

A Grid for Particle Physics - Managing the Unmanageable

D.Britton, Imperial College London, A.T.Doyle, University of Glasgow,
S.L.Lloyd, Queen Mary, University of London.
(On behalf of the GridPP Collaboration)

Abstract

GridPP is a £33m, 6 year project that aims to establish a Grid for UK Particle Physics in time for the turn on of the CERN LHC in 2007. GridPP is embedded in a complex environment involving many other large organisations. Now, at the end of the first phase of the project we review the management issues of GridPP associated with establishing a prototype Grid that will provide the basis for the robust, secure, large scale Grid required for the analysis of LHC data.

1. Challenges

The GridPP project has various unusual characteristics that have posed a number of management challenges. The cutting-edge and complex nature of the work means that it has been hard to define many details of the work very far in advance. This problem has been compounded by the fact that GridPP is a small part of a much larger international context with which it has been necessary to maintain alignment. Finally, as an academic-based project distributed across 20 institutions, there are significant issues concerning line-management and culture. Various novel tools and techniques have been developed to manage these challenges.

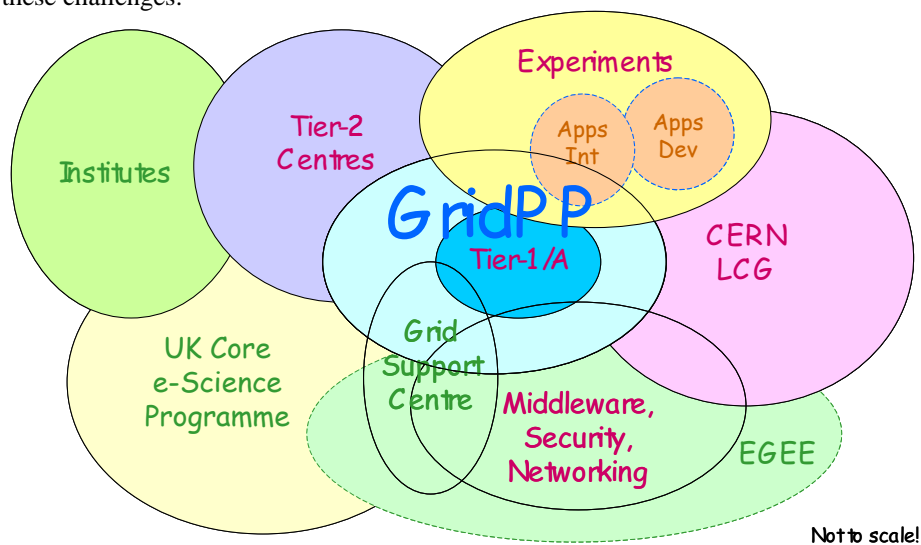


Figure 1: A schematic view of the environment in which GridPP operates.

2. Management Structure

The central management body in GridPP is a small executive Project Management Board (PMB). The Collaboration Board (CB) is the governing body and serves as a means of strategic consultation and communication at a senior level within the community. The Technical Board (TB) coordinates the middleware development and the deployment of various testbeds. The Experiments Board (EB) focuses on the development of Grid interfaces for the HEP experiments, promoting common solutions, and schedules the allocation of hardware resources. Management roles within the project are mainly taken by particle physicists from organisations in the collaboration.

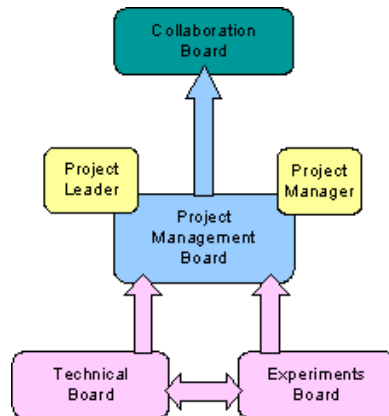


Figure 2: A schematic view of the GridPP management structure.

3. Management Tools

The complicated international context in which GridPP is embedded and the complex nature of the work programme lead to a project definition that is “incomplete”, in that GridPP does not produce, or even sub-contract, all the components needed to produce the final product. Instead, it relies on developments in other related projects. The project definition is “dynamic”, in that there are frequent changes at all levels (including entire paradigm shifts such as OGSA, web-services, and WSRF), and to some extent the project is “devolved” because much of the work done by people employed by GridPP is in the context of other projects (such as EDG, EGEE, LCG, and the experiment applications). The distributed academic setting poses an additional challenge to the formal management of the project. GridPP was fortunate to benefit at an early stage from guidance from an Oversight Committee that lead us to develop unique project management tools to match these challenges.

3.1. Project Map

The main tool developed is called the ProjectMap (see Figure 3) that provides a means of defining a complex project; a means of presenting the status of the project at any moment at a reasonably high level; and a record of the project achievements linked to associated documentation. The tool is based on an Excel spreadsheet with a number of custom macros written in Visual Basic that provide functionality, in particular in the management of the many hundred hyperlinks used to provide navigation.

The top level of the ProjectMap provides a simple overview of the project elements but clicking on any of these brings up a more detailed view where each element is now divided up into a sequence of numbered milestones or metrics. The metrics are simply attributes that may change throughout the life of the project (such as the number of unfilled posts); the milestones are defined by a description, a due date, and a metric (a quantitative way of measuring success). The details of the milestones and metrics are contained at the bottom level of the ProjectMap and are displayed by clicking on the numbered boxes. When a milestone is completed, then a link to supporting documentation, normally an external URL, is added to the bottom-level of the ProjectMap.

The button labelled “Update” runs a macro that examines the current status date (circled in purple in the detailed view shown in Figure 3) and the status of each of the project milestones and metrics. If the milestone has been met, it is coloured green; if it is overdue then it appears red; if it is due within a specifiable period (in this example, within 60 days of the status date), then it appears yellow. The metrics have a binary colour scheme: blue if they are within the specification; purple if they are not.

At a glance, the ProjectMap shows the status of the project and automatically flags areas of concern. The first two columns labelled CERN and DataGrid, actually monitor deliverables from the related LCG and EDG projects and address the “incomplete” nature of the GridPP project: Success of these related projects is essential to the success of GridPP.

The “dynamic” nature of the project is handled by a mechanism of Change Forms that allow the managers of the individual elements (a box or group of boxes containing milestones in Figure 3) to make changes to the deliverables during the project. In many cases, these changes are due to changes outside of GridPP, such as a slippage of dates, a change in functionality, and even the wholesale abandonment of a particular

technology. The Change Form asks for an assessment of the consequences of the change and any modification to the scope of the work being attempted. Change forms are submitted to the Project Manager who implements the request or seeks further information or consultation if necessary. This process is assisted by the identification of risks and dependencies in each of the milestone definitions at the bottom level of the ProjectMap. It is clear that a more traditional resource-driven work breakdown schedule would be hard to maintain in this environment. In practice, the level of effort available is pre-defined, its location is pre-determined to a particular institute and area, and is not a management variable.

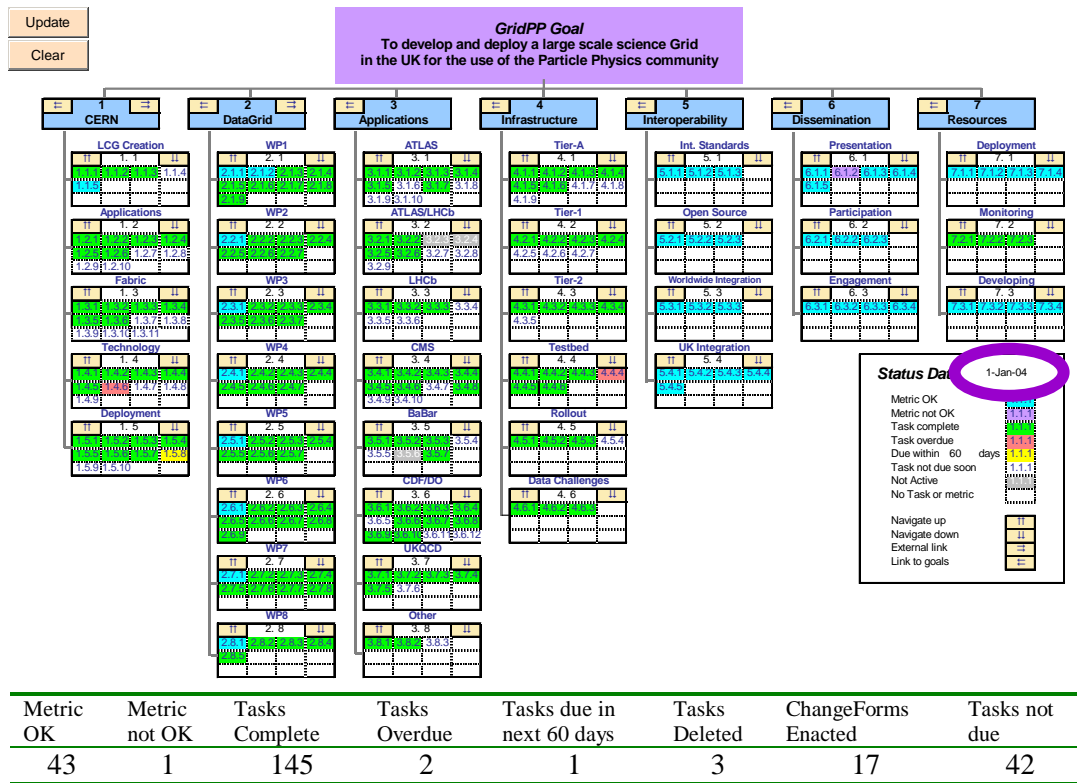


Figure 3: The GridPP ProjectMap.

3.2. Risk Register

In addition to specific risks associated with individual milestones, the management process considers a more global risk assessment. The tool developed here, referred to as the RiskRegister, is similar to the ProjectMap and, for convenience, is part of the same Excel Workbook. The RiskRegister, shown in Figure 4, identifies 42 risks (labelled in the left hand column), which apply to one or more of the project areas labelled in the first row (the GridPP label implies the risk applies to the project as a whole rather than a specific area). The RiskRegister is a sparse matrix of risk, defined in terms of a Likelihood (a number between 1 and 4) and an Impact (between 1 and 4) with the overall Risk being the product of these two numbers. The overall risk is coloured automatically according to the following scheme: High risks (9–16) are red; medium risks (5–8) are orange; low risks (0–4) are green. The risk numbers need not be integer. Clicking on any of the coloured boxes in the RiskRegister brings up the risk assessment form that provides a detailed description; current and future options for managing the risk; and a crisis strategy. The risks are re-assessed on a six monthly basis and new risks may be added. The RiskRegister provides, at a glance, guidance as to any current areas of concern.

3.3. Project Plan

The GridPP ProjectPlan is a detailed picture of the project resources (financial and effort). The heart of this spreadsheet is a table that contains the effort expected from each post for each quarter (3 month period) of the project. This is calculated automatically by entering a start-date, a duration, and the fraction of full time work, for each post. This effort is then translated to cost using a look up from a table of salary scales. The central table also contains the non-staff costs, providing a complete picture of the financial

planning. This table is maintained as the project progresses and shows the actual spend to date. Each line in the table can be attributed a number of attributes (a category; a sub-category; affiliation to an institute and an experiment, etc) and summary tables make extensive use of the Conditional Sum utility in Excel to extract different financial views of the project. Thus, by maintaining this central table the project can be viewed in terms of resources allocated to different areas, different institutions, different experiments, and so on, without the need to re-generate these views. This automation reduces significantly the amount of cross-checking required.

ID	Name	GridPP			LCG			EDG			Apps			Infrast.			Interop.		
		Li	Im	Risk	Li	Im	Risk	Li	Im	Risk	Li	Im	Risk	Li	Im	Risk	Li	Im	Risk
R1	Recruitment/retention difficulties				1	2	2	2	2	4	2	2	4	2	2	4			
R2	Sudden loss of key staff				1	4	4	1	3	3	1	3	3	1	3	3			
R3	Minimal Contingency							1	2	2	2	2	4	1	2	2			
R4	GridPP deliverables late							1	3	3	4	2	8						
R5	Sub-components not delivered to project	2	4	8	2	3	6	2	3	6				2	3	6			
R6	Non take-up of project results				2	1	2	1	4	4	1	4	4						
R7	Change in project scope				1	1	1	2	2	4	1	1	1						
R8	Bad publicity	2	2	4	1	3	3	1	3	3									
R9	No publicity	2	1	2															
R10	External software dependence	3	3	9	2	4	8	1	4	4	2	3	6						
R11	Lack of monitoring of staff				1	2	2	2	2	4	1	3	3	1	2	2			
R12	Withdrawal of an experiment	1	4	4	1	4	4	1	2	2									
R13	Lack of cooperation between Tier centres				1	4	4							1	2	2			
R14	Scalability problems				3	3	9	2	3	6				1	3	3			
R15	Software maintainability problems				3	2	6	1	3	3									
R16	Technology shifts	1	3	3	1	4	4	1	3	3									
R17	Repetition of research	3	2	6															
R18	Lack of funding to meet LCG PH-1 goals				1	3	3												
R19	Adequate persistency solution not ready										1	3	3						
R20	Conflicting software requirements										4	1	4	2	2	4			
R21	Tier-A hardware fails to meet requirements										1	3	3						
R22	Other Hardware fails to meet requirements													1	1	1			
R23	Hardware physical risk (large scale)													1	4	4			
R24	Hardware physical risk (small scale)													2	2	4			
R25	Hardware procurement problems													2	2	4			
R26	LAN Bottlenecks													2	1	2			
R27	Tier-2 organisation fails													2	2	4			
R28	Tier-2 hardware not used as planned													2	1	2			
R29	SYSMAN effort inadequate													2	3	6			
R30	Firewalls interfere with Grid													1	3	3			
R31	Inability to establish trust relationships													2	2	4			
R32	Security inadequate to operate Grid													1	3	3			
R33	GGF does not establish standards																1	3	3
R34	Minimal open source code development																2	2	4
R35	Failure of international cooperation				1	4	4										1	4	4
R36	e-Science and GridPP divergence																2	2	4
R37	Institutes do not embrace Grid	1	3	3															
R38	Grid is not stable enough for use	3	3	9															
R39	Delay of the LHC	2	2	4															
R40	Lack of future funding	1	4	4				2	2	4									
R41	Network backbone failure													0	4	0.4			
R42	Network backbone bottleneck													1	2	2			
Alt-ir																			

Figure 4: The GridPP Risk Register.

3.4. Reporting Mechanisms

Following the lead of the EDG project, GridPP requests a quarterly report from each area of the project that details the effort actually provided and the status of the deliverables. This information on deliverables is used to update the ProjectMap and the information on effort is entered into the ProjectPlan with the help of various macros. The delivered effort is then compared with the expected effort and discrepancies are investigated by the Project Manager. Frequently, this signals the inevitable delays in re-filling a post but may also signal other problems that need attention. The EDG system was followed so that the significant number of GridPP people working on the EDG project did not have to report twice using completely different mechanisms. This was part of the process for managing the “devolved” part of the project where GridPP personnel were working on EDG deliverables. Quarterly reports were not required from the CERN part of the project: although GridPP funded 25 positions here, the management of these posts was devolved to the LCG project.