

A Grid for Particle Physics - Managing the Unmanageable

Dave Britton, Imperial College London

Tony Doyle, University of Glasgow

Steve Lloyd, Queen Mary, University of London.

(On behalf of the GridPP Collaboration)

Abstract

GridPP has recently been extended to become a £33m, 6 year project, funded by PPARC. It aims to establish a Grid for UK Particle Physics in time for the turn on of the CERN Large Hadron Collider (LHC) in 2007. GridPP involves 19 UK Universities plus CCLRC and CERN and is embedded in a complex environment involving many other large organisations such as the CERN LHC Computing Grid (LCG) Project, the European DataGrid/EGEE, US Grid Projects and the large international particle physics experiments. Now, at the end of the first phase of the project, "From Web to Grid", we review the management issues and achievements of GridPP in establishing a prototype Grid and look forward to the second phase, "From Prototype to Production", that will provide the robust, secure, large scale Grid required for the analysis of LHC data.

Overview

The first phase of the project, GridPP1, will be complete in September 2004. It focused on creating a prototype Grid involving four main areas: support for the CERN LHC Computing Grid (LCG); middleware development as part of the European DataGrid (EDG); the development of particle physics applications for the LHC and US experiments; and the construction of infrastructure in the UK. The latter consists of a UK Tier-1 Regional Centre for LHC, which also serves as a Tier-A Centre for the BaBar experiment, and four regional Tier-2 Centres to harness the significant resources available at the UK Universities. Although the Tier-1/A Centre is wholly under GridPP's control, in all the other areas GridPP is a small part of a much larger enterprise as shown schematically in Figure 1.

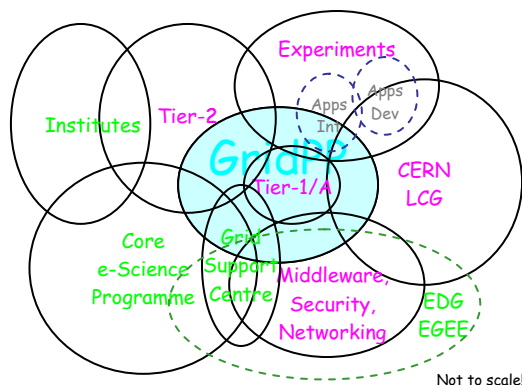


Figure 1: A schematic view of the environment in which GridPP operates.

In addition, all personnel funded through GridPP are based at CERN, the universities or CCLRC, and GridPP does not have direct line management of the post holders. These factors pose distinctive management challenges which we believe have been successfully overcome.

GridPP Management Structure

The GridPP Collaboration Board, consisting of the group leaders from each of the GridPP institutes, provides the high level overview and acts as a 'governing body'. The Project Leader is responsible for the day to day running of the project with the help of the Project Manager who is responsible for financial control over the project and for the tracking of deliverables and milestones. In GridPP1, the Technical Board is responsible for middleware development and testbed rollout and the Experiments Board is responsible for application development and computing resource requests from the experiments. Other bodies are convened as necessary such as the Hardware Advisory Group that recommends Tier-1/A hardware purchases. The Project Management Board (PMB), comprised of the Project Leader and Manager and the chairs of the other boards plus other specialists such as the Dissemination Officer and the UK EDG/EGEE Leader, meets weekly by video-conference and monthly face to face to oversee the project. In the second year of the project, 2002-2003, there were 47 meetings that generated 205 actions on PMB members.

GridPP Management Tools

In order to manage a project of this size and complexity and to be able to report its progress in a succinct manner to the PPARC Grid Steering Committee and others, a number of specialised management tools have been developed.

Figure 2 shows the **Project Map**, which provides a high level overview of the whole project with the ability to drill down to the details of any area.



Figure 2: Project Map showing project areas and elements.

The project map initially defined 233 tasks across the whole project of which 145 (62%) had been successfully completed by January 2004, including all of the EDG tasks. Because GridPP is a development project the definition of the tasks needs to be allowed to evolve with time and this is accomplished using a 'Change Form' that documents why a deliverable needs to be modified or in rare cases deleted. In addition to the tasks there are metrics, such as the number of papers produced or workshops held, mostly in the areas of Interoperability, Dissemination and Resources. All but one of a total of 44 such metrics were within specification in January 2004.

Embedded within the Project Map, the second tool that has been developed is the **Risk Register** shown in Figure 3.

ID	Name	GridPP		LCG		EDG		Apps		Infrastr.		Interop.	
		L1	Risk	L1	Risk	L1	Risk	L1	Risk	L1	Risk	L1	Risk
R1	Recruitment/retention difficulties			1	2	1	2	1	2	1	2	1	2
R2	Sudden loss of key staff			1	4	1	3	3	1	3	3	1	3
R3	Minimal Contingency					1	2	2	2	4	1	2	2
R4	GridPP Deliverables late					1	3	3	4	2	3	1	3
R5	Sub-components not delivered to project	2	4	2	3	2	3	2	3	2	3	2	3
R6	Non take-up of project results			2	1	2	1	4	1	4	1	4	1
R7	Change in project scope			1	1	1	2	3	4	1	1	1	1
R8	Bad publicity	2	2	4	1	3	3	1	3	3	1	3	3
R9	No publicity	2	1	2	1	2	1	2	1	2	1	2	1
R10	External software dependence	3	3	2	4	3	1	4	2	3	2	3	2
R11	Lack of monitoring of staff			1	2	2	2	4	1	3	3	1	2
R12	Withdrawal of an experiment	1	4	4	1	4	1	2	2	1	3	3	1
R13	Lack of cooperation between Tier centres			1	4	4	1	1	2	2	1	2	2
R14	Scalability problems			3	3	2	3	2	3	2	1	3	3
R15	Software maintainability problems			3	2	4	1	3	3	1	3	3	3
R16	Technology shifts	1	3	3	1	4	1	3	3	1	3	3	3
R17	Replication of research	3	2	2	4	1	3	3	1	3	3	1	3
R18	Lack of funding to meet LCG PH-1 goals			1	3	3	1	3	3	1	3	3	3
R19	Adequate persistency solution not ready							4	1	4	2	2	4
R20	Conflicting software requirements							1	3	3	1	3	3
R21	Tier-A hardware fails to meet requirements							1	3	3	1	3	3
R22	Other Hardware fails to meet requirements							1	3	3	1	3	3
R23	Hardware physical risk (large scale)							2	2	2	4	1	4
R24	Hardware physical risk (small scale)							2	2	2	4	1	4
R25	Hardware procurement problems							2	2	2	4	1	4
R26	LAN Bottlenecks							2	1	2	2	2	4
R27	Tier-2 organisation fails							2	1	2	2	2	4
R28	Tier-2 hardware not used as planned							2	1	2	2	2	4
R29	SYSMAN effort inadequate							2	1	2	2	2	4
R30	Firewalls interfere with Grid							1	3	3	1	3	3
R31	Inability to establish trust relationships							2	2	2	4	1	4
R32	Security inadequate to operate Grid							1	3	3	1	3	3
R33	GGF does not establish standards							1	3	3	1	3	3
R34	Minimal open source code development							2	2	2	4	1	4
R35	Failure of international cooperation							1	4	4	1	4	4
R36	Science and GridPP divergence							1	4	4	1	4	4
R37	Institutes do not embrace Grid	1	3	3	1	4	4	1	4	4	1	4	4
R38	Grid is not stable enough for use	3	3	3	1	4	4	1	4	4	1	4	4
R39	Delay of the LHC	2	2	4	1	2	2	4	1	2	2	4	1
R40	Lack of future funding	1	2	4	1	2	2	4	1	2	2	4	1
R41	Network backbone failure							1	4	4	1	4	4
R42	Network backbone bottleneck							1	2	2	4	1	4

Figure 3: The GridPP Risk Register.

There are currently 42 categories of risk that apply across some or all areas of the project. For each relevant area the risk is given a likelihood and an impact rating, both on a scale of 1 to 4. These are multiplied together to give an overall

risk and colour coded on the register to immediately highlight potential problems. The ratings are reviewed 6-monthly by the PMB as the project evolves. New risks, such as Red Hat licensing problems, are added when identified.

Evolution to GridPP2

In the second phase of the project, GridPP2, running from September 2004 to 2007, the focus will be on dramatically increasing the scale of the UK particle physics Grid, and improving its functionality and robustness. The management structure will evolve to reflect this change of emphasis and to take into account the move from EDG to EGEE and the increasing overlap with the Core e-Science programme through, for instance, the Grid Operations Centre.

In order to move to a 'production' environment, the Technical Board, which focused on middleware development, will evolve into the Deployment Board. There will be a dedicated production team led by a Production Manager who has already been appointed. The Experiments Board will become the User Board ensuring that user requirements are prioritised. A Tier-2 Board will ensure that the Tier-2 Centres are fully engaged in GridPP2. A full time Dissemination Officer has also been appointed who will work closely with the training and outreach programme organised by the National e-Science Centre.

Following approval of the GridPP2 proposal a detailed specification of all posts was drawn up and bids were invited from the institutes. The bids were evaluated by a Peer Review Selection Committee made up of members of the PMB and external experts. The detailed specifications and allocations will form the basis of a new Project Map shown in Figure 4 and a new Risk Register will be drawn up.

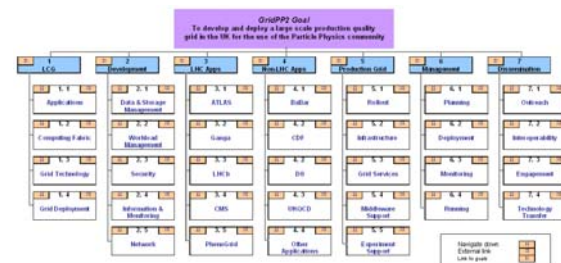


Figure 4: Preliminary GridPP2 Project Map.