The LHC Computing Grid Project

Major Risks

POB – CERN – 3 June 2004

Les Robertson – LCG Project Leader
CERN – European Organization for Nuclear Research
Geneva, Switzerland
les.robertson@cern.ch

last update: 01/06/2004 11:44

- This session is concerned with RISK and the process for managing risk
- I will talk about the process for mitigating the major risks and what we would do in the event of a crisis

- This will necessarily have a "negative" flavour
- In some cases there is also a positive strategy for bypassing the problem –
  but these are not mentioned as from a risk point of view they may be considered optimistic.

Risk Register

Likelihood	*Impact*
1 - never expected to happen	1 - we can deal with it, no problem
2 - could happen but very unlikely	2 - a bit of a hassle but not too bad
3 - could well happen at some point	3 - clearly can be dealt with, but with significant effort
4 - will probably happen	4 - crisis

Risk factor = *likelihood* x *impact*

Low –	**1-5**
Medium –	**6-8**
High –	**9-12**
Unacceptable -	**>12**

Update on the Major Risks identified in August 2003

	risk factor
R14 Inadequate or late 3rd party software	**9**
Performance shortfalls:	
R19 Reliability problems	**9**
R20 Scalability problems	**9**
R21 Grid Middleware/Infrastructure failure	**9**
R28 Site security requirements too restrictive	**12**
R31 Inadequate power and hvac arrangements	**9**

Risk Number: R14-LCG **Owner: LMR**	
Risk Name: Inadequate or late 3rd party software	
Description of Risk: Late or non-delivery of software packages provided by groups external to the LCG project—e.g. EDG components or software from the Globus and Condor project.	
Risk Likelihood: 3	**3**
Risk Impact: 3	**3**
Overall Risk: 9	**9**
Current Process for Managing Risk: Avoid dependence on promised new products or functionality by using packages that already exist. Careful selection of external products from suppliers with apparent long term viability. Continued contact and negotiation with those suppliers.	
Future Options for Managing Risk: Direct participation, where possible, in the management or oversight of key suppliers.	
Crisis Strategy: Discussion of the situation by the SC2. Review impact on overall schedule, availability of alternative products or strategies to alleviate impact (especially in case of late delivery). Consider reassignment of project staff to provide temporary or permanent replacement package.	
Actions Required:	

The nature of this risk has changed as we have moved into a production mode – for the middleware package, the testing and certification process and investment in debugging and error correction (LCG collaborators and NSF funding) has produced an acceptable (for now) level of reliability with some confidence that this can be maintained. However, the functionality is not what was planned a year ago. In effect we have now moved into the process defined to mitigate this risk – only use products that have been demonstrated to exist and for which we can see the support model. The implication is that we have to adapt the requirements to the tools that we can be sure will be available. The EGEE *fast-prototyping* + *ARDA focus* will help us to see what is coming – but we shall not plan to use their products until we can touch convincing pilot versions.

Action required: In the light of the data challenges agree a prioritised plan for providing essential improvements in functionality/performance.

Risk Number: R19-LCG **Owner: LMR**	
Risk Name: Reliability Problems	
Description of Risk: Reliability of the grid operation is insufficient for efficient exploitation. This may arise due to reliability problems with middleware, or due to difficulties in the operational tools and processes.	
Risk Likelihood: 3	2
Risk Impact: 3	3
Overall Risk: 9	6
Current Process for Managing Risk: Middleware: Short term - building up middleware expertise in the grid deployment teams. Longer term - Participation in the EGEE middleware activity, negotiation with VDT as a supplier of US middleware. Operations: Building operations collaboration through the GDB, to enable reliability issues to be tackled at a management level.	
Future Options for Managing Risk:	
Crisis Strategy: Middleware: Escalate to the management of the teams responsible for supporting the middleware components; if solutions are not forthcoming reduce the functionality offered by the service (not so obvious once we are in operation). Operations: Escalate to Regional Centre Management through GDB.	
Actions Required:	

So far the mitigation process seems to have worked fairly well:
Middleware - GDA investment in testing/certification and in a team of expert systems programmers; participation in the European Globus support team; closer relations with the VDT team; support agreements with the authors of EDG components. The result is a level of reliability that is so far rather good compared with the expectation of this time last year.
Operation – Well-planned operations monitoring system implemented at RAL and deployed also at AS/Taipei places us in a good position to detect operations problems. However, these are very early days – so far problems are largely teething troubles. We have little/no experience of subtle operations problems to test our ability for diagnosis and collaborative response.

Risk Number: R20-LCG **Owner: LMR**	
Risk Name: Scalability problems	
Description of Risk: Solutions, programs or packages that work successfully in development and demonstration environments fail when deployed in the production LHC service.	
Risk Likelihood: 3	**3**
Risk Impact: 3	**3**
Overall Risk: 9	**9**
Current Process for Managing Risk: The LCG project will develop successively larger prototype services, both locally and in terms of an HEP grid, between now and the end of 2006. Within the limits of available funding, the scale on which solutions, programs and packages will be tested will be chosen to give reasonable confidence that they will function correctly in the full production service.	
Future Options for Managing Risk:	
Crisis Strategy: If scalability problems are encountered during LCG phase 1, review the situation in the SC2 committee and adapt the programme accordingly - e.g. data challenge programme, resource allocation. If scalability problems are encountered post-2007, review the situation with IT management, experiment experts and regional centre representatives, reassigning resources as necessary to address the failing component.	
Actions Required:	

This remains a major risk for most components and systems. A programme is in place with Alice to test scalability of the components of data recording.

Actions required:
Now that the basic grid service has been deployed, a series of service challenges will be defined that progressively test scalability in key areas – network, mass storage, database, grid catalogues, grid scheduling issues.

Risk Number: R21-LCG **Owner: LMR**		
Risk Name: Grid Middleware/Infrastructure failure		
Description of Risk: The grid fails to deliver the software and infrastructure needed to operate a distributed analysis facility.		
Risk Likelihood: 3	**3**	
Risk Impact: 3	**3**	
Overall Risk: 9	**9**	
Current Process for Managing Risk: Early deployment of a grid using early versions of the applications tools and interconnecting regional centres running in a production mode, to identify what works, what does not work, and what the fundamental problems are. In mid-2004 an assessment will be made of the size of this risk. In parallel, an active evaluation programme of emerging grid middleware will be undertaken.		
Future Options for Managing Risk:		
Crisis Strategy: Work with SC2 to define an alternative low-functionality solution.		
Actions Required:		

The deployment has now started and the fundamental problems are beginning to emerge. While missing functionality of the "middleware" is easy to identify, we do not yet have enough experience to understand the operational and infrastructure issues. The assessment scheduled for mid-2004 will not be completed until later in the year when the full set of data challenge experience is available. This should result in a statement of what can be considered as the realistic expectation for LHC startup.

The EGEE project has a significant middleware activity, and is also investing heavily in grid operations in several sites in Europe. However, the complexity of the EGEE project and its 2 or possibly 4 year lifetime pose themselves significant risks.

Risk Number: R28-Deploy **Owner: IGB**	
Risk Name: Site security requirements too restrictive	
Description of Risk: The local security requirements at a Regional Centre do not allow the flexibility of access needed to operate a grid. Unrealistic requirements on registration, or access by users of certain nationalities will cause complex operational problems.	
Risk Likelihood: 3	2
Risk Impact: 4	3
Overall Risk: 12	6
Current Process for Managing Risk: Standing Security Group established by the GDB to understand and monitor the situation, and to define the requirements such that they can be understood and agreed to by Regional Centre management.	
Future Options for Managing Risk:	
Crisis Strategy: At the project level it will not be critical if some Tier 2+ Regional Centres are not able to participate in the LCG. The situation will be critical for Tier 1 Centres, where the strategy would be to immediately involve senior managers from CERN and the appropriate institutes in an attempt to find a solution. (Note that significant attention is paid to this issue by the Security Group at Tier 1 Centres). If such issues become very serious they must be escalated to the CERN directorate and the funding agencies.	
Actions Required:	

So far the situation is better than expected, and the Security Group has negotiated full access rights for all VO members. However, the grid has only recently become operational and may not yet be a major target of the hacking community. A serious security incident could trigger less flexible attitudes by the Regional Centre security officers.
A more general security risk is introduced and this specific risk downgraded. The mitigation strategy remains careful work by the Security Group. This must be reviewed each year in the light of experience. At present the experiments have de facto backup strategies in the form of a reversion to their previous operational models. If/when the grid becomes the established operation mode consideration will have to be given to the preparation of contingency plans.

Risk Number: R31-LCG **Owner: AC**	
Risk Name: Inadequate power and hvac arrangements	
Description of Risk: The planned level of power for the CERN computer centre is insufficient for the required computing capacity.	
Risk Likelihood: 3	2
Risk Impact: 3	3
Overall Risk: 9	6
Current Process for Managing Risk: Re-assessment of resource requirements completed in 2003. Monitoring of power characteristics of new processors, evaluation of power saving features.	
Future Options for Managing Risk: Upgrade from 2MW to 4MW (initial cost study concluded that this should not be done in a first phase); alternative location for a second computer installation (e.g. using a building with an existing power supply vacated after LHC installation is complete).	
Crisis Strategy: This will not come suddenly as a crisis, and there will be time to plan alternatives, BUT not without additional funding.	
Actions Required:	

The planning for the CERN computer centre is now to provide 2.5 MW maximum power. This is now a general concern in the industry and there is much discussion of technology changes and directions aimed at containing the power consumed by processors and systems. It is too soon to know what this will mean in practice, especially as our concern is power consumed per effective SI2000, not per system. However this is no longer seen as a critical risk.

No change in mitigation process or future options. Note that this is a risk, financial rather than technical, that we will see coming some years ahead

Revised list of Major Project Risks

		risk factor
R14	Inadequate or late 3rd party software	**9**
R20	Scalability problems	**9**
R21	Grid Middleware/Infrastructure failure	**9**
R42	Security problems which lead to shutdown of parts of the grid	**12**
R43	Fragmentation of the LCG into incompatible grids	**9**

May 2004

Risk Number: R42-Deploy **Owner: IGB**

Risk Name: **Security problems that lead to shutdown of sections of the grid**

Description of Risk:

Grid-level security incidents are not solved rapidly enough, occur with too high a frequency, or have very serious impact on regional centres – triggering major regional centres to withdraw from the grid on a medium or longer term basis.

Risk Likelihood: 3 **Risk Impact**: 4 **Overall Risk**: 12

Current Process for Managing Risk:

Standing Security Group established by the GDB to understand and monitor the situation, to define processes and policies to reduce the probability of major incidents, and to ensure good communication between site security officers. The LCG Security Group is very closely integrated with other security groups in the community.

Future Options for Managing Risk:

When the grid becomes the established operation mode for experiments consideration will have to be given to the preparation of contingency plans.

Crisis Strategy:

At the project level it will not be critical if some Tier 2+ Regional Centres are not able to participate in the LCG. The situation will be critical for Tier 1 Centres, where the strategy would be to immediately involve senior managers from CERN and the appropriate institutes in an attempt to find a solution. If such issues become very serious they must be escalated to the CERN directorate and the funding agencies.

Actions Required:

Issue of contingency plans to be addressed in the TDR

Risk Number: R43 - LCG **Owner: LMR**

Risk Name: Fragmentation of the LCG into many incompatible grids

Description of Risk: National, regional or other pressures lead to LHC Regional Centres being fragmented into many different incompatible grids, requiring each experiment to divert key resources to implement its own meta-grid. The probability of this will remain high so long as other sciences require only regional or national computing infrastructures, and if competing grid technologies with substantially different functionality appear.

Risk Likelihood: 3 **Risk Impact**: 3 **Overall Risk**: 9

Current Process for Managing Risk: At present there are two major grids involved in LHC – the EGEE group and the US (Grid3/OSG) group. A smaller group exists in the Nordic countries. The EGEE and US groups use very similar middleware. EGEE provides a force for compatibility within Europe and the non-European EGEE collaborators. A reliable and well established LCG/EGEE service will provide further impetus. Active discussion with US group, being formalised as a responsibility of the GDA management team. First meeting scheduled with the Nordic group.

Future Options for Managing Risk:
With time it is likely that standards will emerge enabling inter-working between grids.

Crisis Strategy:
Two grids will be a nuisance not a catastrophe. It will be a significant problem if there are many different grids, even within one country. It will be a major cost to experiments if the Tier-1 and major Tier-2 centres use incompatible middleware. This is something that will evolve slowly, and should be watched by the GDB and the POB.

Actions Required: