GridPP Project Management Board

# GridPP – Strategic Review

| | |
|---|---|
| Document identifier : | **GridPP-PMB-170-GridPP5_StrategicReview** |
| Date: | **20/2/14** |
| Version: | **18** |
| Document status: | **Final** |
| Author | **PMB** |

# STFC Programme Proposal

## The UK Grid for Particle Physics Collaboration

University of Birmingham
University of Bristol
Brunel University
CERN, European Organisation for Nuclear Research
University of Cambridge
University of Durham
University of Edinburgh
University of Glasgow
Imperial College London
Lancaster University
University of Liverpool
University of Manchester
University of Oxford
Queen Mary, University of London
Royal Holloway, University of London
Rutherford Appleton Laboratory
University of Sheffield
University of Sussex
University of Warwick
Swansea University
University College London

**Contacts**

Prof. David Britton –David.Britton@glasgow.ac.uk (Project Leader)
Prof. Steve Lloyd – S.L.Lloyd@qmul.ac.uk (Collaboration Board Chairman)
Mr. Peter Gronbech – p.gronbech1@physics.ox.ac.uk (Project Manager)

## Abstract

This document contains input from GridPP to the Strategic Review Panel. We respond to the request for "Flat-Cash" and "50% of Flat-Cash" scenarios to cover the provision of Grid Computing for the LHC experiments and other groups, for the period April 2015 to March 2019. This period covers the LHC Run-2 where an increase in computing resources of a factor of 2.5x is required to exploit the last two decades of investment in the LHC programme.

.

## Contents

# 1 Foreword

Grid for Particle Physics (GridPP) provides the UK share of the computing resources and infrastructure required to exploit the LHC to the Worldwide LHC Computing Grid (WLCG). The level of resources provided reflects the size of the UK's involvement in the LHC experiments and the manner in which they are provided, and the level of service with which they are delivered, fully meeting the requirements of an international Memorandum of Understanding (MoU)[1] signed by STFC. In principle, funding for GridPP5 should follow as a consequence of existing strategic decisions about the UK's involvement in exploiting the data from the LHC Run-2 and the level of funding required is well defined by the MoU requirements; It makes little sense to perform a strategic review of GridPP as if it were decoupled from the on-going level of support for ALICE, ATLAS, CMS and, LHCb in the UK. Similarly, it makes even less sense to tension GridPP5 against the Upgrade Projects: GridPP5 is about exploiting years of investments already made in building the current LHC detectors and the data to be delivered during LHC Run-2. The Upgrades are about the future beyond Run-2 and commitments to them must be with the knowledge that they too will require appropriate levels of computing, at the appropriate time, with which to exploit the data that they will generate.

Although the prime objective for GridPP5 is to fulfil the international commitment to deliver the UK contribution to WLCG as defined by the MoU, the project has a subsidiary objective to deliver resources in addition to those specified in the MoU to the UK HEP community to provide a competitive advantage to the LHC users and to support non-LHC groups. Finally, GridPP5 also intends to contribute to the sustainability of the UK e-infrastructure eco-system by exploring synergies with other projects and aligning where possible; and to respond to the STFC Impact agenda.

With that in mind, this document contains the GridPP input to the Strategic Review. We present a Flat-Cash scenario, which enables the UK to deliver the resources at the levels required for exploiting the data from Run-2. We also address the consequences of de-scoping to a 50% scenario, showing how the UK would progressively fail to deliver the required resources and services as the funding reduces.

This document was written based on a set of background documents[2] that address various aspects of GridPP5 planning.

# 2 Outline

The GridPP5 project is composed of the following work-packages:

- WP-A: Infrastructure and manpower at a national Tier-1 centre;
- WP-B: Infrastructure and manpower at the UK Tier-2 sites;
- WP-C: Grid Deployment, Operations and Support;
- WP-D: Management, Travel, Administration, and Impact plans.

The remainder of this document is organised as follows:

**Section 3** provides a brief summary of the experimental motivation for this proposal. We have not provided a detailed justification of the physics case, which is well established.

**Section 4** introduces the background and context, including the WLCG, the LHC Run-2, the evolution of the experiment computing models and the UK and European contexts.

**Section 5** introduces GridPP, briefly describing its history, current status and future plans.

**Section 6** details the requirements of the experiments for the UK Production Grid, including the hardware costing.

---

[1] Memorandum of Understanding for Collaboration in the Deployment and Exploitation of the World Wide LHC Computing Grid: http://wlcg.web.cern.ch/collaboration/mou (In the case of the UK the MoU was agreed and signed by PPARC in 2006).

[2] Username = review; Password = gridpp5.

**Section 7** describes how GridPP intends to meet the requirements of the experiments in the UK by means of a coherent Tier-1 and Tier-2 infrastructure (WP-A and WP-B) bound into a Grid by the work of the deployment/operation/support teams (WP-C) and a light-weight management structure (WP-D).

**Section 8** addresses the Flat-Cash and reduced scenarios.

**Section 9** concludes.

# 3 Motivation

***The top ranked science of the LHC cannot be achieved without the Grid infrastructure***

The discovery of the Higgs boson by ATLAS and CMS, announced in July 2012, was a dramatic vindication of enormous investments over the last 20 years in the LHC machine, the detectors and, more recently, the WLCG. The UK holds preeminent roles in all areas of this scientific endeavour; including past and present spokesmen of the General Purpose Detectors (GPDs), ATLAS and CMS, and of the LHCb experiment. Of particular relevance, UK investment via GridPP of £5.6m in the period 2001-2003 created the basis from which the WLCG collaboration developed to deliver the computing infrastructure so successfully. The UK has an unsurpassed international reputation and credibility in all parts of this quest for the Higgs boson and it is fitting that the UK shared the 2013 Nobel Prize for Physics when it was awarded to Peter Higgs and François Englert.

From a scientific point of view, the Higgs discovery marks the starting point for a decade of work that will tease out the details of the Higgs sector and possibly make other stunning discoveries. Particle physics will be dominated for the foreseeable future by the LHC at CERN with the three year Run-2, starting mid 2015 designed to produce 2-3x the annual data obtained in 2012. Run-3, starting in 2020, and the High Luminosity LHC (HL-LHC) starting in about 2026, are already being planned and the data they produce will be an order of magnitude larger. The science case for GridPP has already been established and is reflected by the top ranking of the GPDs in the Programmatic Review. The science, however, cannot be achieved without the computing infrastructure, any more than it would be feasible without the detectors or the LHC accelerator - as was highlighted by the CERN Director General at the seminar announcing the Higgs discovery.

Although this proposal focuses on the provision of computing resources for the LHC experiments, there has been considerable (and growing) uptake of the Grid paradigm by other HEP and non-HEP communities. In particular, in 2013 the T2K collaboration reached the limit of the storage GridPP could provide until an injection of additional capital allowed us to extend capacity at the Tier-1. We also anticipate growing requirements from the ILC and NA62 collaborations if these projects continue to develop in the UK. Extending the GridPP capacity and capability is a much more cost-effective way of meeting new requirements than building new infrastructures.

# 4 Background and Context

***GridPP provides hardware and services in the context of an international collaboration***

WLCG is much more than a collection of resources, it is an international collaboration that provides a coherent set of linked resources and associated services that are constantly evolving to address increasing requirements; changes in the software, middleware and operating systems; and advances in computer hardware. This evolution takes place whilst maintaining a 24x7 high availability (98%) service. The worldwide collaborative nature of the project has enabled large-scale resources to be harnessed together from many different funding sources. The corollary to this is that it requires collaborators to provide resources in a compatible manner and to contribute to the continuous evolution that is gradually increasing the capabilities and reducing the complexity of the service.

The contributions expected from the UK consist of computing resources delivered at various, specified, levels of service, plus contributions to the running and development of the infrastructure. The roles of many GridPP staff are multifaceted, containing elements related to running the hardware; providing services and support; upgrading the infrastructure; testing future changes; and developing longer term solutions aimed at increasing the performance and decreasing the staff cost. More explicitly, the UK contributes services that are used by the whole collaboration such as the APEL accounting system and the Grid Operations Centre Database (GOCDB) and uses services offered by other countries, such as the Global Grid User Support (GGUS) ticketing system. The UK also takes its share of international leadership roles, for example in security operations, the development of security policy, federated identity management and a number of WLCG and HEPiX technical working groups. In short, GridPP is the UK part of an international collaboration, with the accompanying advantages and responsibilities.

The level of global resources required by the experiments are in effect determined by the science programmes approved for each experiment. Once the trigger rate is fixed, the processing and storage requirements then follow, with little scope for variation within the constraints of the current computing models. These resource requests are scrutinised and approved by the CERN Computing Resource Review Board (CRRB). Over the last decade, the UK has pledged and delivered resources to WLCG based on the UK authorship fraction in each experiment. The UK is part of all four of the main LHC experiments, with about 2% of ALICE, 10% of ATLAS, 4% of CMS, and 19% of LHCb authors and there is an explicit expectation in the WLCG MoU that the UK will continue to contribute the commensurate share of the required computing.

## 4.1 Run-2 at the LHC

***The computing models must evolve to mitigate the otherwise needed 6x increase in resources***

The running conditions during Run-2 (2015-2018) will differ considerably from those in Run-1 (2011-2012). The centre of mass energy of the machine will almost double to 13TeV and the intensity of the beams will be increased to achieve instantaneous luminosities more than triple those delivered at the end of 2012. Consequently, events recorded will be significantly larger and more complex because they will contain more pile-up[3] interactions: More storage will be required and more CPU will be needed to reconstruct and analyse this data.

The LHC experiments have estimated that the increased events' size and complexity in Run-2 will require an increase in computing resources by about a factor of 2.5. In addition, to exploit fully the physics that the upgraded LHC can deliver, the experiments plan to increase the rate at which data is taken from around 400Hz in 2012 to approximately 1KHz in 2015; a further increase of 2.5x. Taking both factors into account, computing resources would need to increase by a factor greater than 6 if everything were done as in 2012. However, to mitigate this increase, the LHC experiments are evolving their computing models and improving their software, which in turn requires GridPP to adapt the Grid infrastructure to the new systems. This directly illustrates why WLCG and GridPP are not static infrastructures but are constantly evolving.

## 4.2 Experiment Computing Models

***A more effective and versatile infrastructure is developing, moderating the rising requirements***

The original hierarchical computing models for the LHC experiments have served us well throughout Run-1 of the LHC. However, the more challenging environment of Run-2 has driven the experiments to make significant updates to their computing models, exploiting new technology where appropriate. The computing models were established nearly a decade ago and since then there have been massive improvements in wide area networks and Run-1 has provided the experiments with a much better understanding of their data. These factors are complemented by the agility that virtualisation and cloud computing is beginning to bring to sites. The effects of these advances, individually and collectively, have had considerable impact on how the experiments plan to perform their computing in Run-2.

The original computing models were characterised by a very strict hierarchical structure with a rigidly defined set of tasks being performed at the Tier-0 at CERN, a different set of tasks being performed at the National Tier-1 centres and finally a third, generally less coordinated, set of tasks being performed at the Tier-2s. Custodial copies of the data were kept (on tape) at the Tier-0 and at (at least) one Tier-1. Primary reconstruction was carried out at the Tier-0, and re-reconstruction was primarily carried out at the Tier-1 centres. The Tier-1 centres also carried out such tasks as skimming. The Tier-2 sites

[3] Pile-up occurs when multiple independent collisions take place in the same bunch-crossing; the recorded event, therefore, contains overlaid information from the different collisions, which has to be disentangled.

generally carried out user analysis (although all LHCb user analysis was originally only at the Tier-1s) and Monte Carlo generation tasks.

Improved networks, together with improved capability/reliability of the sites coupled with the increasingly sophisticated experiment workflow management systems, now allow this model to be relaxed and the GPDs have split the functionality of the Tier-1s into a custodial tape archive and a processing site that is geographically co-located but in other ways quite separate. The processing aspects of the Tier-1s are now very similar to those provided by the Tier-2 sites, except that they are provided, in principle, with a higher reliability and availability than at a Tier-2 site. The wide area network now allows these Tier-1 processing sites (and potentially some of the larger Tier-2 sites) to become remote arms of the Tier-0 at times when there is spare capacity at the Tier-1 centres but insufficient capacity for the Tier-0 to keep up with the prompt reconstruction of the data being taken. This blurring of the distinction between the different tiers allows for a more efficient use of the computational resources. A further demonstration of the use of wide area networking is the splitting of the CERN Tier-0 (and other compute resources) between the CERN site in Meyrin and the new Wigner Data Centre in Hungary.

The hierarchical models initially used by the experiments envisaged data being distributed to specified sites on the Grid, and the middleware layer dispatching jobs to the data. To be effective, multiple replicas of key data sets were required. The improvement of the wide area network and the use of protocols such as xrootd have enabled an analysis job running at one site to analyse data stored at another site directly over the wide area network without having to download a copy to local storage. Both the GPDs have major programmes to extend the access to data at remote sites. For example, the 2013 re-processing of the 2011 datasets by CMS, which traditionally would have been done just at the Tier-1s, was shared between all three tiers. The sharing of the data between sites means that fewer copies of the data are needed, which will save considerable disk space.

The data collected by the experiments typically goes through several stages of reconstruction. The end of this process is a much reduced data format containing “physics objects” that is most suitable for physics analysis. However, the intermediate stages are required for calibration and analyses that require greater access to low-level data entities. A better understanding of the data, the detector, and the calibration process gained from Run-1, will allow the intermediate stages to become transitory during Run-2, with the data only being kept for a limited number of months. This will help to further reduce the disk space requirements of the experiments.

During GridPP4, computer technology has remained relatively stable. However, in order to remain efficient, particle physics must embrace the new computer architectures and techniques that are starting to appear. Architecturally the two biggest changes are the rise of multi-core CPUs (which will be accompanied by coprocessors in the next generation of machines) and the growth of many-core GPU cards. At the moment GPU usage is confined to niche areas within particle physics (such as the calculation of oscillation probability in the T2K experiment), however in these areas they can provide two orders of magnitude increases in performance and their usage is set to grow. Unfortunately, they are not suited to many traditional particle physics codes and those to which they are suited often require considerable efforts to port to them. Particle physics codes have traditionally been of the “single-core, embarrassingly parallel” variety and have been well suited to single-core x86 CPUs. The efficient use of multi-core CPUs (and later coprocessors) has required the LHC experiments to start modifying their code to use simultaneously all the cores in a given CPU. Tier-1 and Tier-2 sites are beginning to open batch queues that schedule multi-core jobs and because these jobs share software components they represent a considerable saving in the amount of memory required per core.

Machine Virtualisation and cloud technologies are also beginning to have a significant effect on computing within particle physics. Virtualisation of machines running services has for a number of years provided a more resilient level of service, which can be decoupled from the failings of individual pieces of hardware, and increase the agility with which services can be deployed. Cloud interfaces provide an automated way of deploying virtualised resources, which is becoming particularly attractive in conjunction with the development of private clouds. Although public clouds, such as Amazon and Google, remain unattractive in terms of cost, functionality and ease-of-use for our applications, the internal provision of resources via a private cloud interface, which may or may not be directly exposed to the user communities, is a development that potentially simplifies the infrastructure (saving costs) at a small price in efficiency. CERN plans to make all of their resources available via a cloud layer only in

2014 and the Tier-1 at RAL plans to offer a similar prototype during 2014 and deploy into production early in GridPP5.

The demand for compute resources has led the LHC experiments to develop the use of "opportunistic" resources. These are resources that are not dedicated to a particular experiment's offline computing but which become available to the experiment for intermittent periods. While these include a variety of resources such as temporary access to supercomputing centres, one of the biggest sources of transitory resources available to the experiments are their own High Level Trigger (HLT) farms. For the GPDs these are comparable in size to their total Tier-1 resources. During data taking the HLTs are completely consumed by their role as part of the experimental triggers, however between running periods these are large unused resources. Both the GPDs are utilising virtualisation and cloud technologies to provide an overlay layer on top of the bare machines of their HLTs. Here the flexibility provided by this layer allows the HLTs to transition from being dedicated trigger farms to being general-purpose resources on a timescale of tens of minutes, and to make the reverse transition when stable beams are approaching. GridPP has been engaged in (leading, in the case of CMS) developments in these areas. Virtualisation and cloud technologies will also be implemented in the future to allow more versatile use of GridPP resources by non-LHC groups.

## 4.3 The UK Context

***A key component of the UK e-infrastructure ecosystem***

Since its inception, GridPP has closely coordinated its activities with many of STFC's investments in e-Science – particularly those of the e-Science Centre at RAL (now Scientific Computing Department) and the UK NGI (National Grid Initiative) whose services (such as GOCDB, APEL and the UK Certificate Authority) GridPP depends on (and partly funds). With the end of the National Grid Service (NGS) GridPP has largely assumed the role of the UK NGI in relating to the European Grid Infrastructure (EGI).

Delivery of the Tier-1 is the responsibility of STFC's Scientific Computing Department and products emerging from the Tier-1 such as the CASTOR storage system are now key components of the data pipelines of STFC's facilities such as Diamond and shortly CLF (as well as NERC's JASMIN/CEMS service). Much of expertise gained at the Tier-1 in delivering high availability scientific compute services is now being integrated into the Hartree Centre operations and in turn the Tier-1 plans to investigate how to exploit resources made available by Hartree.

The GridPP Tier-1 in the context of the Scientific Computing Department could play an important future role in developing the EU-T0 collaboration that is currently being discussed by European funding agencies and CERN.

GridPP has increasing contacts with the DiRAC High Performance Computing consortium in the UK, both at the technical level, for example providing advice on data-transfer techniques, and at the management level where both projects work in the context of the UK Project Directors Group to develop and integrate the UK e-infrastructure.

## 4.4 The European Context

***Translating the UK leadership roles and leverage of funding from EGI to H2020***

WLCG works with a number of international projects, namely the Open Science Grid (OSG) and the European Grid Infrastructure (EGI), to coordinate delivery of the basic functionality upon which both GridPP and WLCG depend. These international production Grids coordinate middleware deployment, operational support, support infrastructure, monitoring, security and many other areas. Over the last decade, the EGI has built a distributed computing and data infrastructure across Europe, Asia, Australia and Africa. It has grown from the prototype European DataGrid in 2001, through three phases of EGEE (Enabling Grids for E-SciencE) and now delivers an unprecedented data analysis capability to over 21,000 researchers, of which the High Energy Physics community is the largest stakeholder.

The UK has taken a number of leading roles within EGEE and EGI, reflecting UK expertise and ensuring UK community influence on key areas of operational development. Currently the UK chairs the Project Management Board and Project Administrative Committee (EGI-InSPIRE) and has fielded one of the seven EGI Executive Board positions since the beginning of EGI in 2010. The EGI Collaboration is composed of representatives from the National Grid Initiatives (NGIs) and supported research communities. The UK NGI was formed from a combination of GridPP and the National Grid Service (NGS). Since the NGS ended in 2013 GridPP has led the UK NGI. Although GridPP must remain focused on the requirements of the LHC experiments it has gained from being in the NGI and the European landscape. There are three main reasons for this: (a) WLCG relies on EGI so it is in the interests of particle physics that the UK NGI succeeds. Strong participation from GridPP is the best way to achieve this; (b) the core NGI services required by distributed computing projects (e.g. authentication and accounting) are not yet supported independently in the UK, as there is currently no major source of funding for an NGI apart from existing projects. We believe that if the UK is to participate in EGI then GridPP should, where appropriate, align work in areas of strength in order to support and exploit the future development of an NGI; (c) GridPP has technical leadership in WLCG and EGI in specific areas (accounting, configuration databases, security policy and operations, and IPv6 transition). These activities are required by WLCG and have provided significant leverage for GridPP including international leadership opportunities and co-funding. If GridPP does not maintain these areas, then there are risks of severe disruption to WLCG and EGI.

Looking forward to Horizon 2020 (H2020), there are opportunities for GridPP to leverage financial support and influence by collaborating with other NGIs and EGI, especially in two calls. EINFRA-1-2014 (line 6) closing September 2014 has a maximum contribution of €8M and will support the evolution of EGI towards a flexible infrastructure capable of federating resources of any kind, to offer services to the whole European scientific community. Accounting, security and a service repository (GOCDB) are co-developed for EGI by GridPP and are essential components envisaged to be a part of such a bid. EINFRA-7-2014 closing September 2014 regards identify federation as a core service across European e-infrastructures. An EGI-TERENA proposal is under discussion and could include authentication and security elements from GridPP. EGI is also co-ordinating proposals for cloud federation (Oxford University-lead, April 2014 closing date) and data management (in discussion with EUDAT and APARSEN, September 2014).

All proposals are currently being drafted. For the September 2014 calls NGIs are currently being asked for input and expertise. Drafts will be created for May and revised so that a fully costed model is available for July and submission in September. The EC review is single staged with the evaluation outcome being known by February 2015 and grant agreements being signed by May 2015. Whilst opportunities in H2020 are possible, the funding available for the EGI call (EINFRA-1-2014) is significantly lower (€8m, likely over 2 years) than for the current FP7 funded EGI-InSPIRE project (€25m over 4 years) and a funding gap will exist between the end of the current project and the start of a future one (12 months), although some activities will be funded through EGI membership fees.

Whilst the future of EGI funding is uncertain, the GridPP strategy to leverage influence through leadership in the international context remains unchanged. However, expectations must be tempered by the reality of steadily falling EC funding over the last decade. The UK received 1.9M€ per year in 2006-8, which fell to 800K€ per year in the period 2008-10, and then further reduced to 400K€ per year for 2013-2014. It is likely to fall further under H2020 as the EC pushes infrastructures to become self-sustaining and concentrates the remaining funds on less domain-specific projects.

Finally, the recently formed “EU-T0” initiative, signed by major EU funding agencies including STFC, provides a clear opportunity to bring together the largest data producers, and to federate national centres into a *European Computing Centre for Experimental Data Management*. The EU-T0 could potentially be ideally placed to confront H2020 opportunities and the Tier-1 at RAL is one of the obvious potential components.

# 5 The UK Grid for Particle Physics

***GridPP is a long running successful project that has underpinned the Higgs discovery***

The GridPP Collaboration was established on 1st September 2001. It is an internationally recognised collaboration of particle physicists and computing scientists from 19 UK universities, STFC and CERN, who built a Grid for particle physics, delivering the UK's contribution to WLCG. The first phase of the project (GridPP1) established a highly functional prototype across the UK and enabled the formation of the WLCG project by providing £5.6m funding for staff at CERN over a three-year period. The second phase (GridPP2) started in September 2004 and deployed a working worldwide Grid in close association with the EGEE and WLCG projects. After a short transition period (GridPP2+) to align the funding cycles, GridPP3 started in April 2008 and ran until March 2011, successfully participating in a series of data challenges, such as STEP09, before receiving the first data from the LHC. GridPP4 commenced in April 2011 and has seen nearly 30 $fb^{-1}$ of data acquired and processed by the LHC experiments in 2011 and 2012, which led to the discovery of the Higgs boson. GridPP has delivered at the highest level, on-time, on-budget, and above specification (the success of the LHC and the experiments in producing data beyond the levels expected have pushed bandwidths and volumes beyond those specified in the original design).

Progress in GridPP4 is monitored by a set of milestones, which mark specific deliverables, together with a set of metrics that monitor the on-going performance of the project against a challenging set of goals. As of the 3rd quarter of 2013, which is just over half way through the project, 54 of the 81 milestones were complete and over 92% of the actively measured metrics were at their target levels. Progress is reviewed by the GridPP Oversight Committee that last met in November 2013 when they *"acknowledged that GridPP had delivered its programme well and that GridPP4 also seemed to have made good progress."*

## 5.1 GridPP5: From Run-1 to Run-3

***Evolving a more agile and effective infrastructure for Run-2 and preparing for Run 3***

The primary purpose of this document is to provide input to a strategic review of GridPP, which will define the future scope of the project. Two very different funding scenarios have been requested, which will lead to very different outcomes. We have explained that the level of resource requirement is determined by the physics programme and approved by the CRRB. Only the Flat-Cash scenario set out below will enable the UK to deliver the resources that are expected and formalised in the MoU. The 50% scenario will require STFC to enter into negotiations with CERN and our international peers in order to reduce the UK contribution below the agreed MoU levels, which may have the knock-on effect of increasing costs to the UK elsewhere such as M&O[4] contributions. There is also a risk of the UK setting a precedent that could weaken the overall LHC programme.

Maintaining the current scope of GridPP in a Flat-Cash scenario is challenging but it is a tractable problem. Moore's law and the evolution of the experiment computing models partly compensates for the growth of the resource requirements, but manpower must also be reduced by continuing to adopt new tools and techniques to automate and simplify the infrastructure. Significant progress has been made in this area during GridPP4 and we are confident that, with adequate resources to drive this forward, the process can continue in GridPP5.

In the Flat-Cash scenario, GridPP5 would deliver the committed increase in Tier-1 and Tier-2 resources at the current levels of service, as specified in the MoU. In the context of the new computing models, the Tier-1 and Tier-2 centres would take on new roles previously reserved for high-level tiers, providing an increasingly agile and flexible infrastructure that will allow more efficient use of the overall resources. This evolution will require experienced and dedicated staff; continuity will be particularly important as will stable operations throughout.

Looking towards the future, as Run-2 progresses, GridPP5 will need to make the preparations required for Run-3. No detailed planning is available yet, however it is probable that although the current WLCG computing model will meet the requirements of Run-2, beyond that LHC computing will need to establish new ways of working[5]. In order to handle the projected high data volumes, the community will need to exploit tools and technologies emerging from the world of Big Data. The highly tuned, domain specific, bespoke middleware currently deployed will need to be replaced by generic "open" solutions with a wide support base. The increased use of standard interfaces may allow increased synergy with other high throughput Big Data projects and there is the potential for sharing infrastructure costs and expertise. As the middleware evolves, the challenges around the long-term curation of the older data

[4] M&O is the Maintenance and Operations contributions that the UK makes to the central costs of running the experiments.

[5] Computing at the HL-LHC (Predrag Buncic - ECFA High Luminosity LHC Experiments Workshop – October 2013)

will increasingly become an issue. It is essential that GridPP has the ability to contribute to, and develop, these areas. There is a high risk that short-term savings on GridPP5 manpower that prevent engagement with future developments will result in longer-term costs.

Looking more widely, GridPP supports a significant community of non-LHC users. As the infrastructure evolves to adopt more standard interfaces and less domain-specific middleware, there are increased opportunities for this user-base to expand. In the Flat-Cash scenario, GridPP would continue to support and expand these communities.

### 5.1.1 GridPP5 Project Structure

The remainder of this document describes in more detail the components required for GridPP5 as a 4-year project starting in April 2015. Details are provided about how significant evolution of, and an increase in, the delivered resources will be achieved in a Flat-Cash scenario. The hard choices that have to be made about what can and what cannot be delivered in a 50% Flat-Cash scenario will then be addressed, identifying the funding points where step-changes in functionality occur.

The components of GridPP5 are given by the following work-packages:

| | |
|---|---|
| **WP-A**<br>Tier-1 Infrastructure | The set of hardware and services that make up the UK Tier-1 centre at RAL. These include basic resources like CPU, storage and network and higher-level services like an Information Service and File Transfer Service. *The Tier-1 is essential for the UK to meet international obligations to receive, process and curate LHC data for the global LHC collaborations.* |
| **WP-B**<br>Tier-2 Infrastructure | The set of hardware and services that make up the UK Tier-2s. For ATLAS and CMS the UK Tier-2s provide the front-line data-analysis capability and for all experiments the Tier-2s are required for simulation production. *The Tier-2s meet international obligations to perform scheduled analysis and are essential for UK physicists to compete for global leadership in extracting the core physics from the LHC.* |
| **WP-C**<br>Grid Deployment, Operations and Support | Coordination across Grid operational tasks including infrastructure-wide monitoring, data handling, security, release management, training, documenting and support of institute-based system experts. Support and communication at the critical boundary between Grid-enabled experiment software and generic Grid services. *The GridPP "Ops-Team" and NGI-based operations will be essential for the coherent and secure operation and evolution of the infrastructure. The support for data handling is mission-critical for the efficient extraction of physics results. Experiment-specific support has proved pivotal for success.* |
| **WP-D**<br>Management, Admin, Impact, Travel | Management of a complex project in the national and international contexts, across 21 institutions; the ability to respond to the impact-agenda; and the ability to collaborate and communicate. *An experienced management team ensures good governance and value-for-money. Sharing knowledge and building trust underlies success.* |

This breakdown of work packages is based on many years of experience and reflects the WLCG environment in which GridPP operates. Work packages A and B, contain the hardware and manpower at the Tier-1 and Tier-2 sites respectively. The deployment, operation, and support of the UK Grid is shared between dedicated staff, described in work package C, and the Tier-2 staff covered in work package B. This is focused around the GridPP Operations-Team that is described later and ultimately transforms the collection of UK Tier-1 and Tier-2 resources into a functional Grid. Work package D addresses the project-level management, administration, travel, and the impact-agenda.

# 6 Experiment Requirements for the UK Production Grid

In this section we summarise the resource and service requirements that the UK computing centres need to meet, in the context of the international requirements.

## 6.1 LHC Global Resource Requirements

***Determined by the approved physics programme, trigger rates, and computing models***

The evolution of the computing models to mitigate the more than 6x increase in computing resources nominally required during Run-2 has been described in Section 4.2. Three things underpin this evolution: Firstly, high bandwidth and reliable networks have allowed the rigid Tier-0-1-2 hierarchy of sites to be replaced by a much more versatile and adaptive configuration with functionality shared across two or more categories and, as data becomes more mobile, fewer duplicate copies are required. Secondly, an improved understanding of the detectors and the experimental data has allowed some intermediate datasets to become transient, further reducing the overall storage requirements. Thirdly, the adoption of new technologies and techniques such as virtualisation and cloud interfaces has enabled the use of temporary and opportunistic resources such as the High Level Trigger farms at CERN. The net effect of these changes is that the growth in the overall requirements, shown in Figure-1 is contained to about 20% per year.

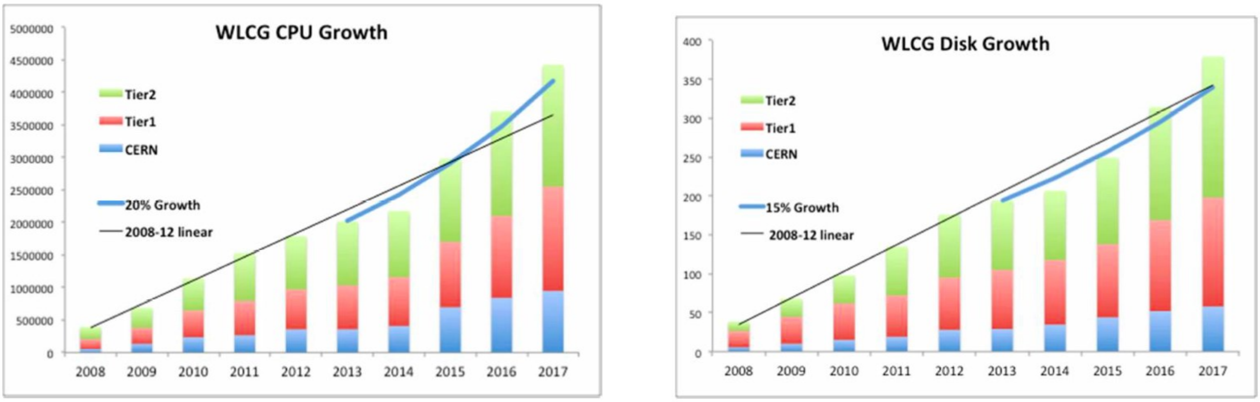



*Figure-1: Predicted growth in the required WLCG resource during Run-2.*

As noted earlier, global resource requirements are estimated for each Tier by the LHC experiments based on the expected machine performance, the experiment trigger rates, and the experiment-computing model. These requirements are iterated with the Computing Resource Scrutiny Group (C-RSG), which assures a common set of assumptions across the experiments, and are approved by the Computing Resource Review Board (CRRB). The UK is represented on the CRRB by STFC.

To meet the global requirements for Tier-1 and Tier-2 resources, individual countries including the UK, are asked annually to confirm their contributions within the framework of the MoU; these so-called “pledges” are finalised by October each year in order to ensure the requirements for the coming year can be met in a timely manner. Updates on what is “planned to be pledged” for years beyond the immediately upcoming year are also requested each autumn.

## 6.2 UK LHC Resource Requirements

***Delivering a proportionate share of the global requirements***

The UK contribution to the global resource requirements is based on the size of the UK involvement in each experiment. For the Tier-1 we take the number of UK authors divided by the total number of authors at Tier-1 hosting countries leading to UK contributions, of 2%, 12.5%, 8% and 31.5% for ALICE, ATLAS, CMS and LHCb respectively. For CMS, the number should be 7% but in previous phases of GridPP the UK has agreed to provide 8% as this is the minimum viable size for a CMS Tier-1 site. For the Tier-2s, the denominator is the total number of authors[6] leading to UK pledges of 2%, 10%, 4% and 21.5% of the global resource requirements. However, GridPP provides additional Tier-2 resources to ATLAS and CMS for local (i.e. UK) analysis (LHCb use the Tier-1) and the Tier-2 fractions actually provided are 2%, 12.5%, 5% and 21.5% respectively.

The LHC experiments have presented their computing requirements for 2014 and 2015 to the Computing Resource Scrutiny Group (C-RSG). The numbers have been endorsed and included in the REBUS system[7]; however, following an agreed schedule-change to solve an approval/pledge timing problem, the new CRRB cycle will confirm (possibly updated) 2015 requests in spring 2014. For 2016 and 2017, preliminary estimates of the resource requirements by the LHC experiments are documented[8] in a review of the LHC Experiment Computing Models, presented to the LHC Computing Committee (LHCC) in December 2013. The numbers contained in Version-2.5 of that document have been used in the preparation of the GridPP5 resource requirements, together with a more recent update from LHCb containing numbers that will be submitted to the C-RSG in 2014.

The LHC Run-2 starts in the spring of 2015 and extends to mid-2018. However, the GridPP5 proposal covers the period up to March 2019 and must, therefore, provision the hardware that will need to be procured in 2018 for use in 2019. To estimate the hardware requirements for 2018, we have linearly extrapolated the growth estimated from 2015 to 2017 and assumed a 25% increment is needed. Very roughly, this is a flat-cash increment, if one assumes a 2-year Moore's Law growth, and is consistent with the increments between 2015-16 and 2016-17. For 2019 when the LHC will not be running, we have assumed a 5% increment.

## 6.3 Non-LHC Resource Requirements

***A cost effective way of meeting the needs of other parts of the STFC portfolio***

In addition to supporting the four LHC experiments, GridPP has traditionally provided computing resources to other particle physics experiments and, contributing to STFC's impact agenda, to other sciences and to commercial entities. Historically, we have been funded to provide 5-10% of overall resources to non-LHC groups and this has proved to be a good match to demand: Over the last 10-years, for example, 9.8% of CPU in the UK went to non-LHC VOs. Storage usage has generally been lower but has been increasing more recently. In estimating the hardware for GridPP5, we contacted all the non LHC-VOs and received some explicit guidance on resource requirements from T2K, ILC and NA62. These have been included explicitly in the planning. For other users, and potential users, we have assumed that resources at the level of 10% of the LHC Tier-1 and Tier-2 resources should be provisioned.

---

[6] Excluding CERN authors for LHCb; this is logically correct, as there is no CERN Tier-2; but a small effect for ATLAS/CMS.

[7] The WLCG REsource Balance and USage system is the official repository of requirements, capacities, and pledges.

[8] https://indico.cern.ch/getFile.py/access?contribId=0&resId=1&materialId=1&confId=212502

## 6.4 Total UK Resource Requirements

The full table of the estimated hardware requirements is shown in Table-1.

| | | From Draft Planning Doc | | | Extrapolations | | | From Draft Planning Doc | | | Extrapolations | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Tier-1 Resources | | | 1.25 | 1.05 | | Tier-2 Resources | | | 1.25 | 1.05 |
| | | **2015** | **2016** | **2017** | **2018** | **2019** | | **2015** | **2016** | **2017** | **2018** | **2019** |
| ALICE (UK) | CPU [KHS06] | 2.4 | 3.2 | 4.2 | 5.3 | 5.5 | | 4.0 | 4.8 | 5.4 | 6.8 | 7.1 |
| | Disk [PB] | 0.4 | 0.4 | 0.4 | 0.5 | 0.6 | | 0.4 | 0.5 | 0.6 | 0.8 | 0.8 |
| | Tape [PB] | 0.2 | 0.3 | 0.4 | 0.5 | 0.5 | | | | | | |
| ATLAS (UK) | CPU [KHS06] | 56.3 | 69.0 | 86.4 | 108.0 | 113.4 | | 65.0 | 76.0 | 91.5 | 114.4 | 120.1 |
| | Disk [PB] | 4.6 | 6.1 | 7.3 | 9.1 | 9.5 | | 6.5 | 9.4 | 12.3 | 15.3 | 16.1 |
| | Tape [PB] | 8.1 | 10.5 | 13.5 | 16.9 | 17.7 | | | | | | |
| CMS (UK) | CPU [KHS06] | 24.0 | 32.0 | 42.0 | 52.5 | 55.1 | | 25.0 | 35.0 | 40.0 | 50.0 | 52.5 |
| | Disk [PB] | 2.1 | 2.8 | 3.6 | 4.5 | 4.7 | | 1.5 | 2.0 | 2.4 | 3.0 | 3.2 |
| | Tape [PB] | 5.9 | 8.0 | 10.8 | 13.5 | 14.2 | | | | | | |
| LHCb (UK) | CPU [KHS06] | 41.6 | 49.8 | 59.9 | 74.8 | 78.6 | | 15.9 | 19.1 | 23.0 | 28.8 | 30.2 |
| | Disk [PB] | 3.7 | 4.7 | 5.1 | 6.4 | 6.7 | | 0.5 | 0.9 | 1.2 | 1.5 | 1.6 |
| | Tape [PB] | 8.5 | 15.7 | 22.1 | 27.6 | 29.0 | | | | | | |
| LHC Total | CPU [KHS06] | 124.2 | 154.0 | 192.4 | 240.5 | 252.6 | | 109.9 | 134.9 | 159.9 | 199.9 | 209.9 |
| | Disk [PB] | 10.7 | 14.0 | 16.4 | 20.5 | 21.6 | | 8.9 | 12.8 | 16.5 | 20.6 | 21.6 |
| | Tape [PB] | 22.8 | 34.5 | 46.8 | 58.5 | 61.4 | | | | | | |
| T2K | CPU [KHS06] | 0.7 | 0.8 | 1.0 | 1.5 | 2.0 | | 0.1 | 0.3 | 0.5 | 0.7 | 0.8 |
| | Disk [PB] | 0.7 | 1.0 | 1.5 | 2.5 | 3.5 | | 0.1 | 0.3 | 0.5 | 0.7 | 0.8 |
| | Tape [PB] | 1.0 | 1.2 | 1.5 | 2.0 | 2.5 | | | | | | |
| ILC | CPU [KHS06] | 0.3 | 0.5 | 0.7 | 0.9 | 1.0 | | 0.3 | 0.6 | 1.0 | 2.0 | 3.0 |
| | Disk [PB] | 0.1 | 0.2 | 0.3 | 0.4 | 0.4 | | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
| | Tape [PB] | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | | | | | | |
| NA62 | CPU [KHS06] | 0.5 | 0.7 | 0.9 | 1.1 | 1.1 | | 1.0 | 1.2 | 1.5 | 1.8 | 2.0 |
| | Disk [PB] | 0.1 | 0.2 | 0.2 | 0.2 | 0.2 | | 0.1 | 0.2 | 0.2 | 0.3 | 0.3 |
| | Tape [PB] | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | | | | | | |
| Others | CPU [KHS06] | 12.4 | 15.4 | 19.2 | 24.1 | 25.3 | | 11.0 | 13.5 | 16.0 | 20.0 | 21.0 |
| | Disk [PB] | 1.1 | 1.4 | 1.6 | 2.1 | 2.2 | | 0.9 | 1.3 | 1.6 | 2.1 | 2.2 |
| | Tape [PB] | 2.3 | 3.5 | 4.7 | 5.8 | 6.1 | | | | | | |
| Operation | CPU [KHS06] | 4.1 | 5.1 | 6.4 | 8.0 | 8.5 | | 3.7 | 4.5 | 5.4 | 6.7 | 7.1 |
| | Disk [PB] | 1.3 | 1.7 | 2.0 | 2.6 | 2.8 | | 1.0 | 1.5 | 1.9 | 2.4 | 2.5 |
| | Tape [PB] | 1.3 | 2.0 | 2.7 | 3.4 | 3.6 | | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Grand Total | CPU [KHS06] | 142 | 177 | 221 | 276 | 290 | | 126 | 155 | 184 | 231 | 244 |
| | Disk [PB] | 14 | 19 | 22 | 28 | 31 | | 11 | 16 | 21 | 26 | 28 |
| | Tape [PB] | 28 | 42 | 56 | 71 | 75 | | | | | | |

*Table-1: UK Resource Requirements.*

## 6.5 Hardware Costing

The WLCG computer model update document also contains a forward look at technology and estimates that there will be a 25% price/performance improvement per year for CPU and a 20% price/performance improvement for disk over the next few years. We have used these assumptions in our calculation of costs, although we have some concerns that there will be a net decrease in the efficiency with which the new multi-core CPU architectures can be used. GridPP data accumulated from 12 years of procurement at RAL show a similar trend (except for a kink in the disk plot that is correlated with a large shift in exchange rates in 2008/9).

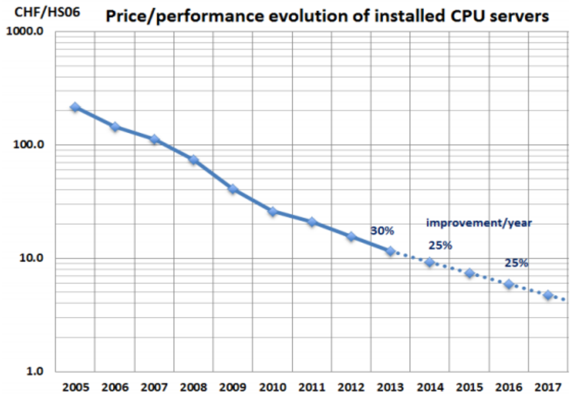



Figure 40: Evolution of price/performance of server systems, example of the CERN computer centre

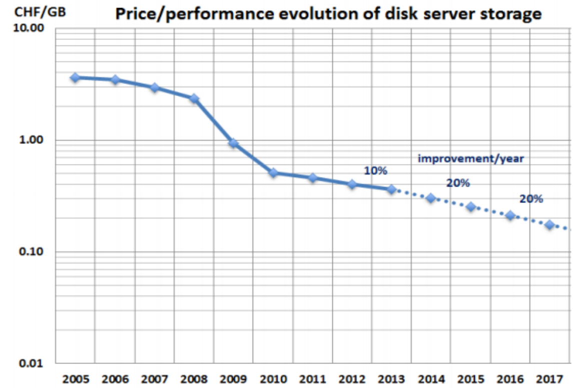



Figure 43: Evolution of price/performance of disk, example of CERN computer centre

*Figure-2: Price/performance evolution of CPU (left) and Disk (Right)*

The resources presented in Table-1 (p11) are costed in Table-2 (below). For disk and CPU we start from our planned 2014 costs and use the CERN assumption that these will decrease by 20%/year for disk and 25%/year for CPU. We also assume a resource lifetime of 4-years for disk and 5-years for CPU (i.e. slightly longer lifetimes than currently, but consistent with the trends we observe). For tape, we assume, based on publically available information together with some information obtained under non-disclosure agreements, that we can manage, for the likely period of GridPP5, with the current robot installation at the Tier-1 and thus that the major cost will be for media. The indicative costing of the tape plan is £1.8m.

Table-2 also contains a line for non-capacity Tier-1 hardware: These are the systems that provide the various services and databases, elements of the internal network and other such overheads. There is a networking line that covers the cost of the OPN (optical-private-network) that connects RAL to CERN. We expect that we will saturate the current links at some point in GridPP5 and that there will be an increase in cost for a significantly larger bandwidth link. At the Tier-2s, the “DRI” funding in 2012 has allowed us to future-proof most sites until about 2015/6. At that point we assume that some investment in upgrading network links will be needed for the following two years. Similarly, the core network at RAL will need to be upgraded with 100Gb/s-capable routers in 2017.

| | FY15 | FY16 | FY17 | FY18 | Total |
|---|---|---|---|---|---|
| **Tier-1 Disk and CPU** | £1.2m | £1.0m | £1.1m | £0.7m | £4.0m |
| **Tier-1 Tape** | £0.5m | £0.5m | £0.5m | £0.5m | £1.8m |
| **Tier-1 Non-Capacity** | £0.3m | £0.3m | £0.3m | £0.3m | £1.1m |
| **Tier-1 OPN Network** | £0.1m | £0.1m | £0.3m | £0.3m | £0.9m |
| **Tier-1 Core Network** | - | - | £0.2m | - | £0.2m |
| **Tier-2 Disk and CPU** | £1.1m | £1.0m | £1.0m | £0.6m | £3.7m |
| **Tier-2 Networking** | - | - | £0.6m | - | £0.6m |
| **TOTAL** | **£3.2m** | **£2.9m** | **£3.9m** | **£2.3m** | **£12.3m** |

*Table-2: Estimated hardware costs for GridPP5 (2015 – 2018).*

We conclude that the cost of hardware for GridPP5 is about £12.3m, though this number probably has a 20% uncertainty due to the uncertainties in future costs and requirements. The profile may need to be adjusted once the running schedule and running conditions are better established.

## 6.6 Network Requirements

***Excellent network connectivity is the bedrock of the GridPP infrastructure***

The UK network requirements for Tier-2 sites are primarily met by the Janet 6 (J6) national research and education network. GridPP, and particle physics in general, maintains excellent relations and frequent interaction with JANET. The new J6 backbone operates at up to 100 Gbit/s and the provision to almost all of the GridPP sites is with multiple 10 Gbit/s connections. In many cases JANET places significant points-of-presence in or near GridPP institutions. The STFC network groups at the Rutherford Appleton and Daresbury labs also ensure that the J6 connections are ahead of demand and, as a result, the Tier-1 is well connected with all other tier sites. GridPP provides, on an annual basis, a formal forward look estimation of network requirements to JANET, who welcomes this level of planning. GridPP also makes use of the J6 dedicated light-path service to connect the Tier-1 site with CERN and the other Tier-1s (the so called LHC Optical Private Network or LHCOPN).

Looking to the future we project that the Tier-1-to-CERN LHCOPN connection will have to approximately double in capacity from 2x10 Gbit/s to at least 4x10 Gbits/s in the GridPP5 era. The RAL site connection to J6 is planned also to increase in line with demand. There will be a modest requirement for some network equipment for the Tier-1 itself, to enable these increased capacities to be connected, and for some equivalent equipment at a few Tier-2 sites.

## 6.7 International Service Commitments

***Resources must be delivered with a well-defined service level***

The requirements on the computing services offered at WLCG tier centres are fully described in the WLCG Technical Design Report (WLCG TDR[9]) and are summarised in the WLCG MoU. In the latter, the service requirements for each of the tier centres are described in terms of target response times to operational problems and meeting these requirements determines the level of manpower requested at the Tier-1 and Tier-2 sites. We note here that the requirements for both high-throughput and high-reliability service provision place extreme demands upon the global network of Tier-1s and requires the ability to deliver a 24 x 7 response at RAL for almost all services. It is important to understand that the UK Tier-1 at RAL forms part of the ‘machinery’ of each experiment, providing common services for the entire experiment. In this sense, the provision of a Tier-1 resource in the UK is no different to the contribution and operation of a sub-detector system. Reliability of the UK Tier-1 is paramount for the operation of the LHC experiments and is an international obligation.

For the Tier-2s, although the service requirements are less demanding than those on the Tier-1 centre, it should be noted that the integrated size of the UK Tier-2s exceeds that of the Tier-1 and, for ATLAS and CMS, a well-founded and well-functioning Tier-2 infrastructure is as much an international obligation and part of the ‘machinery’, as is the Tier-1. In addition, the Tier-2 infrastructure is vital in ensuring that UK physicists can compete for global leadership by extracting the core physics from the LHC.

# 7 Meeting WLCG Commitments and Experiment Requirements

The UK's computing requirements—and the required high level of service—at the Tier-1 and Tier-2s between 2015 and 2018 have been described in the previous section. The need for GridPP to support the wider UK community and to evolve the infrastructure to take advantage of new hardware and

[9] “LHCb Computing Technical Design Report”, lhcc-2005-019.pdf.

software/middleware were also described, together with the need for GridPP to be in a position to start deploying the computing required for Run-3 by 2019.

In this section, we describe how GridPP5 will achieve this. Each work package is addressed in turn and we outline what is required and how it could be delivered with Flat Cash funding. We address the reduced scenario in Section-8.

## 7.1 WP-A: The UK Tier-1

***The Tier-1 provides a flagship international service***

The UK Tier-1 is a mature, effective and reliable service meeting the WLCG's requirements for large, national computer centres with substantial processing and storage capacity and delivering round-the-clock support. Together with the other Tier-1s, it is responsible for the safe-keeping of a proportional share of raw and reconstructed data for the LHC experiments, for large-scale reprocessing and safe-keeping of the corresponding output, for distribution of data to Tier-2s and for the safe-keeping of a share of simulated data produced at these Tier-2s.

To fulfil the formal service level commitments laid down in the WLCG MoU, a Tier-1 must have significant expertise in the management of large storage systems and must provide long-term data archival capability (30 years or more). Data must be actively curated with migrations between generations of tape drive and robot technologies, and with regular assurance that data is still accessible. Tier-1s are expected to operate 24x7 and must respond in a timely manner to issues that occur out of normal business hours. The full service is required to average an availability of 98% throughout the year. An availability of 99% is expected for the acceptance of data from CERN.

The Tier-1s are essential contributors of expertise and infrastructure to support the wider WLCG project. Tier-1s are expected to run a share of global, core services in addition to meeting their national responsibilities. They carry out testing and early deployment of new core services and major upgrades; and contribute staff to lead or support WLCG project teams.

### 7.1.1 The GridPP Tier-1 During LHC Run-1

***A Virtuous Circle: Service Improvements free effort to concentrate on future improvements***

The UK Tier-1 operated extremely successfully in LHC Run-1, meeting its service level commitments and making a substantial contribution to the successful Higgs search. As the LHC Run-1 progressed, the UK Tier-1:

- gained agility and made efficiency savings by automating and simplifying system management;
- evaluated and tested new solutions to deliver increasingly agile operations;
- developed a culture of quality management and increased professionalism, deploying new middleware components without disruption;
- improved communication with the experiments enabling better anticipation of future needs and closer collaboration to resolve problems;
- improved resilience at all levels, reducing the workload of the out-of-hours team and raising overall service availability;
- developed watertight procedures to ensure efficient and reliable operations, minimising wasted effort whilst maximising uptime; and
- carried out an open, continuous service improvement process, bearing down on problems and learning from mistakes.

The result has been a service that has increased in reliability from 92% in 2009 to 99% in 2013, and one that has become substantially more adaptable: Effort has been released from routine tasks in order to address the need to grow and evolve.

Where the Tier-1 has had a strategic interest in new solutions it has taken the lead in global WLCG deployment teams. Taking just recent examples, the team has jointly led the CVMFS[10] deployment project; operated the main FTS3[11] test service; and radically simplified its compute farm setting an example for improvement in turn at GridPP Tier-2 sites.

The improved efficiency and reliability has allowed the service to be operated satisfactorily in steady state at 18 FTE (below the level of 19.5 FTE originally planned for GridPP4), whilst making some progress on various projects. In GridPP5 we expect to be able to achieve almost the same operational level of service with 17.5 FTE, whilst continuing to make slow progress on future developments. We note, however, that at this staffing level progress on Next Generation Disk Storage and IPv6 deployment has not progressed as we would have hoped during GridPP4 and there is a concern that the reduced staffing level requested for GridPP5 will not be sufficient to meet all of the development needs that will be required.

### 7.1.2 The Tier-1 in GridPP5 (Flat Cash Scenario; 17.5 FTE)

***More resources to be delivered and more changes required in GridPP5 than in GridPP4***

The Tier-1 operates around 2,000 separate and interrelated systems running on commodity equipment with the Linux operating system. The core of the service is the EGI supported Grid Middleware providing services such as file catalogues and transfer services, workload management and batch job submission together with the CASTOR Storage Manager (which provides access to disk and tape). Considerable reliability and resilience is necessary at many levels in order to meet WLCG's challenging availability targets, which is achieved by the rigorous application of processes to manage and maintain the service quality. These include monitoring, on-call, change-management, incident reviews, continuous process improvement and others.

The GridPP4 and baseline GridPP5 staffing level for the Tier-1 is 19.5 FTE. In February 2014, as the service prepares for Run-2, the staffing level at the Tier-1 is 19.2 FTE and it is projected to average near that level through 2014. In the Flat-Cash funding scenario effort would be reduced to 17.5 FTE. The rationale for this is that the service was successfully operated at approximately this level during 2013, although at the cost of a lack of progress in a number of important development areas. With the gains in efficiency achieved during the final year of GridPP4 it should be possible to operate the Tier-1 with 17.5 FTE and to make slow progress on necessary developments.

The challenges for GridPP5 are to:

- operate a highly reliable and resilient service able to meet the substantially increased network load and transaction rates of Run-2;
- fulfil our national and international obligations to deliver regional and global services as well as contribute to the collaboration's wider middleware deployment activities;
- meet challenges from necessary evolution of technology;
- evolve the configuration management system and agile infrastructure to reduce operating costs, and better engage (and potentially share costs) with other projects;
- evolve the service gradually towards more open *de facto* interfaces to allow greater collaboration with other large scale, high throughput, big data scientific projects; and to
- prepare to meet the huge data volume challenges of LHC Run-3 and beyond.

Apart from the normal business of day-to-day operation, the following major developments will need to take place in order to stay relevant to WLCG during the period of GridPP5:

- migration of the network and middleware to support IPv6;
- an upgrade of the core network to a 100Gb backbone with increased WAN capacity;
- introduction of a further tape drive generation into operation;

[10] CVMFS is a wide area network file-system which removes the need for sites to run local fileservers to host experiment software repositories.

[11] FTS provides a reliable data movement service for transferring files between Grid Storage Elements.

- deployment of a new disk storage system to meet the projected high transaction rates;
- consolidation of the CASTOR tape service (and back-end databases) to reduce cost of operation;
- deployment of a production quality private cloud service to further increase agility and follow *de facto* standards;
- investigation of new commodity hardware solutions to deliver cost savings; and
- energy efficiency improvements to deliver a reduction in electricity usage at the Tier-1.

Many of these activities represent major projects that will directly and indirectly benefit other UK research infrastructures and, judging from the experience gained during GridPP4, will require a significant input of sustained effort.

The Tier-1 team structure and detailed effort breakdown to deliver the Tier-1 at 17.5 FTE is presented in the following table:

| Team | 2013 Staff Level | Proposed Effort |
|---|---|---|
| Storage Team | 3.3 | 3.1 |
| | Maintain and operate the CASTOR disk and tape pool management system that consists of 5 production CASTOR instances (ATLAS, CMS, LHCb, a shared CASTOR instance for small VOs and a background instance for tape repacking). There are 30 separate core servers and a further 35 tape servers. The team manages the tape robot and procures hardware such as media and drives when necessary. They provide close support to the different experiments optimising the configuration to meet the experiments' requirements. They carry out performance tuning and fault diagnosis in close liaison with the database team at RAL, and in conjunction with CASTOR and database staff at CERN. In order to minimise upgrade problems, extensive testing is carried out on two further CASTOR instances (test and certification). The Storage team will also manage and operate the "Next Generation Disk Store" once this is ready for production. | |
| Middleware Team | 4.0 | 4.0 |
| | Maintain and operate the Tier-1's Grid and Cloud interfaces together with other application layer tools and services. These provide many mission-critical Grid services, such as BDII, LFC, FTS, MyProxy, CEs and CONDOR, WMS, CVMFS, Frontier, Squids and ARGUS. Several of these are critical for LHC production activities for the whole of the UK. About 30 distinct node configurations are deployed across about 65 hosts, many in resilient and VO-specific instances. Manage and develop the QUATTOR configuration management system. A testbed to deploy phased rollouts of new middleware components is maintained (about 50 hosts). The team also manage the service resources, experiment shares, pledges and resource accounting. They work closely with the experiment support staff to provide the experiments' Tier-1-specific infrastructure such as the VO boxes. | |
| Database Team | 1.5 | 1.5 |
| | Support the Oracle databases for CASTOR, the 3D conditions databases, the File Transfer Service (FTS) and LFC file catalogue. This presently requires 132 schemas running on 9 databases hosted by 24 database servers distributed over 7 Oracle RACs and 4 single instances. The team are responsible for: the routine operation and maintenance of the databases; performance tuning and debugging; upgrade testing, resilience and recovery exercises (all crucial for such critical systems); and new service development. They directly manage the 3D conditions streaming service from CERN. | |

| | | |
|---|---|---|
| Production team | 2.9 | 2.9 |
| | Ensure that WLCG service level commitments are fully met, including appropriate communications on operational matters with CERN and the experiments. The team have responsibility for planning and coordinating scheduled interventions and for managing the response to operational incidents and service exceptions. During the working day the team provides an Admin on Duty (AoD) whose task is to monitor continuously the operation of the service, receiving high level exceptions via pager, and who tracks lower level issues via a variety of dashboards and other monitoring tools. The AoD also triages RAL Tier-1 GGUS trouble tickets and end user queries ensuring prompt response. Maintain and develop exception and performance monitoring tools such as nagios, ganglia, the dashboard, and the local helpdesk. Operationally responsible for managing security incidents. Routine operations on the batch farm, including: managing rolling operating system and middleware updates; dealing with day-to-day batch scheduling; node and service exceptions. Day-to-day operation of the virtualisation infrastructure, load-balancing servers, handling exceptions etc. Responsible for the on-call processes and workflow itself and for driving forward the Tier-1's continuous improvement processes. | |
| Fabric Team | 5.3 | 5.0 |
| | Manage the large, high performance local network consisting of about 200 network switches together with the management and monitoring layer. Support the specialist database hardware and storage area network. Responsible for the maintenance of the large scale commodity hardware infrastructure and Linux operating system. Physically maintain and support about 700 batch workers and 400 disk servers, about 10,000 disk drives. Manage maintenance and warranty. Diagnose hardware faults, manage engineer callouts or repair locally when necessary. Manage the core infrastructure systems, file servers, authentication servers and consoles. Provide fabric support for the other teams' specialised servers. Manage and develop the HyperV virtualisation layer. Carry out the annual hardware procurement cycle and tender evaluation; manage the physical rack infrastructure and local power distribution, carry out installation and commissioning of the hardware and at end-of-life arrange disposal. | |
| Management | 1.0 | 1.0 |
| | The Tier-1 Service Manager is responsible for the delivery and operation of the Tier-1 centre ensuring it meets MoU commitments to GridPP and WLCG. The Tier-1 service is a large and complex facility. Effort is needed to project manage the internal service, liaise with GridPP and also maintain high-level links with WLCG and other peer organisations. The Tier-1 is required to routinely report project management information to STFC, the Scientific Computing Department and GridPP. The manager is responsible for the Tier-1 finances and staff effort, managing recruitments and oversees procurements when necessary. | |
| **Totals** | **18.0** | **17.5** |

*Table-3: The Tier-1 team structure and roles.*

### 7.1.3 Host Laboratory Costs

***Reducing cost of providing the underpinning infrastructure for the Tier-1***

Staffing levels at RAL for infrastructure support funded by the laboratory have fallen gradually from 4 FTE at the start of GridPP4 to 2.1 FTE currently. The reduction in support levels reflects reduced machine room operations costs, efficiency gains as reliability improved, and the end of CASTOR software development work as the product has matured. Staffing levels for machine room and network support will fall yet further, reaching 1.65 FTE in GridPP5 as the Tier-1's footprint in the machine room falls from 60% to 40% as other major Scientific Computing Department (SCD) services expand.

Electricity costs are estimated based on the expected STFC 2014 electricity price, projected forward using the average electricity price inflation seen at RAL since 2008 (5.9%). The total cost of electricity over the period 2015-2018 is estimated to be £2,204K, similar to the cost during GridPP4 (£2,229K).

## 7.2 WP-B: The UK Tier-2s

***Delivering a distributed infrastructure, levering additional resources, and supporting users***

Over the last 12 years, the UK has established four successful distributed Tier-2s (ScotGrid, NorthGrid, SouthGrid and the London Grid). This organisational structure has been used to inform models elsewhere in WLCG, for example in France and Germany. The strength of this model is that it combines the best attributes of local facilities (such as on-site user support and local engagement) with the advantages of larger managed facilities (such as quality of service and access to expertise). In addition, its distributed nature, combined with the fact that GridPP funds hardware resources in proportion to past delivery, has resulted in contributions of additional resources and support worth millions of pounds from the institutes involved. As discussed in Section 4.2, the hierarchical Tier-0/1/2 model has been changing and will continue to evolve during GridPP5 with ATLAS and CMS increasingly relying on Tier-2s to provide a wider range of services than just analysis.

The UK Tier-2s currently provide:

- CPU and disk resources to meet the UK's WLCG Tier-2 MoU requirements;
- manpower to support Group Analysis Sites, with large amounts of disk and excellent network connections, for ATLAS and CMS;
- manpower to provide all experiments with CPU and disk for opportunistic user analysis and Monte Carlo simulation;
- a distributed ecosystem to support UK physicists doing their analysis;
- the majority of the Deployment, Operations, and Support personnel who transform the distributed collection of resources into a coherent Grid infrastructure (Section 7.3);
- hardware resources for testing middleware releases, new computing approaches and for running some core infrastructure services;
- a successful framework for leveraging local resources and support; and
- opportunities for reaching out to other communities in the universities and beyond to support STFC's impact agenda.

### 7.2.1 Tier-2 Requirements of the LHC Experiments

**ATLAS –** During GridPP4, ATLAS has made extensive use of Tier-2 resources for simulation, group analysis and individual user analysis. There is an increasing use of major Tier-2s in dynamic data placement and reprocessing activities; the latter will be a more crucial role in GridPP5, following the revised computing model.

ATLAS analysis is organised into eight working groups (Standard Model, Higgs, SUSY etc.), supported by six combined performance groups (Tracking, e/γ, Jets/Missing $E_T$ etc.). In order to balance loads and data throughput, a region with the UK's available resources is expected to host at least ten of these sixteen groups. The optimal configuration in GridPP5 would thus be to maintain the five major UK ATLAS sites (known as T2Ds, Tier-2s with Data), all with good network connectivity to Tier-1 centres around the world, each supporting two of the analysis groups together with other sites running less demanding, more CPU-bound work. The major sites would host pre-placed real and simulated data and also large caches for on-demand replicas for load balancing and optimisation. Additional space for some groups at the Tier-1 would aid the data distribution activity.

In GridPP4, two FTE were allocated to each of five major ATLAS sites in order to provide continuity of operations and support. Analysis requires large amounts of disk with excellent tuned bandwidth to the local CPUs and good inbound and outbound network connections to other sites plus many additional services. These sites must be highly responsive to the needs of the supported experiments and their computing teams. If a CPU fails then it can be changed some time later, but if a disk fails it has to be replaced almost immediately and possible data-loss addressed.

ATLAS globally has investigated distributed computing implementations other than the current Grid model, including the limited use of commercial cloud resources and the opportunistic use of HPC resources. While most workflows have been tested and made to work at some level, cloud resources are only feasible and efficient for about 10% of the total work, and HPC resources for even less. The transfer to the Tier-2s of workflows that were formerly done at the Tier-1, such as reprocessing, reinforces the conclusion that the current distributed Tier-2 structure remains a balanced and resilient solution for the ATLAS Tier-2 requirement.

Although the five major sites are staffed by 10 FTE, about 25% of this effort contributes directly to deployment, operations and support of the Grid infrastructure, as described in Section 7.3, which benefits not just ATLAS but all the user communities and enables the operation of smaller sites with reduced levels of effort. This distribution of staff also provides the appropriate level of direct support for the analysis groups hosted, and often enables engagement with local interests. Looking at the global picture, ATLAS has estimated that a region the size of the UK would require a minimum of 10.5 FTE to run an adequate Tier-2 infrastructure and, in fact, the UK has one of the smallest levels of deployment and operational staff, despite being one of the largest ATLAS regions.

**CMS -** Within CMS the primary role of the Tier-2s is end-user analysis, which does not take place at the CMS Tier-1s. A secondary role, shared with the Tier-1s, is Monte Carlo Production. GridPP supports three CMS UK Tier-2 sites (RAL PPD, Brunel and Imperial) to perform group analysis tasks associated with five CMS analysis streams (which are similar to the ATLAS analysis groups). These streams or groups reflect the UK's physics interests and include the high profile Higgs and SUSY analysis streams. Distributing these streams across the three sites provides resilience against the loss of a single site; a permanent reduction to just two sites would remove this resilience. As with the ATLAS analysis sites, the CMS sites were supported with 2 FTE to ensure continuity of operation and support (2 FTE funded at Imperial; 1.5 funded at RAL PPD was complimented by 0.5 FTE funded by EGI; and 1.5 FTE funded at Brunel was supported by an addition 0.5 FTE of institute-funded support). The CMS Tier-2s are now taking on additional roles and may even perform prompt reconstruction, currently a Tier-0 task. One example of this was the recent re-reconstruction of the 2011 data in 2013. Traditionally a Tier-1 task, this was shared between all the tier levels, as shown in Figure-3.

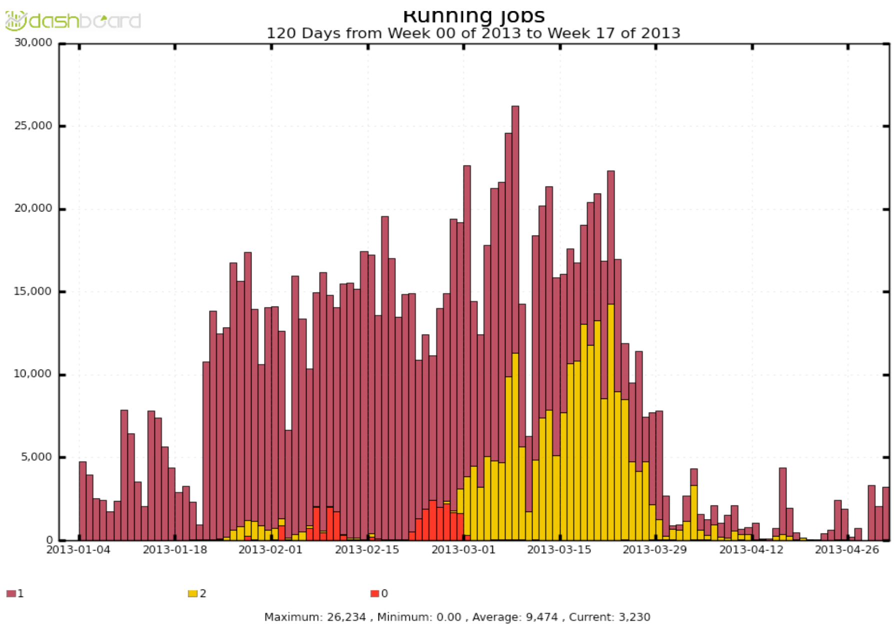



*Figure 3: 2013 reprocessing of the 2011 CMS datasets at the different tiers (Tier-0 = red, Tier-1 = burgundy, Tier-2 = yellow)*

**LHCb -** In 2011 LHCb extended its use of Tier-2 sites to make use of some of the larger, and most reliable, sites for data reconstruction. These so called 'attached' Tier-2 sites drew data from a suitable Tier-1 site, reconstructed it, and then transferred the results back. In 2013 LHCb further extended its use of Tier-2 sites through the introduction of so-called "T2D sites" with dedicated disk resources. Operationally this means that a fraction of the reconstructed (DST) data is now stored at T2Ds and is available for re-processing and analysis in exactly the same way as was previously done only at Tier-1 sites. The primary motivation for T2D sites was to increase contributions from non Tier-1 hosting countries but an important benefit has been to allow countries such as the UK to deliver resources more flexibly, improving the efficiency with which resources at the Tier-1 and the Tier-2s can be exploited and making better use of the available capacity.

For the T2Ds in the UK, LHCb focused on sites that already had adequate effort to support ATLAS and CMS. As the strategy to make more use of Tier-2s develops in GridPP5, LHCb will need a modest increase in support, but this should be deliverable with the effort currently deployed if the distribution and assigned responsibilities of the Tier-2 staff are carefully managed.

### 7.2.2 Current Service Provision

The larger Tier-2 sites were reviewed at the mid-point of GridPP4 and a modest redistribution of effort was performed, in recognition of those sites that had performed best and those that had provided additional resources from external funding. Table-4 summarises a snapshot of the resources at the UK Tier-2s towards the end of 2013 before the current round of procurements started. The colour coding show large sites in green, medium size sites in yellow and small sites in blue (in terms of the amount of disk). The FTE-1 and FTE-2 columns show the amount of effort funded by GridPP4 in the first and second half of the project respectively, with changes highlighted in red.

| Site | CPU [HS06] | Disk [TB] | GridPP FTE-1 | GridPP FTE-2 |
|---|---|---|---|---|
| **Brunel** | 11,600 | 484 | 1.5 | 1.5 |
| **Imperial** | 26,900 | 2,026 | 3.0 | 3.0 |
| **QMUL** | 29,800 | 1,680 | 2.0 | 2.0 |
| **RHUL** | 14,500 | 640 | 0.5 | 1.0 |
| **UCL** | 2,600 | 189 | - | - |
| **Lancaster** | 25,298 | 1,032 | 2.0 | 2.0 |
| **Liverpool** | 11,129 | 550 | 2.0 | 1.0 |
| **Manchester** | 22,809 | 883 | 1.0 | 2.0 |
| **Sheffield** | 5,476 | 360 | 0.5 | 0.5 |
| **Durham** | 8,140 | 50 | 0.25 | 0.25 |
| **Edinburgh** | 17,200 | 355 | 0.5 | 0.5 |
| **Glasgow** | 31,600 | 1,300 | 3.0 | 3.0 |
| **Birmingham** | 8,121 | 320 | 0.5 | 0.5 |
| **Bristol** | 2,247 | 105 | - | - |
| **Cambridge** | 1,600 | 275 | 0.5 | 0.5 |
| **EFDA Jet** | 1,772 | 11 | - | - |
| **Oxford** | 11,770 | 650 | 2.0 | 1.5 |
| **RAL PPD** | 25,975 | 1,260 | 1.5 | 1.5 |
| **Sussex** | 2,365 | 50 | - | - |
| **Total** | **261,000** | **12,200** | **20.75** | **20.75** |

*Table-4: Tier-2 site resources (snapshot at the end of 2013) and site staff effort in the first half (FTE-1) and second half (FTE-2) of GridPP4.*

Some of the UK Tier-2 sites are now very large but are still supported by only 2 FTE. In addition, 10 of the 20.75 posts contribute effort (estimated to be about 6 FTE) to form the essential backbone of the GridPP Operations Team, which is described more fully below, and to engage with the central experiment computing operations. One post at each of Imperial and Glasgow assists some of the smaller sites where there is little or no effort at all. Another key role of the Tier-2 staff is to engage with potential users: almost all of the non-LHC VOs now using GridPP resources became Grid users through contact with, and help from, Tier-2 support staff.

### 7.2.3 Leverage and Delivery in GridPP4

A key advantage of having the UK Tier-2s distributed across almost all of the institutes participating in experimental particle physics is that it enables GridPP to leverage resources, operating costs, knowledge and manpower from a variety of difference sources, such as university computer centres, research infrastructure funding, and local expertise. Similarly, it provides local physicists with ready access to Grid expertise and provides a conduit for the dissemination of information and the engagement with diverse communities outwith particle physics.

Although this leverage is difficult to quantify exactly, a survey[12] of all sites was undertaken in 2009 and again in 2013 and an estimation of the annual leverage is shown below.

| | **Annual Leverage** | |
|---|---|---|
| | **2005-2009** | **2010-2013** |
| **Capital Costs** | £900K | £700K |
| **Hardware** | £650K | £700K |
| **Electricity** | £500K | £650K |
| **Manpower** | £400K | £450K |
| **Total** | £2,400K | £2,400K |

*Table-5: Estimated annual contributions from the Tier-2 host institutes obtained from surveys covering 2005-2009 and 2010-2013.*

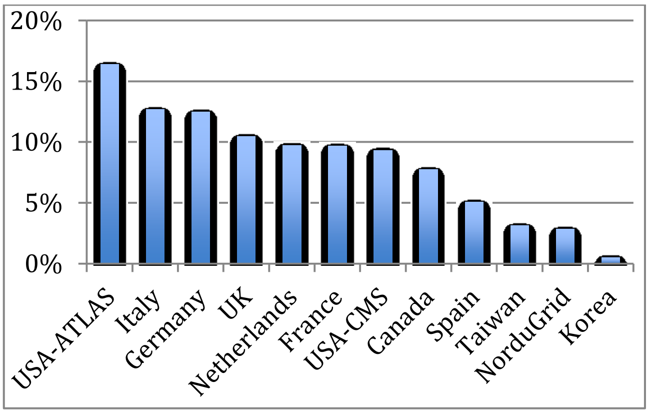
It is clear that the Tier-2 institutes have been able, over a sustained period, to leverage very significant resources above those that STFC, through GridPP, has been able to provide. In the absence of capital available from the SRIF and CIF schemes in the earlier period, the capital has fallen somewhat but electricity costs have increased leaving the total constant at around £2.4m/year for the last 9 years. This leverage will scale not only with the level of resources located at a site but also with the number of sites that contribute. To date, it has been spread across all 17 institutes, which has allowed much of it (such as electricity costs) still to be at a level that institutes are happy to provide. If the number of Tier-2 sites goes down then the leverage will almost certainly go down significantly, not only due to the loss of sites but also because electricity costs will become more visible as resources would be concentrated at fewer places and are more likely to be charged.

In summary, any reduction in the number of Tier-2 sites, therefore, is likely to reduce future leverage significantly, with investment in server rooms wasted; physicists losing local support and becoming less efficient; and lost opportunities for both outreach and impact activities.

Another demonstration of the impact of leverage obtained from Tier-2 sites is clearly visible in Figure-4b, which shows that, at 15%, the UK was the largest contributor of Tier-2 resources to WLCG after the US. This compares with the Tier-1 contribution shown in Figure-4a, of 10.4%. The difference in the

[12] Details are given in the background documents referenced on page-1.

fractions is due to the delivery of leveraged Tier-2 resources, which has been of direct benefit to the LHC experiments and the UK LHC community. It is a concrete example of the effectiveness of the GridPP strategy for the delivery of Tier-2 resources.



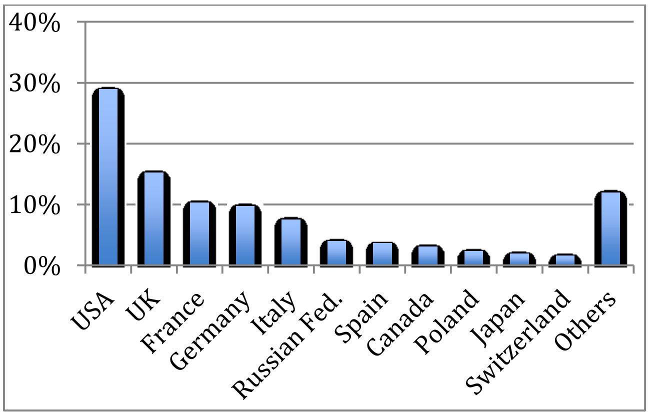
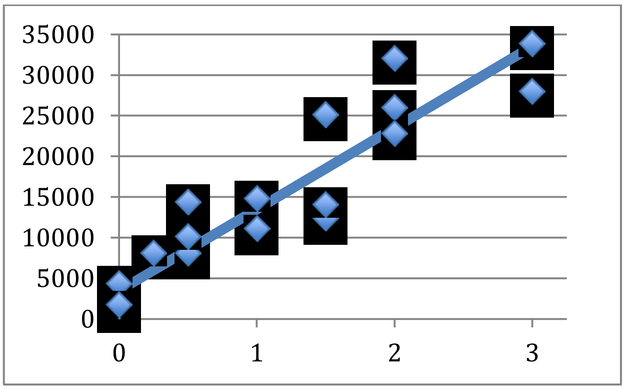




*Figure-4: (Left) Tier-1 Resources delivered in 2011-2013 and (Right) Tier-2 resources delivered over the same period, by country.*

### 7.2.4 The Tier-2s in GridPP5 (Flat-Cash Scenario – 19.5 FTE)

***Balancing economies of scales with the leverage, support and engagement of a distributed Grid***

During GridPP4 site performance was closely monitored through a number of key metrics and both the staff levels (at the mid-point) and the hardware grants were adjusted accordingly. We would expect to repeat this exercise at the mid point of GridPP5 to ensure that manpower remains optimally deployed. At a high level, the staffing resources at each site broadly reflect the resources delivered, but as noted earlier, there is also a distinction between the difficulty of running a designated analysis site, and the easier case of running a site concentrating on Monte Carlo production work. Figure-5 shows a snapshot of the correlation between Tier-2 site manpower and the resources provided at the end of 2013.



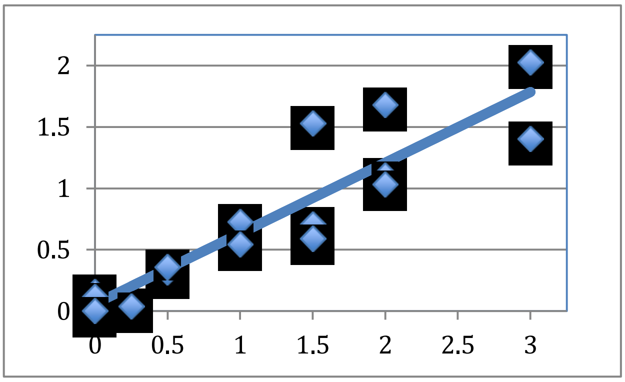




*Figure-5: Staff effort (FTE on horizontal axis) scaling with CPU (HS06, left) and disk (PB, right) at the Tier-2s*

Using an arbitrary but sensible weighting between disk and CPU, each FTE of Tier-2 effort effectively delivers something like 625TB of disk and 12,500 HEPSPEC06 of CPU. For comparison, the Tier-1 capacity of about 10PB of storage and 100,000 HEPSPEC06 would translate to 12 FTE. Of course, it is not possible to directly compare these numbers because, currently with 19.2 FTE, the Tier-1 delivers the custodial tape storage plus many other services in addition to the disk and CPU at a higher level of service. Nevertheless, it suggests that the distributed Tier-2 system is efficient, especially when one considers that many of the Tier-2 staff have additional national and international responsibilities for deployment, operations and support (see Section 7.3.1.1). This type of algorithm will be the starting point for allocation of Tier-2 manpower in GridPP5, but the final allocation will also take into account the designation of analysis roles for some sites; future site plans; and any additional site staff roles within GridPP and WLCG. In the Flat-Cash scenario, Tier-2 staffing will need to decrease by 1.25FTE and at least one additional site (compared to GridPP4) is likely to receive no funded effort.

In summary, the Tier-2 sites will need to deliver 2-3x the integrated resources during GridPP5 compared to GridPP4 and take on new workflows as the LHC community better exploits the provided resources. The infrastructure will need to continue to evolve, matching the computing models and taking advantage of advances in technology. We have shown that the leverage obtained from the Tier-2 sites is substantial and sustained; we hope it is clear that cutting the number of sites with funded staff risks incurring large capital costs and significant recurrent costs for electricity, as well as impairing the substantial second-order benefits such as local support for physicists; opportunities for engagement with new partners to respond to the STFC impact agenda; and resilience of the infrastructure.
Given these potential risks, we do not consider a reduction in Tier-2 staffing levels beyond the 10% we propose, to be sensible. We hope it has been demonstrated that the integrated level of staff delivering the expected UK Tier-2 contribution to WLCG, is cost effective, especially given that these staff have significant national and international roles as will be further substantiated in Section-7.3.1.1 below. In the Flat-Cash Scenario we tread the line between economies of scale and the benefits of a distributed system.

## 7.3 WP-C: Grid Deployment, Operations and Support

***Transforming the collection of UK Tier-1 and Tier-2 resources into a Grid infrastructure***

To transform a collection of distributed resources into a coherent Grid requires common middleware, security policies, accounting, monitoring and support. These must be applied across geographically separated sites running a heterogeneous collection of hardware managed under local policies, but in a way that is consistent and compatible with the international WLCG context. To keep the Grid running requires active management, careful scheduling of deployment and upgrade priorities, excellent communication, and a responsive support system. The essential tasks are performed by some of the Tier-2 staff described in Section 7.2, augmented by experts in specific key areas.

Deployment covers the activities that integrate the hardware in physically distributed machine rooms into a usable distributed infrastructure. This starts at the fabric level with the implementation of cluster management tools and then moves on to the generic services, such as batch scheduling and service monitoring, and finally onto the higher level services and experiment-specific middleware and software. Deployment is a multifaceted and essential activity. GridPP works with partner projects, such as WLCG and HEPiX, and has often been at the leading edge of such work being, for example, the first to fully implement technologies such as CVMFS and have sites transitioned to SL6. This leadership has allowed GridPP to benefit more quickly from the new technologies, mirroring the 'virtuous circle' of progress at the Tier-1 described in Section 7.1.1, wherein effort invested in improving the system reduces the operational effort, which in turn allows further effort to be invested in additional improvements.

Operations describes a wide variety of work which ensures that the deployed distributed infrastructure is usable and functioning optimally. This is achieved through various levels of monitoring, tracking of problems and pre-empting issues. When major interventions are required, the operations team liaise with sites to plan and schedule work to minimise disruption to users and are constantly on the alert for security vulnerabilities and incidents, developing and implementing responses as appropriate.

Support is a key function of any successful production service. For the Grid it is sub-divided into operational support, deployment support and user/VO support. The formal support process uses tickets submitted via a helpdesk that are triaged and assigned to people within the project for follow-up. The GridPP site staff, operations team, and area experts, respond to a wide range of support tickets.

As GridPP operates a computing infrastructure our 'users' are, to first order, the experiments rather than individuals; that is, GridPP supports a range of workflows and services for each experiment. GridPP has three experiment support posts (one for each of ATLAS, CMS and LHCb) based at the

Tier-1 that provide operational support and maintain a formal channel of direct communication with the experiments. The post holders have an expert understanding of the individual experiment computing models and workflows from an implementation perspective.

Although GridPP's main work is in support of the experiments, direct user support does also take place in order to host specific physics groups, individual and group analysis work, and outputs at a Tier-2. Staff at GridPP sites work with local users to help them get the most out of the infrastructure, and where necessary escalate issues to operations meetings - these meetings have standing agenda items to review experiment and user problems and issues. At first it can be unclear whether a problem lies with the user, the experiment, or the Grid domain, and local expert help is invaluable in identifying where the problem lies.

## 7.3.1 The Deployment, Operations and Support Tasks

### 7.3.1.1 Core Team Tasks

The core Operations Team (ops-team) is led by the GridPP Production Manager supported by eight Tier-2 staff (one from each of the sites supporting experiment analysis activities) who lead the tasks identified below. EGI funded effort, which is now coming to an end, also contributed to some areas. The core tasks are:

**Staged rollout** – Updated middleware has to be extensively tested, integrated and documented before it can be widely deployed. GridPP contributes to this international task as part of our responsibilities to WLCG. This provides the UK Tier-2 staff with expert knowledge and access to the latest advances.

**On-duty** – One person (rotated on a weekly basis) undertakes a series of tasks each day, such as reviewing dashboards and raising tickets against observed problems. This is proactive work to maintain the quality of the service.

**Ticket follow-up** – There are typically 25-50 live tickets at any one time across GridPP sites, ranging from user support requests to requested infrastructure updates. One member of the Operations Team takes responsibility for reviewing open tickets on a regular basis and the Operations Team as a whole contributes ideas to resolving more difficult problems. This is reactive work to maintain the quality of the service.

**Regional tools** – The staff who run the Tier-2 sites rely upon a number of tools and services to spot and diagnose problems. One member of the team is responsible for ensuring that the monitoring framework is running and up-to-date, and represents GridPP internationally in discussions aimed at evolving the tools. The Operations Team also work to provide new tools to assist Tier-2 personnel and users to increase their productivity.

**Security** – Maintaining a secure infrastructure is critical, and several Operations Team members support the GridPP officer in this activity. The team meet on a bi-weekly basis to agree strategies for the UK as a whole to follow (for example in coordinating rollout of authentication services), to review and follow-up on incidents, to form a view on vulnerabilities, and to discuss and recommend updates to security policies. This security activity needs to be carried out in an international context. The UK Security Officer is required to perform regular weekly duties as part of an operational rota at the EGI CSIRT level, this CSIRT being the operational security team used by WLCG in Europe. In the event of an incident the team will ensure a response coordination structure is in place. Other tasks include training, checking site patching and reviewing automated security test results.

**Documentation** – Large quickly evolving projects produce varying qualities of documentation that can quickly become obsolete. The Operations Team assigns responsibilities for documents across its members and also provides tools to encourage regular review of critical documentation. One member of the team regularly reviews documents and follows up on missing or out-of-date areas.

**Monitoring** – GridPP Operations Team members have helped develop monitoring tools and take a part in a WLCG monitoring consolidation task force aimed at simplifying and improving the monitoring infrastructure as part of our responsibilities to WLCG.

**Accounting** – Usage needs to be tracked for GridPP, WLCG and EGI. The integrity of usage data needs to be constantly monitored and resources need to be benchmarked periodically. For example, the upgrade from SL5 to SL6 changed the performance by varying amounts depending on the hardware specification.

**Core services** – Various national services must run to support the infrastructure. Some, such as the perfSONAR network monitoring are run at every site. Others, such as the VO management service (VOMS) and the workload management system (WMS) are run at one or a few sites. The effort to deploy, run, maintain, and develop these services come mainly from the Operations Team.

**Interoperation** – GridPP sites rely upon middleware developed within international projects such as WLCG and EGI/EMI, and within the experiments. Core Operations Team members attend relevant meetings and workshops to coordinate the UK participation.

**Wider VO support** – GridPP endeavours to support other communities outside of its core LHC VO user communities. These range from HEP organisations to other academic disciplines through to industrial partnerships. This core-task explicitly drives liaison with these communities, providing advice on tools and techniques, and follows up on any issues encountered.

#### 7.3.1.2 Specialist Tasks

The core tasks listed above, performed by the core Operations Team that draws upon a subset of the personnel at the Tier-2 sites, are complemented by a number of explicitly funded operational, deployment, and support tasks undertaken by specialists:

**Experiment Support Posts** – The three experiment support posts (for ATLAS, CMS and LHCb) have ensured excellent communication and engagement at the coordination and technical levels. These posts have acted as the primary channel by which GridPP can understand operational problems experienced by the experiments and are a key mechanism to ensure that GridPP delivers and evolves the infrastructure in the manner required by our primary clients. These posts have ensured appropriate UK involvement as the experiments have evolved their computing models and have given guidance for strategic developments at the Tier-1. The posts have been invaluable in communicating issues to the experiments and driving their resolution. As we move towards Run-2, and as the computing models place increasing emphasis on the role of Tier-2 sites, these experiment support roles must expand in order to complement the embedded Tier-2 effort, bringing their wealth of experience to those Tier-2 sites that take on more responsibilities.

In GridPP4, there have also been two additional posts, one to support the non-LHC experiments with a particular role in assisting them to get started on the Grid (for example, NA62 and MICE), and one to support the Ganga interface used by ATLAS, LHCb and non-LHC VOs.

**Data storage and management** – The role of the storage and data management team is to support the infrastructure for storing and moving data in the UK. This includes troubleshooting; the testing of new releases to ensure upgrades do not affect the services; and performance tuning. These roles underpin the ability of GridPP to run small sites with a small amount of effort. GridPP's storage and data management group has been highly successful in supporting a "Big Data" infrastructure and have also led the WLCG storage "evolution" task force. Their work is often presented at HEPiX and CHEP meetings. Internationally recognised, the GridPP's storage and data management support list now contains approximately 80 Tier-2 site members in 10 different countries. The core team is split across three-and-a-half posts, whose duties, in addition to the core duties of supporting storage at sites and responding to incidents, are:

1. Data storage support: Supporting filesystems.
2. Investigating next generation storage to optimise cost, performance and durability in the future;
3. DPM support: Disk Pool Manager is used at the majority of the Tier-2 sites, GridPP contributes effort as part of an international collaboration supporting this storage middleware. Operational support is provided to sites and the UK has developed an administrative toolkit for the maintenance and troubleshooting of a DPM installation;
4. Data transfer support, Tier-2 coordination, and experiment liaison: This work optimises, tunes and monitors GridPP network transfers, supporting data transfers from Tier-2s to the RAL T1, or to Tier-1s or other Tier-2s outwith the UK.
5. Data support for non-LHC VOs: Developing techniques and workflows for data movement and data management.

**Operational Security –** The GridPP security officer advises upon and manages security risks that may threaten the availability of GridPP services or generate negative publicity, and participates actively in the operational security team used by WLCG in Europe (EGI CSIRT). Aside from duties as part of the regular weekly operations rota within the EGI CSIRT, the tasks include the investigation of security issues and formulating the GridPP/WLCG/EGI responses; for example, the assessment of cloud security models. As discussed above, the GridPP security team members assist the Security officer in the operational elements of his role.

**Security Policy and International Coordination** – Distributed computing infrastructures rely on trust being maintained between institutes, countries, providers, users and developers. This trust is established and maintained by well-thought-out security policies and by appropriate security operations and procedures. As the infrastructure evolves and new technologies are deployed it is essential to further develop, evolve and disseminate security policy in conjunction with national and international partners. GridPP has been a leader in this role for many years and funds 0.5 FTE in this area. All of the current security teams used by WLCG (security operations, vulnerability handling, policy coordination, IGTF/EUGridPMA for Authentication) were proposed and started by the GridPP security team and this has allowed GridPP, at many levels, to ensure that developments meet the needs of the UK.

**Security Vulnerabilities** – Unless proactively addressed, software vulnerabilities can lead to security incidents and service failures. The security vulnerabilities group reviews vulnerabilities as they are identified and issues clear distributed alerts that are key to securing the infrastructure. Site and developer progress towards addressing the vulnerabilities are tracked. GridPP created this role as a major contribution to WLCG/EGEE and EGI but it is now funded by EGI.eu/EGI-InSPIRE, not GridPP. The work is a UK responsibility for the benefit of WLCG but it is not yet clear how this required work will be continued in GridPP5 as it would increase the portfolio of funded tasks against a shrinking budget.

**GOCDB maintenance** - Developing and maintaining the Grid Operations Centre Database (used by all of WLCG and EGI, and being adopted more widely) has been a UK contribution to WLCG. The GOCDB provides an essential repository of relatively 'static' information that helps to identify sites, their contacts and services. It is used for publishing outage information and is actively queried by many Grid services. GridPP4 funded 0.5 FTE matched by EGI. There is no assurance that the latter funding will be available in the GridPP5 era.

**APEL maintenance** - the Accounting Processor for Event Logs (APEL) accounting is used in multiple ways by WLCG and EGI, from checking how resources are being utilised across countries and experiments, to comparing site contributions with MoUs. The UK developed and continues to maintain this distributed software infrastructure as part of an international contribution to WLCG. GridPP4 funded 0.5 FTE matched by EGI. There is no assurance that the latter funding will be available in the GridPP5 era.

**UK Certificate Authority** (CA) – The CA maintains a signing service that acts as part of a global trust federation to authenticate UK users. GridPP depends upon this service, originally developed by GridPP, but now operated and funded by STFC.

**Networking** - The distributed data distribution within WLCG computing, coupled with the high i/o demands of many tasks, has led to a consistent need to review site LAN capability and configurations, institute connectivity to JANET, and wider considerations about network peering. Work done within this area extends beyond the core Operations Team and covers areas such as a rolling network forward look, network monitoring, and the evaluation of technologies such as IPv6. Better than expected network connectivity is one of the driving factors behind improvements in the experiment computing models and networking as a core operational area is of increasing importance. The area task has for much of GridPP4 been covered within the Operations Team under Core Services.

## 7.3.2 Deployment, Operations and Support (Flat-Cash Scenario – 9.25 FTE)

As described in the previous section, deployment, operations and support comprises many individual tasks. The core tasks are performed by the Operations Team using effort funded at the larger Tier-2 sites. This is estimated as 6 FTE out of a total of 20.75 FTE funded at the Tier-2s in GridPP4. In the GridPP5 Flat-Cash scenario, we propose to reduce the effort at the Tier-2 sites to 19.5 FTE, which may result in a small loss of effort available for the core tasks. The other Deployment Operation and Support tasks are currently performed by 12 FTE funded by GridPP4 and by 4 FTE funded from other sources (see Table-6). In the Flat-Cash scenario, we propose to protect at least three out of the five experiment support posts and make balanced reductions to the other tasks to reduce the total support from 12 to 9.25 FTE. However, since this area currently depends partly on external funding, the future of which is not yet established (for example H2020 opportunities), it is not possible to specify the exact distribution of future effort in the right-hand columns of Table-6: Any reduction in support for international responsibilities would have to involve negotiations with WLCG and other partners. It is also extremely probable that external funding will be reduced.

| | **Current Funding** | | **Future Funding** | |
|---|---|---|---|---|
| | GridPP4 | Other | GridPP5 | Other |
| Production Manager | 1.0 | | | |
| Experiment Support | 5.0 | | >=3.0 | |
| Data Management | 3.5 | | | |
| Security | 1.5 | 1.0 | | |
| GOCDB | 0.5 | 0.5 | | |
| APEL | 0.5 | 0.5 | | |
| CA | 0.0 | 1.5 | | |
| Networking | 0.0 | 0.5 | | |
| **Total** | **12** | **4** | **9.25** | **??** |
| Core Ops Team Effort | ~6.0 | | ~~6.0 | |

*Table-6: Deployment operations and support effort (FTE).*

## 7.4 WP-D: Management, Administration, Impact, and Travel

Work package D contains the management and administrative parts of the project, together with impact, knowledge-exchange, outreach, and the travel budget.

### 7.4.1 Management

We plan to continue the effective management structure and processes that have been developed over the last twelve years and which have been regularly endorsed by the STFC Oversight Committee. The key body is the Project Management Board (PMB), see Table-7, which has direct representation from the LHC experiments to ensure that the project is driven at all times by the needs of the principal clients. At the start of GridPP4, 3FTE of management was funded, which has now been reduced to 2.5FTE.

| PMB Member | Additional Roles |
|---|---|
| Project Leader | Oversight of all areas. WLCG MB member. |
| Deputy Project Leader | Security Policy; Represents SouthGrid |
| Project Manager | HEPSYSMAN Chair. Represents non-LHC VOs. |
| Collaboration Board Chair | Tier-2 Oversight and leads dissemination/impact group. |
| ATLAS Liaison | Also represents NorthGrid |
| CMS Liaison | Also GridPP Technical Director. Also represents London Grid. |
| LHCb Liaison | Also represents ScotGrid and Network Liaison with JANET |
| EGI Liaison | EGI PMB Chair and UK NGI |
| Tier-1 Manager | Link to Tier-1 operation teams. |
| Production Manager | Link to Operations team. |
| LCG Representative | Link to CERN and Tier-0. |

*Table-7: PMB members*

The PMB consults with the GridPP Collaboration Board, formed from the group leaders of GridPP institutes, on matters of high-level policy and strategy. Strategic input and guidance is also received from the STFC Oversight Committee. The main conduits of communication to and from the technical level are via the Tier-1 Manager who interfaces to the internal Tier-1 structure, and the Production Manager who leads the Operations Team and chairs their weekly meeting. The GridPP user-board has been phased-out over GridPP4: the three LHC experiments with significant UK involvement are now directly represented on the PMB and the introduction of a weekly Experiment Liaison meeting, coupled with quarterly Tier-1 resource review meetings, have provided appropriate and responsive fora for addressing the needs of other virtual organisations.

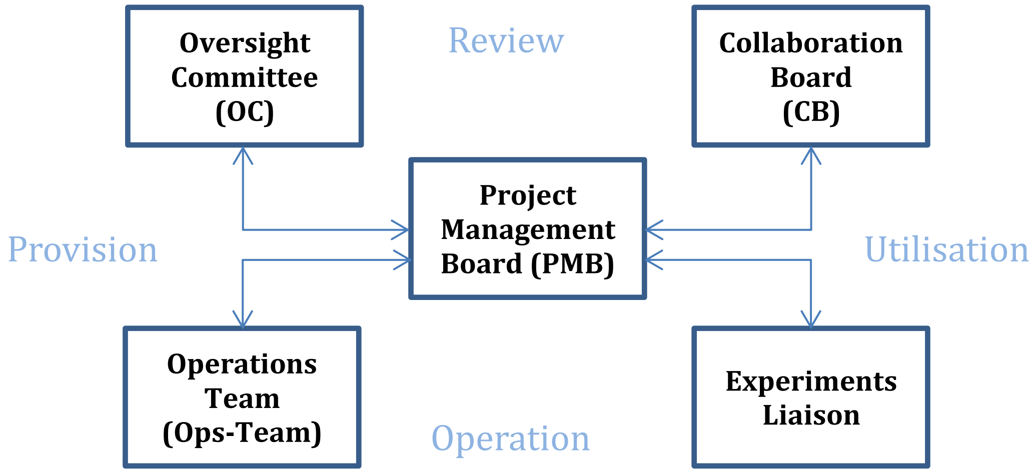



*Figure 6: Proposed GridPP5 management structures*

**Project Schedule -** GridPP5 is the fifth phase of what has become an operational project to provide Grid computing resources and support to the LHC experiments and other groups. As it has moved into this steady state, the project has increasingly been monitored by on-going metrics (such as “number of vacant posts” or “average service availability”, etc). In addition, a cyclic set of milestones is monitored that reflects the yearly process of pledging, procurement and delivery of resources at the Tier-1 (shown in Table-8); another set monitors the two-yearly hardware cycle at the Tier-2 institutes. These are complimented by additional, one-off, milestones that mark the evolution of the infrastructure. For example, in GridPP5 we will need to perform a review of our IPv6 readiness; and a review of our options for the next generation of storage software around the mid-point of the project. In addition, two site security challenges and updates to policy documents will be scheduled during the project.

| | |
|---|---|
| March | Internal Tier-1 strategic and technical review. |
| May | Tier-1 external review: Performance, delivery, future issues. |
| May | Hardware Advisory Group meets to review requirements. |
| June | Procurement milestones established and procurement started. |
| October | UK pledges to WLCG for following year (April 1st). |
| January | Experiments report to PMB on delivery during the past year. |
| March | New hardware testing completed to meet April 1st pledge date. |

*Table-8: The annual cycle of Tier-1 milestones*

Milestones and metrics are monitored by means of the GridPP ProjectMap[13], which for GridPP4, contains 182 metrics and 81 milestones. Many of these are directly applicable to GridPP5 and the ProjectMap will be adapted to fully cover GridPP5 once the scope of the project is defined. Associated with the ProjectMap is a flexible system of change-requests that allow us to adjust the project monitoring to address new issues that develop.

[13] The GridPP4 ProjectMap is available at http://www.gridpp.ac.uk/pmb/ProjectManagement/GridPP4_ProjectMap_9.xlsx

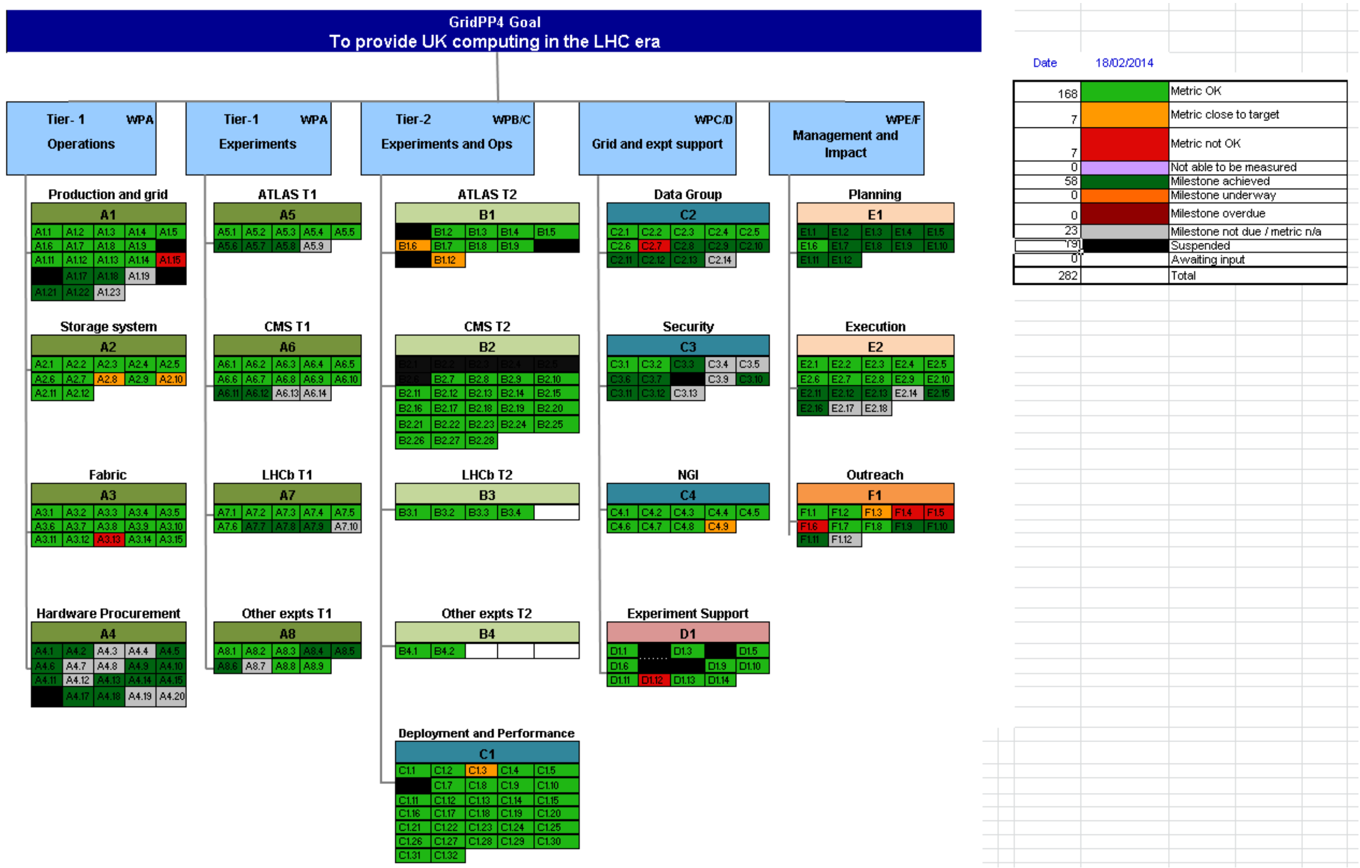



*Figure 7: The GridPP4 ProjectMap that will be extended to cover the GridPP5 project*

**Project Risks -** GridPP tracks risks by means of a Risk Register[14], which is reviewed regularly to ensure adequate steps are taken to mitigate any foreseen problems. Many, if not all, of the 31 risks in the current Risk Register are applicable to GridPP5. Many of our risks are quite dynamic and in the Risk Register we assess the risk-levels over the next 6-12 months. The main objective of GridPP5 is to deliver resources as defined in the MoU, so for GridPP5 the key additional risks to those in the Risk Register are:

1) The current financial uncertainty drags on and causes us to lose key staff, even if adequate funding is eventually awarded;
2) GridPP5 is funded at less than the Flat-Cash level and the project fails to deliver according to the MoU requirements;
3) Funding stops for externally-funded services upon which GridPP depends;
4) The impact of administrative constraints such as freezes on hiring at STFC; inflexibility with the capital/resource split; or procurement hurdles due to government directives.

The technical risks in GridPP5 are moderated by the fact that many of them are risks shared by the whole WLCG collaboration (eg: IPv6 and DPM). A few technical risks are more specific to the UK (eg CASTOR and its eventual replacement). Finally, there is a class of UK-specific risks to delivery, such as network-breaks and physical catastrophes, which are well covered in the Risk Register.

**Project Finance -** The financial position of grants issued to the universities and the SLA at RAL is closely tracked in liaison with STFC and monitored by the GridPP Oversight Committee. Depending on the scope approved for GridPP5, up to £12m will be spent on hardware, the majority of which will be categorised as capital, tape-media being the largest exception. Table-2 in Section 6.5 provides a more

[14] The GridPP4 RiskRegister is available at http://www.gridpp.ac.uk/pmb/ProjectManagement/GridPP4_Risk_Register_v1-Q313.xls

detailed breakdown. This will be spent in four annual procurements at the Tier-1 (Work Package A) for capacity hardware (CPU and disk) plus procurements for Tape media and non-capacity hardware. We propose to issue two hardware grants to the Tier-2 institutes (Work Package B) during GridPP5 (2015/6 and 2017/8). Most of these procurements will include "collections" of hardware that amount to more than £25,000 but it is not possible to specify the items in more detail until the procurements are initiated.

### 7.4.2 Impact and Outreach

GridPP impact and outreach activities are mainly based in the Tier-2 institutes and are coordinated by a 0.5 FTE funded GridPP Dissemination Officer based at QMUL. GridPP has responded to the pathways-to-impact programme to increase emphasis on impact and two-way engagement with industry.

The 0.5 FTE post holder is also the STFC Researcher in Residence at the Simon Langton Grammar School for Boys in Canterbury, working for the CERN@school project that conducts research-based and educational activities. GridPP has long been engaged with CERN@school, making Grid resources available to schools to store data and analyse the results from measurements with the detectors, promoting the Grid and modern computing techniques to a new generation of scientists and engineers. We believe that if we can successfully enable schools to routinely use GridPP resources then this will make it much easier to attract SMEs and others to do likewise.

GridPP resources continue to be used by a number of diverse projects and GridPP supports around 50 virtual organisations from non-HEP communities. The top non-HEP users over the last year were Life Sciences, European Earth Sciences, NMR and Structural Biology, Nuclear Fusion and Veterinary Surveillance. Results from work on the Tomato Genome and the similarities between Bats and Dolphins run on GridPP resources were published in Nature.

GridPP has been working with Alex Efimov, Director of Sensors and Instrumentation at the Electronics, Sensors, Photonics KTN. Alex has obtained funding to help GridPP's impact agenda and has written a 'GridPP Capability Guide' for potential industrial contacts. As a result, we are currently prototyping a project with 'Locked Space Technologies' involving scaling tests of a novel method of splitting files across disks and sites. We are starting a pilot study at Imperial, which if successful will be extended to QMUL and then to other GridPP sites. Other projects, including working with a financial institution in the City of London, are being discussed.

GridPP, via ScotGrid, has engaged with TechCube[15], an Edinburgh-based start-up incubator for technology companies, to develop a facility for SMEs to explore Big Data. A satellite Grid site, linked to the GridPP resources, will be set up with a £0.5m grant from Edinburgh City Council to enable pre-competitive work to be scale-tested.

GridPP has close ties with hardware and software manufacturers and suppliers. Recently, the Tier-1 for example, has been working with Data Direct Networks (DDN) to test and evaluate two new products: one hardware (SFA7700) and the other software (WOS). The Tier-1 is also working closely at present with one of the main hard disk manufacturers in order to observe a new disk product in an operational environment. The Tier-1 has also recently been working with Morgan Stanley to deploy Aquilon at the Tier-1 as a first step to making it generally available and usable.

Whilst we cannot predict what will happen in this area, we expect the Flat-Cash scenario for GridPP5 to continue to deliver effective support for STFC's impact agenda.

### 7.4.3 WP-D Resources in GridPP5 (Flat-Cash Scenario – 3 FTE)

In GridPP5 we reduce management effort to 2FTE; the final details of how this will be achieved will be established at a later date. We propose to maintain the 0.5FTE administrative assistant (which, as the only non-academic post in the project, is a cost effective way of supporting the Project Leader and Project Manager) and to continue the 0.5FTE effort of the GridPP Impact Officer in order to drive efforts

[15] http://techcu.be/

to respond to the STFC impact agenda. The total effort requested in WP-D is, therefore, 3FTE in the Flat-Cash scenario.

In the Flat-Cash scenario the GridPP5 travel budget is reduced by 20% reflecting the 10% reduction in the staff count and an additional 10% reduction in the travel budget per head.

# 8 Funding Scenarios

GridPP has received funding since September 2001 in various phases. Since 2008, the project has effectively received flat-cash at just under the £7m/year level. Assuming an average annual inflation of 2.5%, this corresponds to a 15% reduction over the last 6 years. In that period the UK has increased the resources delivered by more than a factor of 10; made large improvements to service reliability, functionality, and efficiency; and increased our client base to include new HEP VOs such as T2K, ILC, and NA62 together with new VOs from the wider community.

The next phase of GridPP covers the period April 2015 to March 2019, which corresponds to Run-2 of the LHC. The resources GridPP needs to deliver against the WLCG MoU increase by factors between 2.0x (CPU) and 2.7x (Tape storage), with the current high level of service needing to be maintained to deliver MoU commitments and to evolve the infrastructure to meet the changing computing models and take advantage of new technology and techniques.

GridPP has been asked to consider two funding scenarios:

- Flat Cash: This corresponds to £6.9m/year, which is about a 10% reduction from GridPP4 in real terms.
- 50% of Flat Cash, which corresponds to about £3.5m/year.

In both cases, the guidelines request that the scope of the bid should encompass operational support, both for the LHC and for other UK particle physics experiments, and the provision of experimental data analysis and storage services through the:

- maintenance and development of the Tier-1 Centre, central to the UK's contribution to the LHC experiments and the Worldwide LHC Computing Grid, and
- provision of Tier-2 centres as part of the integrated UK particle physics Grid computing resource.

GridPP is unable to provide a viable scenario for 50% of flat cash that meets the full scope requested. Inevitably, after many years of flat-cash whilst delivering increased requirements, there is very little, if any, headroom within the project and it is impossible to provide the full project scope for half the money. It is also not possible to present two extreme scenarios between which meaningful interpolations can be made, because there are step-changes in the functionality that can be delivered as the funding envelope is reduced. Therefore, although we will address the two scenarios requested, we will also indicate two additional funding levels where the scope of the project is significantly reduced. In neither of these additional cases, however, do we believe that such a project would enable the UK to deliver an effective, let alone a leadership-level, contribution to LHC computing.

## 8.1 Flat-Cash Scenario

In Section-7 of this document we have described how the GridPP5 project would deliver the WLCG MoU commitments in a Flat-Cash scenario. Staff reductions (summarised in Table-9 below) have been made in all areas.

To meet the Flat-Cash funding envelope of £6.9m/year, reductions have also been made to the hardware budget, compared to the baseline cost presented in Table 2 (page 12). Funding for capacity Disk and CPU are reduced by 20% at the Tier-1 and Tier-2s and funding for Tape Infrastructure at the Tier-1 would be reduced from £1.8m to £1.4m over the project. The total reduction in the hardware budget-line is £2.1m compared to the baseline of £12.3m and the associated risk is that the UK will fail to deliver the agreed MoU resources. We propose to take this risk because balancing factors are: (a) significant uncertainty in future hardware pricing where the cost evolution assumed by CERN is slightly more conservative than the trends we have observed to date in the UK; (b) opportunities for additional capital have appeared on several occasions in recent years, if this continues then the project might be able to benefit in the future; and (c) the historical over-provision by UK Tier-2s of resources compared to WLCG pledges. Nevertheless it must be emphasised that any reduction in funding is at the cost of an increased risk of not meeting international commitments.

The full cost summary for the Flat-Cash scenario is presented in Table-10 in Section 8.3.

## 8.2 De-scoping the Flat-Cash Scenario

GridPP is unable to provide a scenario for 50% of flat cash that meets the full scope requested. Even the minimum threshold hardware budget proposed in the Flat-Cash scenario requires £2.5m/year leaving, in a 50% Flat-Cash budget, funding for just 12 FTE, well below the level required to provide either a Tier-1 or Tier-2 service, let alone both. Consequently, in this scenario, there is no possibility of meeting the current international MoU commitments in full and a choice has to be made as to which parts of the service to retain; a choice that should be driven by the LHC experiments.

Together, ATLAS and CMS, represent about 70% of the GridPP user-base, with LHCb representing another 20%. The final 10% is composed of ALICE, other HEP VOs, and non-HEP users. Discussions were held with ATLAS, CMS and LHCb representatives who consulted with both UK and international management within their respective experiments. ATLAS and CMS regard the current integrated Tier-1 and Tier-2 services as being essential to deliver commitments to collaborations and CERN, and a vital infrastructure for the UK community. Forced to de-scope, both experiments consider that a Tier-2 service, supported by a custodial tape-storage facility at the Tier-1, is the least-worst scenario. They note that the usefulness of such a tape-store would very much depend on maintaining a significant disk-buffer in front of it, effectively turning the RAL Tier-1 into a data-source/sink for the UK Tier-2 infrastructure. In this scenario, the UK would have to withdraw from the commitment to deliver Tier-1 resources but could probably deliver the expected Tier-2 resources. The emphasis on Tier-2 resources reflects the current and expected evolution of the GPD experiments' computing models described in Section-4.2.

LHCb is somewhat different to the GPD experiments in that, historically, it has made more use of Tier-1 resources, with the Tier-2 infrastructure used almost exclusively for Monte Carlo simulation. This has started to change with the recent introduction of LHCb Tier-2 sites hosting significant amounts of disk to support both reprocessing and analysis work. Despite this, LHCb considers the UK Tier-1 service as their top priority. Since the UK is such a large part of LHCb (21%), the loss of the UK Tier-1 would have major international implications. However, as the relative size of ATLAS and CMS would inevitably dictate the choice in a below flat-cash scenario, LHCb would hope to accelerate the recent T2D development by establishing such additional sites in the UK to contribute to all processing, reprocessing and simulation tasks. This would require some staff effort at Tier-2s to be specifically identified to support LHCb in the same way that, at present, Tier-2 staff effort is identified for explicit support for ATLAS and CMS. Naturally LHCb would also need RAL to act as a data source/sink and to provide custodial tape services.

In summary, if there is insufficient funding to enable GridPP to deliver the full scope of UK commitments in the WLCG MoU, we must focus on maintaining a robust Tier-2 service whilst re-focusing the RAL Tier-1 as a custodial tape-store and, if affordable, a sink/source for data. GridPP would continue to deliver resources to ALICE and other VOs, resources permitting, using the Tier-2 infrastructure and the UK would have to negotiate a withdrawal from commitments to deliver Tier-1 resources to WLCG.

In the next sections we explore the costs and risks of de-scoping the flat-cash scenario to the 50% level and try to identify the step-changes that occur as funding is reduced.

### 8.2.1 Minimal Frozen Service

The first major step-change is the point at which the UK could only just continue to meet the MoU commitments and deliver both the Tier-1 and the Tier-2 services in the short term. We define this scenario as that in which there would a reasonable chance that we could meet the MoU requirements in 2015, or if not, then at least get close enough to avoid having to enter into international negotiations in respect of the UK failing to provide its equitable share. We refer to this as the Minimal Frozen Service scenario, which reflects the fact that, at the level of effort proposed, it will be impossible to continue to evolve the infrastructure and by the end of GridPP5 we would almost certainly not be delivering the service required. *This scenario would delay, but not avoid, the requirement to negotiate withdrawal from the obligations of the WLCG MoU*.

The starting point is to maintain the 20% de-scoped level of capacity hardware presented in the Flat-Cash scenario, as this is the minimum necessary to have a reasonable chance of delivering the MoU commitment. However, in the Minimal Frozen scenario, additional cuts would be made to the non-capacity hardware (for example, delaying the upgrade of the network link to CERN by a year and reducing the funding for networking hardware) to save an additional £400,000.

At the Tier-1, the staffing level of 17.5 FTE proposed in the flat-cash scenario is expected to slow the speed at which the Tier-1 can evolve and prepare for the future. By dedicating no effort to longer-term developments, however, we estimate that the current service could be preserved with a minimum of 15 FTE. The key risks here are that the Tier-1 would store up problems for the future and that the service would degrade over the lifetime of GridPP5; that the Tier-1 would not be able to handle the step-change expected for Run-3; and that the Tier-1 would not be able to take advantage of the savings that could ultimately be made in investing in new technologies. It is almost certain that, in this scenario, the Tier-1 would lose its status as one of the more reliable Tier-1 sites and the operation would slip into a perpetual fire-fighting mode.

At the Tier-2s, the Flat-Cash proposal of 19.5 FTE would enable a robust structure to continue, albeit with a reduced number of sites with funded effort. In the Minimal Frozen scenario we consider an absolute minimum structure would comprise 8 large sites, each with funding for 2 FTE, giving a total of 16 FTE. With the loss of an additional 3.5 FTE compared to the Flat Cash scenario, a further seven sites would lose funded effort with the risk that the hardware they have already installed becomes ineffective or unavailable. In de-scoping to this scenario we risk losing nearly half of the GridPP sites along with their current hardware, the associated leverage, engagement and local support. There is also a great danger that significantly increasing the size of the remaining sites to accommodate the capacity hardware may lead to further loss of leverage if individual facilities become too large for the universities to continue to support (at no charge to GridPP) recurrent costs such as electricity. We note that the estimated value of electricity provided by the Tier-2 sites is equivalent to the cost of 8 FTE/year over the lifetime of GridPP5.

In this Minimal Frozen Service scenario, the Deployment/Operations/Support posts and the Experiment Support posts would remain at the level of the Flat-Cash scenario. We do not consider it possible to reduce these further if we are to continue to provide a full service. The management line would be cut back from 2.0 to 1.5 FTE and funding for the 0.5 FTE for an Impact Officer would have to be removed. The 0.5 FTE for an administrative assistant, however, would need to be retained in the light of the reduced management. The risks in this area are that project would not be adequately managed and that GridPP would have no dedicated effort to support STFC's Impact Agenda.

The total effort required for the Minimal Frozen service is 49.25 FTE for a total project cost of £6.2m/year (90% of the Flat-Cash scenario). See tables in Section 8.3 for the full comparison.

### 8.2.2 Partial Tier-1 Service

Following the direction which was set through discussions with the LHC experiments, the next step in de-scoping GridPP would be to focus on maintaining a Tier-2 service while further reducing the scope of the UK Tier-1. *STFC would need to re-negotiate the UK's involvement in WLCG and fully understand the broader implications of delivering less than the MoU commitment.*

This scenario preserves the custodial tape-store at RAL and a disk buffer in front of it of a few PBs (initially, this could be quite large, using the 10PB of disk already at the Tier-1; but, eventually, this would have to be reduced to the level of 2-4 PB). In this configuration, RAL could not function as a Tier-1 but would act as a data-source and data-sink for the UK Tier-2s, and maintain an international role in the custodial storage of LHC data. Given the evolution of the Experiment Computing Models the UK Tier-2s could contribute to what traditionally have been Tier-1 tasks. On the other hand, the UK would be unable to contribute to the trend to share high-level, critical tasks such as prompt reconstruction between CERN and the Tier-1s and would have to rely on our international collaborators to pick up our share.

The Tier-2s would also not be able to provide the Tier-1 level of service but the more distributed nature of the Tier-2 infrastructure would bring some compensation in that, at any one time, it is likely that at least part of the service would be available (modulo single points of failure) and upgrades and testing could be done at different times at different sites.

In this scenario, no Tier-1 capacity hardware is funded at RAL, but 25% additional capacity hardware is funded at the Tier-2s to enable some contribution over and above the strictly-defined Tier-2 tasks. The network connection to CERN and the tape infrastructure at RAL would be funded at the same level as in the Minimal Frozen Service scenario above. The hardware budget is thus reduced by about £2m from the previous scenario.

It is assumed that the tape-store and associated databases can be maintained at RAL with about 4FTE; and that the disk-buffer, networking, and other services (such as file transfers and monitoring) would require another 3FTE for a total of 7FTE. The 3.0FTE of experiment support posts are removed in this scenario: these primarily act as an engagement method between the Tier-1 and the Experiments, so the number of Tier-2 staff is increased from 16 to 18FTE to provide the necessary engagement and, notably, to support LHCb more explicitly on the Tier-2 infrastructure. Management is reduced from 1.5 to 1.25FTE. The total effort required for the Partial Tier-1 service is 42.25FTE and the project cost is £4.9m/year (70% of the Flat-Cash scenario).

In this scenario the consequences (rather than risks) are that UK would be unable to meet the MoU commitments for Tier-1 resources, and the associated level of service. At a practical level, the UK could perhaps, depending on how much non-GridPP funded resource is provided by the Tier-2 institutes, continue to provide some current Tier-1 functionality but without being able to guarantee the expected resource and service levels. As for the Minimal Frozen Service level above, there is a great danger that increasing the amount of funded GridPP Tier-2 hardware, whilst reducing the number of GridPP Tier-2 sites, will lead to loss of leverage if individual facilities become too large for the universities to continue to support infrastructure investment and recurrent costs such as electricity.

### 8.2.3 Fully De-scoped 50% Scenario

The final step in de-scoping to the 50% scenario requires:

- removing any possibility of the UK contributing Tier-1 functionality (although the tape-store at RAL is preserved);
- reducing Tier-2 hardware to 80% of the level required to meet the MoU commitment based upon the UK authorship fraction;
- reducing the number of UK Tier-2 sites to seven and the associated staffing from 18FTE to 13FTE;
- reducing deployment/operations/support staff from 6.25 to 5.5FTE with choices on which international responsibilities to reduce or withdraw from (Grid Operations Centre; APEL accounting; or international security policy) being negotiated with international partners; and
- a management line further reduced to just 1FTE.

We do not present the risks inherent in this scenario: the damage is real and apparent: the UK would do no more than host a minimal Tier-2 infrastructure and endeavour to support local analysis efforts.

The total effort in this fully de-scoped scenario is 24 FTE and the project cost is £3.5m/year.

## 8.3 Scenario Comparison

The three tables below compare the different scenarios in terms of manpower (Table 9), the costs (Table 10), and the ability of the project to meet the current MoU commitments (Table 11).

| | | | | De-scoping to 50% Flat Cash | | |
|---|---|---|---|---|---|---|
| **Work Package** | **Area** | **GridPP4** | **Flat-Cash** | **Minimal Frozen Service** | **Partial Tier-1** | **Tape Only Tier-1** |
| **WP-A** | Tier-1 Staff | 19.50 | 17.50 | 15.00 | 7.00 | 4.00 |
| **WP-B** | Tier-2 Staff | 20.75 | 19.50 | 16.00 | 18.00 | 13.00 |
| **WP-C** | D/O/S Staff | 11.73 | 9.25 | 9.25 | 6.25 | 5.50 |
| **WP-D** | Manage/Admin/KE | 4.00 | 3.00 | 2.00 | 1.75 | 1.50 |
| **TOTAL FTE** | | **56** | **49** | **42** | **33** | **24** |

*Table-9: **Manpower** (FTE) funded by work-package for the different scenarios*

| | | | De-scoping to 50% Flat Cash | | |
|---|---|---|---|---|---|
| **Category** | **GridPP4** | **Flat-Cash** | **Minimal Frozen Service** | **Partial Tier-1** | **Tape Only Tier-1** |
| **Staff Costs** | £17.7m | £17.0m | £14.6m | £11.4m | £8.3m |
| **Hardware Costs** | £8.9m | £10.1m | £9.7m | £7.8m | £5.6m |
| **Travel and Misc.** | £0.80m | £0.65m | £0.56m | £0.44m | £0.27m |
| **Total Cost** | £27.4m | £27.7m | £24.9m | £19.6m | £14.2m |
| **Cost per annum** | **£6.9m** | **£6.9m** | **£6.2m** | **£4.9m** | **£3.5m** |
| **De-scope *wrt* GridPP4** | **0%** | **-10%** | **-20%** | **-36%** | **-54%** |

*Table-10: Project **costs** for the different scenarios*

| MoU Commitment | GridPP4 | Flat Cash | Minimal Frozen Service | Partial Tier-1 | Tape Only Tier-1 |
|---|---|---|---|---|---|
| Tier-1 Capacity Hardware | Yes | ~Yes | ~Yes | No | No |
| Tier-1 Service Level | Yes | Yes | Increasingly degraded | No | No |
| Tier-2 Capacity Hardware | Yes | ~Yes | ~Yes | Yes | ~Yes |
| Tier-2 Service Level | Yes | Yes | Increasingly degraded | Increasingly degraded | No |

*Table-11: Meeting the **MoU commitments** in the different scenarios. The tentative "~Yes" reflects the fact that there is only funding for 80% of the full estimated cost of the hardware. "Increasingly degraded" means that the service level could not be maintained over the full duration of the project.*

In addition to the risks discussed in Section 7.4.1, a reduction in funding from flat-cash progressively adds risks to the project as follows:

**Flat-Cash:**

1) Only 80% of the estimated hardware cost is funded. We hope that other sources of capital funding might become available during the project and/or hardware costs will fall more quickly than suggested by CERN. There is a risk of not fully meeting the MoU commitments. The provision to non-LHC VOs may have to be reduced significantly or entirely.

**Minimal Frozen Service:**

2) Additional risk/certainty that there is insufficient manpower at the Tier-1 to address the required service evolution and after a year or two (depending on speed of evolution) the UK would not be able to maintain the WLCG MoU service requirements. More explicitly, we may not be able to fully test new releases; we may not be able to proactively investigate replacement storage systems for CASTOR that are required in the mid-term; we may fail to prepare adequately for IPv6; we might not contribute appropriately to the task and the leadership in WLCG.

3) There is a significant risk that reducing the number of Tier-2 sites will disproportionately reduce the leverage. Not only will we lose leverage (including access to new opportunistic resources such as campus private clouds and desktop grids) and hardware at the sites concerned, but concentrating the GridPP5 hardware at fewer sites may also lead to a reduction in the support those sites can provide in terms of electricity. Reduction of the number of sites with funded effort will also reduce support and engagement opportunities.

4) It is clear that with threshold manpower GridPP will not be able to fully engage with external developments such as EU-T0, H2020, the development of the broader UK e-infrastructure ecosystem, and the Impact Agenda.

**Partial Tier-1 scenario:**

5) GridPP will fail to deliver the MoU Tier-1 resources. This will require STFC to re-negotiate international agreements. This will destroy the UK's reputation as a reliable partner, which may compromise the UK's position in future international collaborations; undermine past investments in the

LHC detectors and computing; and may have financial consequences that reduce the apparent savings in this scenario.

6) Concentrating the GridPP5 Tier-2 resources at far fewer sites will lose existing hardware and leverage at the sites that become unfunded. There is a high risk that concentrating the GridPP5 hardware at far fewer sites will lead to increased costs for electricity and local infrastructure that the institutes will not support.

7) There is a technical risk that despite the guidance from the GPDs, the UK will have difficulty adapting the experiment computing models to work with the partial Tier-1 architecture and the resources will be used less efficiently.

**Tape-Only Tier-1 scenario:**

8) The UK would, in addition, be unable to provide Tier-2 resources at the service level required by the MoU.

# 9 Conclusion

We have explained how the science programmes approved for each experiment, in effect determine the level of global resources required by the experiments; how, once the trigger rate is fixed, the processing and storage requirements then follow, with little scope for variation; how, equally deterministically, the UK share of the global resources reflects the UK participation in each experiment; how GridPP functions as an integral and significant part of the Worldwide LHC computing Grid—which like all collaborations, brings advantages and obligations—and, finally, how GridPP must, on behalf of the UK, provide the infrastructure at well-defined levels of service and contribute at an appropriate level to the common tasks that are required.

In this document we have presented a plan to deliver the UK MoU commitments in a Flat-Cash scenario. After many years of flat-cash and increasing requirements, this is challenging and not without risk. We have also addressed, as requested, a scenario with half of that level of funding, demonstrating we hope, that such a level of funding would not enable the UK to meet international commitments. Given that we consider a 50% reduction in funding in comparison to GridPP4 to be unrealistic, we have further illustrated how service to the LHC experiments and the WLCG community would be degraded as funding drops progressively below a flat-cash scenario.

GridPP has, for many years, ensured an effective UK contribution to the computing needs of the LHC experiments. We recognise that STFC faces difficult choices in today’s financial situation but enabling effective exploitation of future LHC data by UK physicists requires continued investment in computing resources at least at the level of flat-cash.

# Appendix: Acronyms

| | |
|---|---|
| ALICE | A Large Ion Collider Experiment |
| AoD | Admin on Duty |
| APARSEN | Alliance Permanent Access to the Records of Science in Europe Network |
| APEL | Accounting Processor for Event Logs |
| ATLAS | A Torroidal LHC ApparatuS |
| BDII | Berkeley Database Information Index |
| CA | Certificate Authority |
| CASTOR | CERN Advanced STORage Manager |
| CB | Collaboration Board |
| CE | Computing Element |
| CEMS | Climate and Environmental Monitoring from Space |
| CERN | The European Organisation for Nuclear Research |
| CHEP | Computing in High Energy Physics (largest WLCG conference) |
| CIF | Capital Investment Fund |
| CLF | Central Laser Facility |
| CMS | Compact Muon Solenoid |
| CPU | Central Processing Unit |
| CRRB | Computing Resource Review Board |
| C-RSG | Computing Resource Scrutiny Group |
| CSIRT | Computer Security Incident Response Team |
| CVMFS | CERN Virtual Management File System |
| DiRAC | Distributed Research utilising Advanced Computing |
| DPM | Disk Pool Manager |
| DRI | Digital Research Network Infrastructure |
| DST Data | Data Summary Tape Data |
| EGEE | Enabling Grids for E-SciencE |
| EGI | European Grid Infrastructure |
| EGI-InSPIRE | Integrated Sustainable Pan-European Infrastructure for Researchers in Europe |
| EINFRA | e-Infrastructure funding call |
| EMI | European Middleware Initiative |
| EU | European Union |
| EUDAT | European Data Infrastructure Project |
| EUGridPMA | EU Grid Policy Management Agency |
| FTE | Full-time equivalent |
| FTS | File Transfer Service |
| FY | Financial Year |
| GANGA | Gaudi/Athena and Grid Alliance |
| GGUS | Global Grid User Support |
| GOC | Grid Operations Centre |
| GOCDB | Grid Operations Centre Database |
| GPD | General Purpose Detector |
| GPU | Graphics Processing Unit |
| GridPP | Grid for Particle Physics |

| | |
|---|---|
| H2020 | Horizon 2020 |
| HEP | High Energy Physics |
| HEPiX | The High Energy Physics Unix Information Exchange |
| HEPSPEC | High Energy Physics Standard Performance Evaluation Corporation |
| HL-LHC | High Luminosity Large Hadron Collider |
| HLT | High Level Trigger |
| IGTF | International Grid Trust Federation |
| ILC | International Linear Collider |
| JANET UK | Joint Academic NETwork |
| JASMIN | e-Infrastructure for climate and earth system science |
| KE | Knowledge Exchange |
| LAN | Local Area Network |
| LCG | LHC Computing Grid |
| LFC | LCG File Catalogue |
| LHC OPN | LHC Optical Private Network |
| LHC | Large Hadron Collider |
| LHCb | Large Hadron Collider beauty (one of the experiments) |
| LHCC | LHC Computing Committee |
| M&O | Maintenance and Operations |
| MB | MegaByte |
| MICE | Muon Ionization Cooling Experiment |
| MoU | Memorandum of Understanding |
| NERC | Natural Environment Research Council |
| NGI | National Grid Initiative |
| NGS | National Grid Service |
| OC | Oversight Committee |
| OPN | Optical Private Network |
| OS | Operating System |
| OSG | Open Science Grid |
| PMB | GridPP Project Management Board |
| PPARC | Particle Physics and Astronomy Research Council |
| QMUL | Queen Mary, University of London |
| QUATTOR | system administration tool-suite |
| RAID | Redundant Arrays of Independent Disks |
| RAL PPD | Rutherford Appleton Laboratory Particle Physics Department |
| RAL | Rutherford Appleton Laboratory |
| RAM | Random-access memory |
| RCUK | Research Councils UK |
| REBUS | Resource, Balance and Usage website for WLCG |
| RHUL | Royal Holloway, University of London |
| SCD | Scientific Computing Department (STFC) |
| SL | Scientific Linux |
| SLA | Service Level Agreement (RAL finances) |
| SME | Small and Medium Enterprises |
| SPEC | Standard Performance Evaluation Corporation |

| | |
|---|---|
| SRIF | Science Research Investment Fund |
| STEP | Scale Test for the Experiment Programme |
| STEP09 | Scale Test for the Experiment Programme 2009 |
| STFC | Science and Technology Facilities Council |
| T2D | Tier-2s with Data |
| TB | Terabyte |
| TDR | Technical Design Report |
| TERENA | Trans-European Research and Education Networking Association |
| TeV | Tera electron-volt |
| UB | User Board |
| UCL | University College London |
| UK CA | UK Certificate Authority |
| VO | Virtual Organisation |
| VOMS | VO Management Service |
| WAN | Wide Area Network |
| WLCG TDR | WLCG Technical Design Report |
| WLCG | Worldwide LHC Computing Grid |
| WMS | Workload Management System |
| WP | Work Package |