

DRAFT

CERN/LHCC 2003-065

LHCC-G-064

5 December 2003



LARGE HADRON COLLIDER COMMITTEE

LCG PROJECT COMPREHENSIVE REVIEW

November 2003

1. EXECUTIVE SUMMARY

The LHC Computing Grid (LCG) Project was created by CERN Council in September 2001 with the aim of prototyping and deploying the computing environment for the LHC experiments. The formal launch of the project was at a workshop held in March 2002. Since that time, the LCG Project has made very significant progress towards the realisation of the computing requirements of the experiments in time for LHC operation in April 2007.

It is realistic to expect the LCG Project to have prototyped, built and commissioned the initial LHC computing environment suitable for LHC operation starting in April 2007. The LHCC considers that the LCG Project has made an impressive start to a very complex task in a very short time. In particular, the LCG Project scope has been clearly defined, the first application software products have been delivered to the experiments, the first Grid software suite has been defined and deployed, the initial Grid production service has been put into operation and will get tested by the experiment Data Challenges in 2004 and plans for increased functionality of the Grid services are being currently defined. The organisational structure of the LCG Project has been put in place and is able to adapt itself to the changing needs of the project.

The first annual LHCC Comprehensive Review of the LCG Project took place on 24-25 November 2003. The LHCC referees addressed the following areas: Status Update, Management and Planning; Middleware and Interaction with EDG/EGEE and other Developments in Middleware; Fabric Area; Grid Deployment and Regional Centres; Applications Area. The LHCC acknowledges the considerable amount of work that has gone into the preparation of the LCG Project Comprehensive Review.

The conclusions and concerns of the LHCC are given below. They will help the Committee to follow up outstanding issues and to monitor future progress of this project in forthcoming sessions of the LHCC prior to the next LCG Project Comprehensive Review one year hence.

2. OVERVIEW

The LHCC noted significant progress in the LCG Project since its inception. The Committee took note of the new organisation structure put in place and that a significant majority of the LCG Project's milestones are being met on time. Although resources are continuing to be allocated for the most part of Phase I, **detailed discussions are required leading up to a Memorandum of Understanding for Phase II of the LCG Project.**

The LHCC took note of developments in middleware project focusing on progress in the EGEE, EDG, VDT and ARDA endeavours. The Committee identified several concerns, particularly **the risks associated with lack of product delivery and the difficulties in entering the analysis phase of the project. Further collaboration with the US partners and with all the Regional Centres is recommended.**

The LHCC has no major concerns regarding the Fabric Area. The Fabric Area is considered to be the least controversial component of the LCG Project.

Although delays in the deployment of the LCG-1 software package have been reported, several Regional Centres have now installed successfully the package and integration of the experiments into the LCG Project is in progress. Problems encountered with LCG-1 are being corrected in time for LCG-2 that is to be deployed by the end of 2003. The LHCC underlined that it is only the beginning of the LCG Project and the community is learning to handle the new Grid computing environment.

The LHCC noted the good progress in the Applications Area. Manpower levels are adequate, the scope of the project is properly defined and milestones are being met, including deployment of the first products, e.g. POOL, in the production environment of at least one of the LHC experiments. **The Committee underlined the importance of supporting the Monte Carlo generator codes for the experiments.**

3. LCG STATUS UPDATE, MANAGEMENT AND PLANNING

The LCG Project is a collaboration of the LHC experiments, the Regional Computing Centres and the physics institutes with the aim of preparing and deploying the computing environment that will be used by the LHC experiments to analyse the LHC data. The project includes support for applications and the development and operation of a computing service.

The LCG Project is divided into two phases. Phase I (2002-2005) has the objective of building a service prototype, based on existing Grid middleware, of running a production Grid service and producing the Technical Design Report for the final system. Phase II (2006-2008) will build and commission the initial LHC computing environment. The LCG is not a development project and it relies on other Grid projects for the middleware development and support.

The LCG Project is organised around the Project Overview Board (POB), the Project Execution Board (PEB) and the Software and Computing Committee (SC2). The PEB is responsible for the general coordination and manages the implementation of the system while the role of SC2 is to identify common solutions and to set requirements for the project through Requirements and Technical Assistance Groups (RTAGs). The LHCC took note of the evolution of the project organisation with the increased reality brought about by the system tests. In particular, the new organisation takes into account the shift of focus from the development to the deployment phases and the recognition and inclusion of the experiments in the subsequent phases of the LCG Project. The LHCC provides the scientific review of the LCG Project while the Computing Resource Review Board deliberates on matters concerning resources. However, **the LHCC expressed some reservation on the role and composition of the new SC2 and these issues must be re-visited.**

A new milestone tracking process has been introduced for the LCG Project. The SC2 monitors the Level-1 and Level-2 milestones and of the 188 milestones currently defined, about 90% of them due by October 2003 have been met. **The Level-1 milestones are currently under review by the LCG Project and will be presented to the LHCC for discussion in March 2004.**

The LHCC took note of the status of resources, both in terms of manpower and computing material. **Concern was expressed on the staffing and priorities in the Regional Centres during the initial roll-out period of the LCG Phase I.** As only Phase I of the LCG Project has been discussed in detail and authorised, additional detailed discussions are needed for the transition to Phase II. **The Memorandum of Understanding for Phase II must be developed over the next 12 months.**

The LHCC concludes that the LCG Project has made an impressive start to a very complex task in a very short time. Computing ingredients are becoming available and key tests of the system will be made through the increasingly demanding Data Challenges. The focus for 2004 is the take-up by the experiments of the applications

products and Grid service, the development of computing models and the second round of Grid technology aimed at distributed analysis. A fundamental ingredient for the success of the LCG will be its ability to establish itself as a truly global project.

The LHCC noted significant progress in the LCG Project since its inception. The Committee took note of the new organisation structure put in place and that a significant majority of the LCG Project's milestones are being met on time. Although resources are continuing to be allocated for the most part of Phase I, detailed discussions are required leading up to a Memorandum of Understanding for Phase II of the LCG Project.

4. MIDDLEWARE AND INTERACTION WITH EDG/EGEE AND OTHER DEVELOPMENTS IN MIDDLEWARE

The LCG Project does not develop and therefore does not control the middleware technology evolution. However, it does identify the starting technologies to be deployed, sets-out the evolution strategy and the long-term support technologies and works towards the future coherent, acceptable and supportable middleware solutions.

The LHCC took note of the developments in the Enabling Grids for e-Science in Europe (EGEE) initiative. The undertaking is a European Union project approved to provide partial funding for operation of a general e-Science Grid in Europe and includes the supply of suitable middleware. EGEE provides funding for 70 partners, the majority of whom have strong ties with the high-energy physics community. The LCG Project operates the EGEE infrastructure as a service to EGEE and thus ensures the compatibility between the LCG Project and EGEE Grids.

The EGEE effort includes provision of a middleware package satisfying requirements agreed with the LCG Project and with other sciences. The middleware tools will provide functions that are of general application, including applications beyond high energy physics, and that can reasonably be expected to be available from public or commercial sources such as Internet protocols, UNIX and HTML.

The Committee noted the very tight EGEE delivery timescale dictated by the LCG Project requirements and considers that the existing middleware is generally too complex and under-developed. From past experience with the European Data Grid (EDG), the main risk appears to be the lack of product delivery. The management of both the LCG Project and Grid Technology Area (GTA) are aware of this risk and are emphasising the need for simplicity. The LCG Project looks forward to EGEE using its resources to move to the analysis phase. It was, however, noted that the existing production phase has not been fully mastered and would need to do so before moving into the next phase.

The LHCC considers it very important for the middleware project to ensure tight links and collaboration with the US part of the effort and to establish a close and better collaboration with all the Regional Centres.

The Committee noted that convergence of the LCG Project with the EDG is foreseen with the release of EDG Test-bed 2. At this point, the LCG and the EDG will both be based on the same technology infrastructure of VDT (the Virtual Data Toolkit comprising Globus, Condor and additional components developed in the US) together with EDG components. The actual components used will depend on their maturity level and robustness.

A service-based approach to the provision of middleware – the Architectural Roadmap towards Distributed Analysis (ARDA) - is being developed in parallel to the LCG-1 and LCG-2 production and analysis services but need to be proven as a strategic direction. *The plan for ARDA will be submitted to the SC2 by the end of January 2004.*

The LHCC will continue monitoring progress in the middleware project, particularly given the upcoming Data Challenges.

The LHCC took note of developments in middleware project focusing on progress in the EGEE, EDG, VDT and ARDA endeavours. The Committee identified several concerns, particularly the risks associated with lack of product delivery and the difficulties in

entering the analysis phase of the project. Further collaboration with the US partners and with all the Regional Centres is recommended.

5. LCG FABRIC AREA

The goal of the computing Fabric Area is to prepare the Tier-0 and Tier-1 centres at CERN. The former focuses on the mass storage of the raw data, the first processing phase and the data export to Regional Centres, for example, raw data copies, while the latter concentrates on the data analysis task.

Good progress was demonstrated in the Fabric Area. In particular, the LHCC noted the positive collaboration with various institutes and industry, the successful Data Challenges and that the agreed milestones are being met on time. The LHCC noted no major concerns although it did identify that the overall challenge is with the market development of the technology (network backbone and storage). The timescale is tight but is not considered to be critical. Further work in the Fabric Area includes defining the computing models in more detail. The Fabric Area is considered to be the least controversial component of the LCG Project.

The LHCC has no major concerns regarding the Fabric Area. The Fabric Area is considered to be the least controversial component of the LCG Project.

6. GRID DEPLOYMENT AND REGIONAL CENTRES

The aim of the LCG Grid Deployment Area is to deploy and operate a prototype LHC computing environment. Its scope is to integrate a set of middleware, to coordinate and support its deployment in the Regional Centres, to provide operational services to enable running as a production quality service, to provide assistance to the experiments in integrating their software and to provide direct user support.

The LHCC noted that certain key milestones have not been met on time. Due to the late arrival of middleware and the slow take-up in the Regional Centres, the July 2003 LCG-1 milestone, i.e. the release of the first global Grid service, was 3-months late and included a reduced functionality with respect to the original design. The November 2003 milestone has only partially been met owing to a lack of operational experience and to the delay in the start of testing of the system by the experiments. However, LCG-1 has been deployed successfully at several sites and a number of sites are joining soon or are preparing to do so. Computing power is added according to the requirements from the experiments. Regional Centres have installed successfully the LCG-1 package. LCG-2, which incorporates reliability and performance targets, is nonetheless expected to be deployed as scheduled in December 2003 as its functionality is required for next year's Data Challenges starting with ALICE in January. **Finally, the Committee recommends that the Regional Centres should be queried on how they believe funds will become available to achieve their required computing capacity.**

The LHCC took note of the on-going discussions concerning modifications to the Grid Deployment Board (GDB), which is the collaboration board of the Regional Centres and the experiments. **As the high-level issues have for the most part been solved, the GDB should henceforth ensure that more detailed technical issues are discussed.**

Good progress was reported on the LCG Security Group, whose mandate is to advise and make recommendations to the Grid Deployment Manager and to the GDB on all matters related to LCG-1 security.

The Committee took note of the creation of a task force to define the requirements and to agree on a prototype for a Grid Operations Service and of the MapCenter facility, which performs low-level tests and aggregates these up through several levels, with the highest being individual countries, in order to show the best and worst status at each level.

Integration of the LCG Project into the experiments is proceeding with the aim of helping the experiments implement their production and analysis environment within LCG Project. Difficulties have, however, been reported on the deployment owing to

installations being too complex, and to the slowing down of the propagation of changes due to the various time zones. The LHCC underlined that it is only the beginning of the LCG Project and the community is learning to handle the new Grid computing environment.

Although delays in the deployment of the LCG-1 software package have been reported, several Regional Centres have now installed successfully the package and integration of the experiments into the LCG Project is in progress. Problems encountered with LCG-1 are being corrected in time for LCG-2 that is to be deployed by the end of 2003. The LHCC underlined that it is only the beginning of the LCG Project and the community is learning to handle the new Grid computing environment.

7. LCG APPLICATIONS AREA

The Applications Area develops and maintains that part of the physics applications software and associated infrastructure that is shared among the LHC experiments. The scope includes common applications software infrastructure, frameworks, libraries, and tools, together with common applications such as simulation and analysis toolkits, Grid interfaces to the experiments and assisting the integration and adaptation of physics applications software in the Grid environment. The aims and scope of the Applications Area are being defined properly through the RTAGs. Products such as POOL (Pool Of persistent Objects for the LHC), developed as a common persistency framework for physics applications at the LHC, have already been deployed in the production environment in at least one of the LHC experiments.

Measures of tracking the Applications Area project are in place and are based around a list of agreed milestones. Included in the list of milestones is the integration of the application into the experiments. The LHCC noted that time to integrate POOL and SEAL, which aims to provide the software infrastructure, basic frameworks, libraries and tools that are common among the LHC experiments, into the experiments has generally been underestimated.

The level of existing manpower appears to be adequate. It needs to stay so in order to ensure the timely development and integration of the software. **Consideration must also be given to achieving the continuity of personnel on the long-term.**

The Committee **expressed some concern about the long-term support of products, particularly in the math library area. Moreover, the Committee requests further clarification of how proposals made in the Architect's Forum are to be incorporated in to the Applications Area.** Two Application Area milestones concerning the distributed production environment and the distributed analysis environment have been suspended pending specification of requirements by SC2. Finally, the Committee **stresses the importance to support the Monte Carlo generator codes required by the LHC experiments. Such support appears to fit the scope of the Simulation project.**

The LHCC noted the good progress in the Applications Area. Manpower levels are adequate, the scope of the project is properly defined and milestones are being met, including deployment of the first products, e.g. POOL, in the production environment of at least one of the LHC experiments. The Committee underlined the importance of supporting the Monte Carlo generator codes for the experiments.

Emmanuel Tsesmelis