

Tier-2 storage support document (v 1.1)

Philip J. Clark, Greig A. Cowan, Jamie Ferguson
University of Edinburgh

Jens Jensen, Jiri Mencak, Owen Syngé
Rutherford Appleton Laboratory

October 10, 2005

Abstract

Storage Resource Managers (SRMs) provide a standard interface that facilitates the management of disparate storage resources in a global Grid environment. This document details how the GridPP storage group will facilitate and support the deployment of this interface at UK Tier-2 sites.

Keywords: Storage, SRM, dCache, DPM, Tier-2



Document change log

Issue	Date	Comment	Author
1.0	01/10/2005	Initial Version.	Text by Greig Cowan
1.1	10/10/2005	Incorporated comments from GridPP storage group, documentation officer and project manager.	Text by Greig Cowan

This document will be updated regularly as deployment of SRM interfaces to storage takes place across GridPP Tier-2 sites.

1 Overview

One of the main tasks of the GridPP storage group is to support UK Tier-2 sites in the deployment and management of their storage resources as part of the GridPP project. This document aims to provide a starting point for sites who are attempting to deploy a SRM interface to their storage. Further details can always be found within the regularly updated wiki [1].

2 Support structure

The storage group consists of a set of core personnel (see Table 1) based at Edinburgh and RAL. In addition, there is a wider storage community formed by the members of the GridPP storage mailing list [2] who constitute LCG site administrators and other members of GridPP concerned with the use of storage resources.

Who	Email	Location	Time	Main role in supporting Tier-2's
Phil Clark	p.j.clark@ed.ac.uk	Edinburgh	15%	coordination
Greig Cowan	g.cowan@ed.ac.uk	Edinburgh	100%	dCache, DPM, documentation
Jens Jensen	j.jensen@rl.ac.uk	RAL	50%	coordination
Jiri Mencak	j.mencak@rl.ac.uk	RAL	30%	dCache, DPM, documentation
Owen Syngé	o.m.syngé@rl.ac.uk	RAL	100%	dCache, DPM, documentation, SRM

Table 1: Core GridPP support personnel.

Support to Tier-2 sites is provided via a number of routes.

- GridPP storage wiki [1].
 - Contains information about all subjects related to storage.
- GRIDPP-STORAGE mailing list [2].
 - To join the list, either go to the list homepage and follow the joining instructions or email Jens Jensen with your request. The homepage contains the archive of past postings to the list.
- Weekly phone conferences (Wednesday, 1000-1030).
 - Meeting agendas, minutes and call-in details are announced to the GRIDPP-STORAGE mailing list.
- Storage group bug/action tracker at the support project page [3] provided by CERN Savannah.

3 Tier-2 requirements for SC4

The preparation phase of Service Challenge 4 (SC4) starts now **October 2005**. By 31st January 2006, the basic middleware components should be delivered and in place at each participating site. Integration testing will take place during February and March in preparation for the start of the throughput phase on **April 1st 2006**. The following is taken directly from Ref. [5] and details the specific Tier-2 related requirements for SC4:

The following services shall be provided by each of the Tier2 Centres in respect of the LHC Experiments that they serve, according to policies agreed with these Experiments:

- provision of managed disk storage providing permanent and/or temporary data storage for files and databases;
- provision of access to the stored data by other centres of the LHC Computing Grid and by named analysis facilities as defined in [5];
- operation of an end-user analysis facility;
- provision of other services, such as simulation, according to agreed Experiment requirements;
- provision of network bandwidth and services for data exchange with Tier1 Centres, as part of an overall plan agreed between the Experiments and the Tier1 Centres concerned.

All storage and computational resources shall be *grid enabled* according to standards agreed between the LHC experiments and the regional centres.

Grid enabled storage implies that each site must have a Storage Resource Manager (SRM) interface to their local disks and tape. The deployment of this interface is the subject of this document.

The service phase of SC4 starts on **May 1st 2006** and by **September 1st 2006**, the initial LHC service should be in stable operation. The first LHC event data is expected during the summer of 2007. GridPP has setup a mailing list for all UK sites to join who are involved in the Service Challenge [6].

4 Storage Resource Management

The amount of scientific data produced by modern scientific experiments has reached enormous levels that cannot be economically stored at a single site or country. Instead, collaborating institutes that are geographically distant must share their available storage resources. Middleware components are required to dynamically manage these shared resources and datasets. Storage Resource Managers (SRMs) provide a common interface to all formats of data storage systems including tape and disk, irrespective of hardware and operating system used [7]. This makes SRM suitable for use within a Grid framework. The SRM application program interface itself is a collaboration [8] between Lawrence Berkeley (LBNL), Fermilab (FNAL), Jefferson (JLAB), CERN, and RAL. It is also a GGF working group, GSM-WG [9]. A storage element that supports an SRM interface is often called “an SRM”.

4.1 dCache

dCache is a disk pool management system, able to provide a disk cache to a hierarchical storage system, that has been jointly developed by DESY and Fermilab. Clients can access dCache through a number of protocols, the most important of these being the SRM protocol (currently provides an SRM v1.1 interface). GridPP documentation on the installation and administration of a dCache system can be found in the GridPP wiki [10], which also points to other sources of information.

dCache is a mature product that has been successfully deployed at a number of large institutes around the world (i.e. RAL Tier-1, SARA, FZK, IN2P3, BNL) and more recently at smaller institutes (i.e. GridPP Tier-2 sites involved in SC3).

The high levels of customisation and optimisation that are available in dCache lead to it being a very flexible solution for different site setups. However, the increased complexity can lead to difficulties in understanding and deploying the system. To help with the managing of the system, dCache provides web-based monitoring of its services, in addition to an administration ssh interface with an optional GUI.

From a novices point of view, dCache initially proved difficult to install and configure by hand, primarily due to the lack of user experience the lack of documentation. This situation has been greatly improved through the work of the GridPP storage group. Contributions have been made to LCG YAIM and the GridPP wiki has been populated with significant amounts of documentation detailing the installation, configuration and administration of a dCache system. Installation of dCache now proceeds just like any other LCG node. A basic system comprising of a head node (which provides the SRM interface, GridFTP access, system databases, web monitoring and the administration interfaces) and a single pool node (which provides storage that is accessed via the head node) is simple to setup and publish in the LCG information system. Adding extra storage pools to the basic setup is simple.

The GridPP storage group continue to work on the deployment of dCache within LCG. Evidence of this can be seen in the storage groups page in CERN savannah [3]. The dCache documentation being provided by GridPP is being fed back into dCache's own documentation [4].

4.2 DPM

The Disk Pool Manager (DPM) is a lightweight solution for disk storage management. Its main attribute is that it offers the required SRM interface, without being complicated by other modes of access or complications such as tape storage systems. It has been developed at CERN. GridPP documentation on the installation and administration of a DPM system can be found in the GridPP wiki [11], which also points to other sources of information.

Due to the relative youth of the project, DPM has not seen the same levels of deployment and use as dCache. One example of where DPM has been used successfully was during SC3, where it was deployed at a NorduGrid site [12]. As Tier-2 sites begin their migration from Classic SE to SRM enabled storage, DPM is likely to see significant levels of deployment.

Being developed by CERN, DPM was integrated much more closely with LCG YAIM than dCache initially was. Installation of a basic system (head node and pool node) can be carried out in less than a day. There are relatively few configuration options with DPM, and administration essentially involves checking that the relevant daemons are running.

The GridPP storage group continue to work on the deployment of DPM within LCG. Evidence of this can be seen in the storage groups page in CERN savannah [3].

It should be noted that the DPM developers have created a tool that will migrate the information off of a Classic SE box and into a DPM storage element [13].

4.3 Which SRM should I choose?

As part of the strategy, the GridPP Storage Community will endeavour to support both the DESY/FNAL dCache and the LCG DPM. Both provide disk pool management with an SRM interface.

4.3.1 Basic Summary

Typically, sites should use the following rule when deciding whether to deploy dCache or DPM as their SRM:

dCache was built to support large sites, DPM was built to support smaller sites.

To help qualify this statement, sites should consider the criteria in Sections 4.3.2 and 4.3.3.

4.3.2 What is a small site?

According to LCG's documentation (July 2004) a small site satisfies all of the following:

- Provides 1-10TB of storage, usually system-attached to nodes.
 - Note that this is probably a measure that should be applied to available storage now (September 2005) - in the LHC era even Tier-2s will offer many 10s or 100s of TB. DPM has no intrinsic limit to the amount of disk it can manage.
- No SAN architecture.
- No Grid enabled tape storage.
- No full-time support for storage solutions. Only a fraction of an FTE available to manage the system.

In particular, DPM is intended as a replacement for the “Classic SE”.

4.3.3 How does a Tier-2 choose?

1. Decide whether the site matches the ‘small’ criteria above.
2. Decide whether the site should run dCache, or if it is better for the site to run DPM (as opposed to “I just need an SRM”).
3. Study the documentation for your preferred option.
4. Think about how much manpower/time to learn about the software and maintain it (see Sections 4.1 and 4.2).
5. If in doubt, contact the support group, join the mailing list and read the archives. This will provide the opportunity to speak to other sites who have already deployed an SRM.

See the Grid Storage [1] page for details.

4.3.4 When you have chosen

Since it is unlikely that the storage group will be able to support 20 new sites simultaneously, it is typical that each Tier-2 site and the storage group would agree with GridPP deployment [14] as to when SRM installation should begin. However, the storage group does not discourage individual sites from experimenting with deployment of an SRM whenever they find themselves in a position to do as such. It is *highly* recommended that sites deploying an SRM follow these steps:

1. Contact `storage@gridpp.ac.uk` to inform the GridPP Storage group of your interest.
2. Join the mailing list.
3. Read the documentation and start installing one.
4. Do not hesitate to ask for help when you need it.
5. Attend the phone conferences.

5 Tier-2 resources

Table 2 shows the storage requirements for each site, as taken from each Tier-2 memorandum of understanding (MoU). The table also shows the percentage of these resources that each site has delivered up to the second quarter 2005. Although these figures show that many Tier-2's have met their capacity requirements, it is clear that this capacity is not currently deployed on the Grid, either in the form of a Classic SE or with an SRM front end. This can be seen by comparing the figures in Table 2 to those in the GridPP storage monitoring page (Appendix B, [15]) and the storage information provided by GStat [16]. Meeting the MoU capacity requirements is *insufficient* for sites to participate in SC4 or the production Grid. Sites *must* migrate their storage from having a Classic SE front- end to having an SRM interface.

	MoU				Delivered		Ratio
	2004	2005	2006	2007	1Q05	2Q05	
Brunel	1.0	21.0	21.0	21.0	0.1	0.4	40%
Imperial	16.3	28.3	93.3	93.3	0.4	3.0	18%
QMUL	28.5	28.5	58.5	58.5	25.0	0.1	0%
RHUL	13.2	23.2	23.2	23.2	5.6	7.7	58%
UCL	0.7	0.7	0.7	0.7	0.8	2.0	278%
Lancaster	86.7	86.7	86.7	86.7	2.0	34.0	39%
Liverpool	80.3	80.3	80.3	80.3	0.0	0.0	0%
Manchester	372.6	372.6	372.6	372.6	8.8	12.0	3%
Sheffield	3.0	3.0	3.0	3.0	3.0	2.0	67%
Durham	5.0	5.0	5.0	5.0	5.0	2.0	40%
Edinburgh	70.5	70.5	70.5	70.5	1.0	19.0	27%
Glasgow	14.8	14.8	14.8	14.8	2.5	1.8	12%
Birmingham	9.3	9.3	9.3	9.3	9.0	9.3	100%
Bristol	1.9	1.9	1.9	1.9	1.9	1.9	99%
Cambridge	4.4	4.4	4.4	4.4	3.5	4.4	101%
Oxford	13.5	18.5	24.5	24.5	18.5	18.5	137%
RAL-PPD	5.8	11.8	17.4	24.4	6.0	5.8	101%
London	59.7	101.7	196.7	196.7	31.9	13.2	22%
NorthGrid	542.6	542.6	542.6	542.6	13.8	48.0	9%
ScotGrid	90.3	90.3	90.3	90.3	8.5	22.8	25%
SouthGrid	34.9	45.9	57.5	64.5	38.9	39.9	114%
Total	727.5	780.5	887.1	894.1	93.1	123.9	17%

Table 2: MoU storage figures. Capacity is in TB. Taken from Steve Lloyd report, September 2005.

6 Plan for support

The ability of the GridPP storage group to support Tier-2 sites in deploying and managing an SRM interface to their storage resources depends upon a number of issues:

1. Current available storage resources (see Table 2).
2. Planned procurement of storage resources.
3. Available manpower (number of FTE's per Tier-2 site).

It is our plan to stagger the support of the Tier-2's to allow for a controlled roll-out of SRM across the UK. This model is suitable for the following reasons:

1. Table 2 shows that some sites have already delivered the required storage capacity BUT have yet to deploy an SRM. Other sites have yet to begin procurement. For those sites with available hardware resources, SRM deployment can begin, assuming they have sufficient manpower to facilitate the deployment.
2. A phased rollout will give the GridPP storage community controlled exposure to the support of Tier-2 sites who have little or no knowledge of SRM. This will allow the community to build up a set of best practices for the large scale deployment at all Tier-2's.

6.1 Timetable

This is the provisional order in which we will support those Tier-2 sites who do not already have some form of SRM in operation. As part of the planning for Service Challenge 4 (begins **1st April 2006**) sites are currently being asked to provide information on when they will be able to begin the process of deploying an SRM [17]. They are also being asked to specify which SRM solution they will use.

Table 3 shows which sites have already deployed an SRM. However, it must be remembered that further work will have to be carried out at these sites as updates to the current software are released which include new and required functionality (i.e. security fixes; new releases; updates to the Grid information system).

Tier-2 site	SRM implementation
Imperial	dCache
Lancaster	dCache
Manchester	dCache
Edinburgh	dCache + DPM
Glasgow	DPM
RAL PPD	dCache
Birmingham	DPM

Table 3: Tier-2 sites with a working SRM v1.1 interface (as of October 2005)

As is clear from Table 3, the majority of sites that have deployed an SRM have chosen dCache. This reflects the fact that these sites were the early adopters of SRM (during SC3) at which time dCache was the only viable solution. The situation has now changed as can be seen from the fact that Glasgow have a working DPM install. See also Section 4.2.

The current timetable for SRM rollout at the remaining Tier-2 sites is as follows:

1. October 2005
 - Sheffield
 - Liverpool
2. November 2005
 - Durham
 - Oxford
 - QMUL
3. December 2005/January 2006
 - Brunel
 - RHUL
 - Bristol
 - Cambridge
 - UCL (HEP and CCC)

This timetable is taken in part from the GridPP SC4 section of the wiki [17] and information that has been obtained directly from sites. The timetable is subject to change depending upon available resources (hardware and manpower) at each institute between October 2005 and January 2006.

Table 4 gives a list of responsible personnel at each of the Tier-2's who should join the GridPP storage mailing list. General levels of support will always be provided for all sites via this list and the weekly phone conferences. More specific support (phone calls/site visits) will be provided for the above sites during the months specified above.

Each site should have an operational SRM by **31st January 2006** to allow for integration testing to be completed before the start of SC4.

6.2 Monitoring

The status of SRM deployment for GridPP sites is being tracked via a storage monitoring page [15]. This displays a table showing the available and used space at each site and whether sites are advertising an SRM interface via the various LCG BDII's. See Appendix B.

7 Conclusion

The goal of the GridPP storage group is to support Tier-2 sites in the deployment of an SRM interface to their storage resources. This must be done within the time frame of the LCG service challenges such that a production grid service is available when the LHC begins recording data in 2007. This document has provided information that Tier-2 sites should use when planning their SRM deployment and has detailed how GridPP will support them in this effort.

A Tier-2 personnel

Table 4 presents a list of contacts at GridPP Tier-2 sites who deal with local storage related issues. Any updates to this list should be sent to Greig Cowan (g.cowan@ed.ac.uk).

Tier-2	Contact	Email
Edinburgh	Greig Cowan	g.cowan@ed.ac.uk
Glasgow	Graeme Stewart	g.stewart@physics.gla.ac.uk
Durham	Mark Nelson	mark.nelson@durham.ac.uk
Lancaster	Matt Doidge	matt.doidge@gmail.com
Liverpool	Paul Trepka/Michael George	pat@hep.ph.liv.ac.uk / mpg@hep.ph.liv.ac.uk
Manchester	Alessandra Forti	aforti@hep.man.ac.uk
Sheffield	Andy Beresford	a.j.beresford@sheffield.ac.uk
Birmingham	Yves Coppens	ycr@hep.ph.bham.ac.uk
Bristol	Yves Coppens/Pete Gronbech	sysadmin starting Oct
Cambridge	Santanu Das	santanu@hep.phy.cam.ac.uk
Oxford	Pete Gronbech	gronbech@physics.ox.ac.uk
RAL PPD	Chris Brew	c.a.j.brew@rl.ac.uk
Brunel	Henry Nebrensky	j.nebrensky@brunel.ac.uk
Imperial	Mona Aggarwal	m.aggarwal@imperial.ac.uk
QMUL	Alex Martin	a.j.martin@qmul.ac.uk
RHUL	Simon George	s.george@rhul.ac.uk
UCL-HEP	Gianfranco Sciacca/Ben Waugh	lcg-admin@hep.ucl.ac.uk
UCL-CCC:	William Hay/Alice Fage	is-lcg-support@ucl.ac.uk

Table 4: Tier-2 SRM personnel

B Storage monitoring

Location	Target Capacities (TB)		Current Capacities (TB)				Storage Type	FileSpace Type
	2005	2007	Capacity	Available	Used	Used (%)		
ScotGrid								
Durham	5	5	1.915	1.915	0	0	Classic SE	permanent
Edinburgh	70.5	70.5	30.337	17.251	13.086	43	dpm, Classic SE, Classic SE	volatile, volatile, permanent
Glasgow	14.8	14.8	4.671	3.204	1.467	31	?, dpm	permanent, permanent
ScotGrid Totals	90.3	90.3	36.923	22.37	14.553	39		
NorthGrid								
Lancaster	86.7	86.7	53.17	52.346	0.824	1	Classic SE, dcache	permanent, volatile
Liverpool	80.3	80.3	0.009	0.004	0.005	55	Classic SE	permanent
Manchester	372.6	372.6	0	0	0	0	-	-
Sheffield	3	3	2.315	2.154	0.161	6	-	permanent
NorthGrid Totals	542.6	542.6	55.494	54.504	0.99	1		
SouthGrid								
Birmingham	9.3	9.3	7.04	6.76	0.28	3	dpm	permanent
Bristol	1.9	1.9	0.182	0.182	0	0	Classic SE	permanent
Cambridge	4.4	4.4	2.005	1.691	0.314	15	?	permanent
Oxford	18.5	24.5	1.605	1.145	0.46	28	Classic SE	permanent
RAL PPD	11.8	24.4	0.227	0.057	0.17	74	?, Classic SE, dcache	permanent, permanent, permanent
SouthGrid Totals	45.9	64.5	11.059	9.835	1.224	11		
London								
Brunel University	21	21	0.437	0.43	0.007	1	Classic SE	volatile
Imperial College	28.3	93.3	6.309	6.29	0.019	0	?, dcache	permanent, volatile
Queen Mary University	28.5	58.5	0	0	0	0	-	-
Royal Holloway University	23.2	23.2	8.772	7.785	0.987	11	?	volatile
University College	0.7	0.7	1.072	0.903	0.169	15	?	volatile
London Totals	101.7	196.7	16.59	15.408	1.182	7		
GridPP Tier-2 Totals	780.5	894.1	120.066	102.117	17.949	14		
GridPP Tier-1 RAL	191	1100	136.248	71.661	64.587	47	dcache, dcache	permanent, permanent
GridPP Totals	971.5	1994.1	256.314	173.778	82.536	32		

Figure 1: Screenshot from the GridPP storage monitoring page [15] showing the current status of storage deployment at GridPP sites.

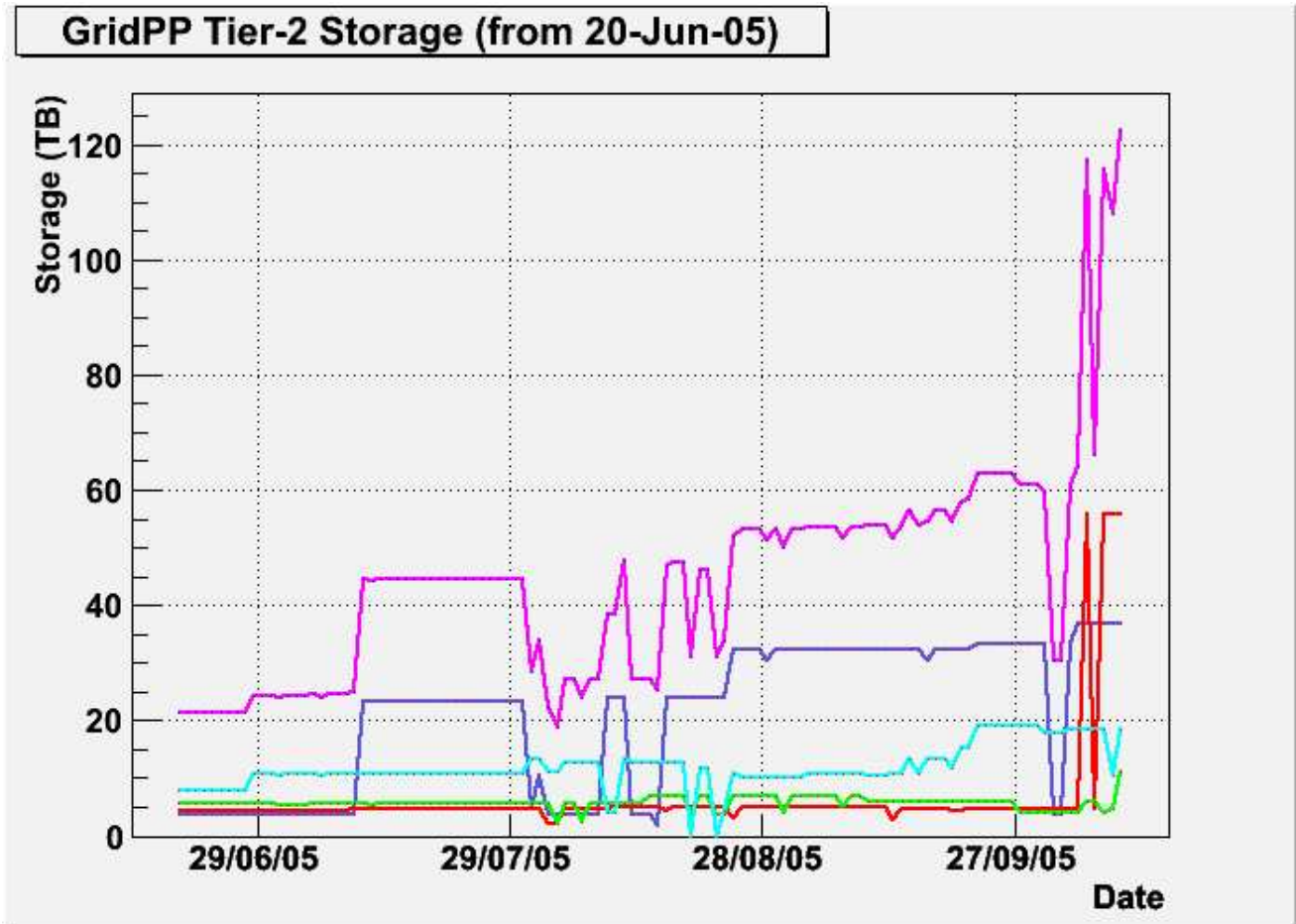


Figure 2: Screenshot from the GridPP storage monitoring page [15] showing the current status of storage deployment at GridPP Tier-2 centres. The key is below.

Magenta	GridPP Tier-2 Total
Blue	ScotGrid
Red	NorthGrid
Green	SouthGrid
Cyan	London

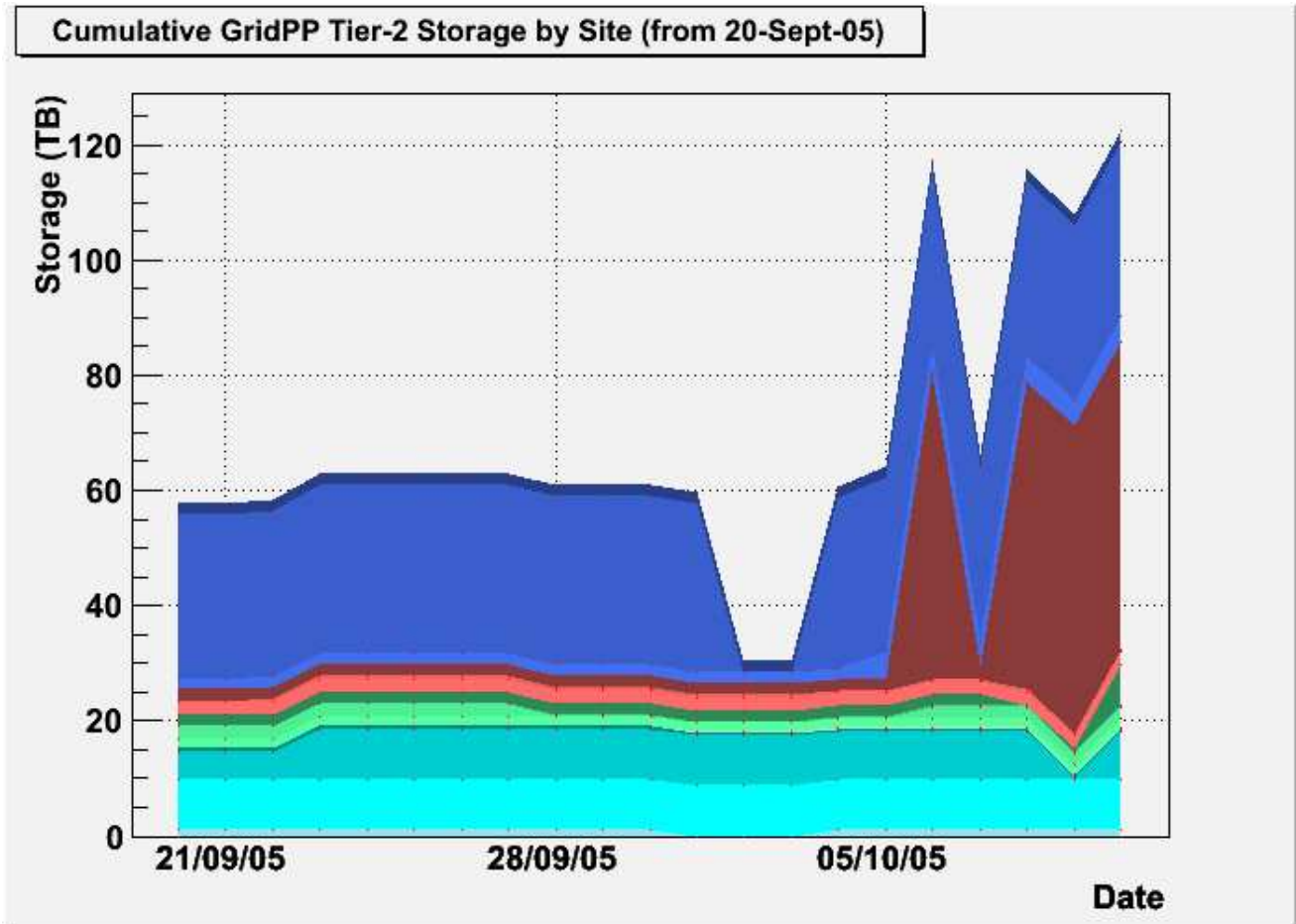


Figure 3: Screenshot from the GridPP storage monitoring page [15] showing the current status of storage deployment at GridPP Tier-2 sites and centres. The key is below, where the Durham is at the top of the plot and UCL is at the bottom. As of October 2005, Edinburgh and Lancaster are reporting the largest available storage capacity.

Royal Blue4	Durham
Royal Blue3	Edinburgh
Royal Blue2	Glasgow
Indian Red4	Lancaster
Indian Red3	Liverpool
Indian Red2	Manchester
Indian Red1	Sheffield
Sea Green4	Birmingham
Sea Green3	Bristol
Sea Green2	Cambridge
Sea Green1	Oxford
Dark Sea Green2	RAL-PPD
Cyan4	Brunel
Cyan3	Imperial College
Cyan2	Queen Mary London
Cyan1	Royal Holloway London
Cadet Blue2	University College London

References

- [1] GridPP storage group wiki
<http://wiki.gridpp.ac.uk/wiki/Grid.Storage>
- [2] GridPP storage group mailing list
<http://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=gridpp-storage&O=A&T=0>
- [3] GridPP storage group action/bug tracker
<http://savannah.cern.ch/bugs/?group=srmsupportuk>
- [4] Official dCache documentation
<http://www.dcache.org/manuals/Book>
dCache forum for users to exchange information
<http://www.dcache.org/maillinglists.shtml>
- [5] *Planning for SC4 and the initial LHC Service Setup*
<https://uimon.cern.ch/twiki/pub/LCG/TalksAndDocuments/SC4-planning-Aug29.pdf>
- [6] GridPP Service Challenge mailing list
<http://www.jiscmail.ac.uk/cgi-bin/webadmin?A0=gridpp-storage&O=A&T=0>
- [7] Technical defintion: SRM is a service orientated API supported over a SOAP web services protocol.
- [8] SRM collaboration project page
<http://sdm.lbl.gov/srm-wg/>
- [9] GGF workin group on SRM
<https://forge.gridforum.org/projects/gsm-wg/>
- [10] GridPP developed dCache documentation
<http://wiki.gridpp.ac.uk/wiki/Grid.Storage/DCache>
- [11] GridPP developed DPM documentation
<http://wiki.gridpp.ac.uk/wiki/Grid.Storage/Disk.Pool.Manager>
- [12] Grid for development, maintenance and support of the Advance Resource Connector Grid middleware
<http://www.nordugrid.org/>
- [13] CVS repository containing the source code for the Classic SE to DPM migration tool.
<http://isscvcs.cern.ch:8180/cgi-bin/cvsweb.cgi/migration-classicSE-DPM/?cvsroot=lcgware&hideattic=0>
- [14] GridPP information about deployment and use of LCG/EGEE middleware
<http://www.gridpp.ac.uk/deployment/>
- [15] GridPP storage monitoring page
<http://www.gridpp.ac.uk/storage/status/gridppDiscStatus.html>
- [16] Current status of LCG sites
<http://goc.grid.sinica.edu.tw/gstat>
- [17] Deployment status of SC4 sites
http://wiki.gridpp.ac.uk/wiki/Service_Challenge_4_Site_Status