, Ignacio);remote(Jeremy/GridPP, Onno/NL-T1, Suijan/ASGC, Angela/FZK, Michael/BNL, Daniele/CMS, Gareth/RAL).

Experiments round table:

- ATLAS - This is the last day of the current throughput tests.The load generator was cranked up yesterday to reach the rate of 3 GB/sec out of CERN and this has been sustained for almost a day. The Tier-0 ATLAS CASTOR pool shows an output rate of 4 GB/sec which includes 1 GB/sec to tape with an input rate of 1 GB/sec coming from the load generator giving confidence that this will sustain real data acquisition. A second part of this exercise was to test FTS 2.2 and here results are not so good as first reported yesterday by Michael at BNL. Gavin then gave his analysis - There is a new design in FTS keeping separate request pools for srm and gridftp with an initial ratio of two srm requests per gridftp request. For both CASTOR and dcache this assigns server resources at each get and put and often when srm tries to process its next request the gridftp process has timed out and the transfer fails. In investigating this the CERN-BNL channel was changed to a ratio of 1.5 srm requests per gridftp and also the total number of concurrent transfers was reduced to 40. The error rate has dropped, though it is still 8-9% of failures, and we are checking which change did the most improvement. Simone queried if going to a 1-1 ratio returned behaviour to that of the previous FTS but the answer was that this was not the case. Simone continued that the ratio CERN-SARA was changed to 1.5 but there are now many transfers stuck in the active state for many hours so these should be checked. Also the ATLAS site getting most data is, correctly, BNL but the second is ASGC which, as a 5% site, should be behind IN2P3 and NL-T1 so this should be checked. Then checksumming has been turned on at some destinations and found not to impact performance notably at NIKHEF which uses dpm with gridftp calculating the checksum on the fly. ATLAS plans for the weekend include a big export of production data so these throughput tests will finish today. Michael pointed out that with on the fly checksumming the file replication rate at BNL was quite low at 1100 per hour and it would be good to better understand the checksumming overhead. Simone suggested a test with larger numbers of smaller files would be good for this - to be discussed at next weeks ATLAS T1/T2/T3 jamboree meeting.

- CMS reports - Not much to report for Tier-1 or services. Seeing many "file could not be opened" errors at some Tier-2 sites (Caltech, Purdue, Florida, Desy) and transfer errors with root cause at Tier-2 at Lisbon, KIT, Nebraska and Florida.

- ALICE - 1) MC and the RAW data production is still stopped due to some (internal) Alice sw issues. In any case the system has been kept busy due to the analysis activities which have continued. 2) All ALICE sites are going to be (transparently) upgraded with a new Alien version by the beginning of the next week. The new version will contain some refinement of the storage packages and the information providers used for the CREAM-CE submissions. 3) By the beginning of next week the glite3.2-VOBOX is expected to be available to all sites according to current deployment plans.

- LHCb reports - For the first time have all types of LHCb jobs running at the same time with a total of 18000 over the grid. 1) A CERN AFS server failure at 10.45 yesterday affected many other servers, including LSF batch and LHCb's DIRAC, causing a one hour outage (FIO will do a post-mortem with platform-LSF to understand this). 2) CERN lcg-ce ce128 and ce129 (those submitting to SLC5) failing large number of pilot jobs since yesterday with gatekeeper configuration errors (understood by FIO - to be fixed). 3) Waiting for the list of lost files from RAL. 4) Large amount of user and stripping jobs failing at dcache sites except PIC. Jobs being killed by dcache watchdog after being stuck in opening a connection or failing to retrieve a tURL. Possibly load related - being investigated before contacting sites. 5) Gavin reported on follow-up to the problem of slow srmrm (srm deletes) reported by LHCb earlier this week. A new CASTOR 2.1.8-13 to fix this problem will be ready at the end of October and this is satisfactory to LHCb.

Sites / Services round table:

- NL-T1: The SARA SRM has a broken disk pool - on older system of 7TB probably used by several experiments. They hope to have copied its data elsewhere by Monday.

- BNL: The HPSS upgrade, to release 7.1.1, has been successfully completed.

- RAL: CASTOR restores are completed and being checked. Services for CMS and ALICE will be brought back soon. Working to produce the lists of lost files for ATLAS and LHCb and will wait for feedback from the experiments (for ATLAS this will be via Frederic Brochou) before opening up for them. Hardware is now ready for the 3D databases but there are issues getting streams replication working again.

- ASGC: For 3D DB will try and synchronise with SARA.

AOB:

- On Monday 12th October, 513 R-068 will be used for a WLCG-ASGC meeting (closed). We will still use the room for the daily operations meeting - so pleased don't be alarmed if you arrive to find a full room! We will break the WLCG-ASGC meeting 14:58 - 15:30 (less if possible) for the daily WLCG ops.

-- JamieShiers - 2009-10-02

This topic: LCG > WLCGDailyMeetingsWeek091005
Topic revision: r19 - 2009-10-09 - HarryRenshall