EGI-InSPIRE

CAPABILITIES OFFERED BY HEAVY USER COMMUNITIES

EU DELIVERABLE: D6.1

Document identifier:	EGI-D6.1-V9.doc
Date:	**24/11/2010**
Activity:	**SA3**
Lead Partner:	**CERN**
Document Status:	**DRAFT**
Dissemination Level:	**PUBLIC**
Document Link:	https://documents.egi.eu/document/154

Abstract

This document describes the capabilities offered by the Heavy User Communities (HUCs) to other communities. This public report illustrates how the functional capabilities being supported by this activity can be re-used by other communities using European Distributed Computing Infrastructures (DCIs). Sufficient technical depth is provided for potential adopters of DCI platforms to make an initial assessment of how they could work with the offered technologies.

I. COPYRIGHT NOTICE

II. DELIVERY SLIP

	Name	Partner/Activity	Date
From	Jamie Shiers	CERN	30/7/ 2010
Reviewed by	**Moderator:** Steven Newhouse **Reviewers:** Gergely Sipos, David Wallom	EGI.eu EGI.eu, University of Oxford	22/9/2010
Approved by	**AMB & PMB**		24/11/2010

III. DOCUMENT LOG

Issue	Date	Comment	Author/Partner
1	July 2010	First draft	IT-ES group, CERN
2	August 2010	Case studies received from partners	SA3
3	August 24, 2010	Merged document with new input, obvious formatting errors fixed.	SA3
4	August 28, 2010	Final input, slight reorganisation, consistent spelling for names / acronyms	SA3
5	October 25, 2010	Reorganisation and corrections in answer to the reviewers' comments	SA3

IV. APPLICATION AREA

This document is a formal deliverable for the European Commission, applicable to all members of the EGI-InSPIRE project, beneficiaries and Joint Research Unit members, as well as its collaborating projects.

V. DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the authors. The procedures documented in the EGI-InSPIRE "Document Management Procedure" will be followed:
https://wiki.egi.eu/wiki/Procedures

VI. TERMINOLOGY

A complete project glossary is provided at the following page: http://www.egi.eu/about/glossary/.

VII. PROJECT SUMMARY

To support science and innovation, a lasting operational model for e-Science is needed − both for coordinating the infrastructure and for delivering integrated services that cross national borders.

The EGI-InSPIRE project will support the transition from a project-based system to a sustainable pan-European e-Infrastructure, by supporting ‘grids’ of high-performance computing (HPC) and high-throughput computing (HTC) resources. EGI-InSPIRE will also be ideally placed to integrate new Distributed Computing Infrastructures (DCIs) such as clouds, supercomputing networks and desktop grids, to benefit user communities within the European Research Area.

EGI-InSPIRE will collect user requirements and provide support for the current and potential new user communities, for example within the ESFRI projects. Additional support will also be given to the current heavy users of the infrastructure, such as high energy physics, computational chemistry and life sciences, as they move their critical services and tools from a centralised support model to one driven by their own individual communities.

The objectives of the project are:

1. The continued operation and expansion of today’s production infrastructure by transitioning to a governance model and operational infrastructure that can be increasingly sustained outside of specific project funding.
2. The continued support of researchers within Europe and their international collaborators that are using the current production infrastructure.
3. The support for current heavy users of the infrastructure in earth science, astronomy and astrophysics, fusion, computational chemistry and materials science technology, life sciences and high energy physics as they move to sustainable support models for their own communities.
4. Interfaces that expand access to new user communities including new potential heavy users of the infrastructure from the ESFRI projects.
5. Mechanisms to integrate existing infrastructure providers in Europe and around the world into the production infrastructure, so as to provide transparent access to all authorised users.
6. Establish processes and procedures to allow the integration of new DCI technologies (e.g. clouds, volunteer desktop grids) and heterogeneous resources (e.g. HTC and HPC) into a seamless production infrastructure as they mature and demonstrate value to the EGI community.

The EGI community is a federation of independent national and community resource providers, whose resources support specific research communities and international collaborators both within Europe and worldwide. EGI.eu, coordinator of EGI-InSPIRE, brings together partner institutions established within the community to provide a set of essential human and technical services that enable secure integrated access to distributed resources on behalf of the community.

The production infrastructure supports Virtual Research Communities (VRCs) − structured international user communities − that are grouped into specific research domains. VRCs are formally represented within EGI at both a technical and strategic level.

VIII. EXECUTIVE SUMMARY

Through a series of case studies covering the full range of disciplines currently supported by the Heavy User Communities, this document explains in concrete terms what is required in order to exploit the grid for a variety of applications. It covers the full lifecycle from first "test-drives" to petascale production use (typically defined as 10^5 cores but in this case also in terms of data volume stored and transferred).

Although the case studies in this document underline the applicability of grid computing (and the services for the Heavy User Communities in particular) to a wide and diverse range of application areas, concrete steps in the grid require close collaboration between application-specific experts and those of grid technology.

This is an annual report that will be updated or rewritten to reflect the then-current status in one and two years from the date of this document.

TABLE OF CONTENTS

1 INTRODUCTION

The purpose of this document is to describe the capabilities offered by the Heavy User Communities (HUCs) to other communities. This public report illustrates how the functional capabilities being supported by this activity can be re-used by other communities using European DCIs. Sufficient technical depth is provided for potential adopters of DCI platforms to make an initial assessment of how they could work with the offered technologies.

The strategy adopted is to present a number of case studies that cover a range of areas and disciplines, including initial “gridification” investigations (the porting of an application to the grid environment) to full production support for a new virtual organisation, with a specific focus on the services supported by this work package.

2 AN OVERVIEW OF GRID COMPUTING

After one decade of grid computing there are now many demonstrations and realisations of production usage, ranging from scavenging of spare cycles on desktop PCs to “petascale computing”, as shown most clearly by the Worldwide LHC Computing Grid (WLCG). Grids have thereby been shown to be suitable for applications ranging from quasi-“single user” cases – for example, when a well-defined application has to be run for a short period, but for which turn-around is critical – to decade-long, multi-VO (Virtual Organization[1]) environments, supporting thousands of users, up to one million jobs per day, data rates of multi-GB per second and multiple petabytes of data and for which the application development cycle is a continuous process. Clearly, the requirements in these cases differ widely and hence a “one-size-fits-all” approach is unlikely to be applicable. In order to make this document as useful as possible, a variety of case studies are presented, covering all the disciplines supported by this activity as well as the full dynamic range of usage. It covers areas that are “linked by the discipline” – e.g. showing re-use amongst various High Energy Physics collaborations – as well as those “linked by the technology” – where sometimes widely different application domains can benefit from the same tools, applications, or even basic concepts.

The case studies are presented in order of increasing complexity – from first steps through full production to advanced use and optimisation.

2.1 What is the grid?

In 2002 Ian Foster – often referred to as one of the founders of grid computing – wrote a definitive article: What is the Grid? A Three Point Checklist. This short document can still be considered valid and highlights the following three points:

1. (A grid) coordinates resources that are not subject to centralized control…
2. … using standard, open, general-purpose protocols and interfaces…
3. …to deliver non-trivial qualities of service.

Thus, a grid is more than a distributed system: it is one with which a well-defined and measurable level of service is associated, resources are coordinated as opposed to centrally managed and it is based on interfaces and standards rather than strictly defined (and typically limited) implementations. These attributes offer a great deal of flexibility that comes at the price of a certain degree of complexity. Despite constant efforts to keep the costs of entry and ownership as low as possible, there are non-negligible start-up and on-going support issues that need to be well understood. For applications that do not require significant computational or storage resources, nor for which the aggregation of resources across multiple sites is a significant advantage, a simpler approach, e.g. based on a localised farm, may be more appropriate. On the other hand, where one or both of these is a requirement – as typified by the needs of large High Energy Physics experiments – grid computing offers a proven solution. Even if this is not the case, some of the basic concepts that are fundamental to grid computing – such as the virtual organisation – may well be of interest. More details are given in the case studies below.

[1] A VO is collections of sites and users from the infrastructure that come together into a collaboration to achieve some common goal (e.g. to perform a simulation, to run an application, etc.).

2.2 Motivation

There is a wide range of applications that require significant computational and storage resources – often beyond what can conveniently be provided at a single site. These applications can be broadly categorized as:

- **Provisioned:** meaning that the resources are needed more or less continuously for a period similar to, or exceeding, the usable lifetime of the necessary hardware
- **Scheduled:** where the resources are required for shorter periods of time and the results are not necessarily time critical (but higher than for the following category)
- **Opportunistic:** where there is no urgent time pressure, but any available resources can be readily soaked up.

Reasons why the resources cannot easily be provided at a single site include those of funding, where international communities are under pressure to spend funds locally to institutes that are part of the collaboration, as well as those of power and cooling – increasingly a problem with high energy prices and concerns over greenhouse gases. To give an explicit example, the WLCG requirement has always been for the order of 100,000 of the fastest commodity PCs on the market (now interpreted as 100K cores) – far above what can realistically be provided by an individual research laboratory today.

Currently, adapting an existing application to the grid environment requires an in-depth understanding not only of the grid computing paradigm but also of the computing model of the application in question. This requires that experts in both the application domain as well as grid technology need to work closely together to assess the suitability of grid technology and perform the necessary porting and infrastructure setup. More details are given below.

2.3 First steps in the grid

2.3.1 General remarks

The current HUC grid infrastructure is primarily developed and supported for the SL(C) – Scientific Linux (CERN) – operating system. Usage with other operating systems may be possible but may result in much more difficult and time-consuming deployment issues, and occasionally incompatibilities. The choice of the operating system has an impact on VO management (deployment of grid services) as well as on the applications (which should be able to run on a chosen OS).

Another important decision to be made by the VO is the choice of middleware, which is provided by the European Middleware Initiative (EMI) for EGI. This decision should evaluate the capabilities of the different middleware stacks and the requirements of the higher level applications and VO users. This report will refer in particular to the gLite middleware that is widely used in WLCG and EGI.

Several of the sections below contain hyperlinks to more detailed documentation. These are clearly identified using the standard notation (blue underlined text).

2.3.2 VO managers

A VO is collections of sites and users from the infrastructure that come together into a collaboration to achieve some common goal (e.g. to perform a simulation, to run an application, etc.).

In general you want to make the life of your users as easy as possible, not exposing them to the unneeded underlying complexity of grid infrastructure. This requires documentation to be provided in the early phases of your project and simplification of the procedures of interest to your users. A list of the key steps to be followed by the VO managers is detailed here:

1. **Registration of a new VO:** Before you go on, check if your activity may be hosted by one of the existing VOs. This will save you the considerable effort of creating your own VO – described below. Otherwise, the online registration form for a new VO is available here. The procedure is fully described in this document and includes security certificates and acceptable use policy to be defined for your VO.
2. **Resources:** New VOs have been provided with minimal set of resources just for basic testing – typically in opportunistic or occasionally (for specific production-like purposes) scheduled fashion. As a VO manager you should find existing grid sites that are willing to contribute the resources to your activity.

   N.B. production grid usage – other than at the level of first test-drives or low-level purely opportunistic usage – implies the contribution of the corresponding resources.
3. **User Interface:** Job submission requires an installation of grid middleware called the User Interface on a computer where users have interactive login access. Re-locatable tarfiles are probably the easiest installation method, which is described here. General gLite installation instructions can be found at this address. The UI installation is nonetheless quite difficult - don't ask your users to do it themselves.

   Portals and other high level environments, built on top of the UI, hide the UI itself from the end users and make the UI installation and maintenance an administrative task, which is handled by the system administrator of the community and not by the members.
4. **Software areas:** For applications that require large and complex installation, you should probably provide pre-installed application code in the grid sites that support your VO (for simple/smaller applications users may just send all required files with every job). This may be achieved using your VO manager credentials and so-called software tags and software areas. The software area is simply a local directory in each site, typically mounted on a network file system and shared with the worker nodes (where the jobs execute) such that jobs may access application files installed there. (N.B. for larger VOs, instabilities or configuration issues with the software area continues to be one of the major reasons for tickets to be opened against sites, with corresponding impact on production activities. Careful design and management of the software area is essential and should not be overlooked.) The software tag is a way of marking a site to have a certain version of application installed, such that users may send jobs to those sites only. An example of how this works is provided here.

2.3.3 VO users

Make sure you understand basic security concepts: digital certificates and public-private keys and follow the registration procedure that is described here.

Grid computing revolves around two key concepts: jobs (computational tasks) and data (access to distributed storage systems). The gLite manual provides an in-depth introduction to these concepts.

The gLite middleware supports different types of jobs including the parametric job type for parameter sweep applications. User may provide job requirements, which specify desirable resource parameters such as memory and CPU time limits, installed application software tags, etc. The system automatically finds the most appropriate resources for job execution and allows the specification of so-called sandbox files, which are transferred from the end-user environment (user interface) to the worker nodes and back. This transfer method is suitable for small files, where a typical limit for all files in a sandbox is of the order of 10 MB. Data management tools support the distribution and transfer of larger files. The key concepts in data management include: Storage Elements (SEs) which are typically mass-storage systems distributed in the grid and File Catalogs (FCs) which allow one to keep track of files and in some cases also metadata lookup (e.g. AMGA).

As raw gLite commands may be difficult to use, job management may be greatly simplified by using higher-level tools, such as Ganga (see Section 6.2), which allow also easy switching between various middleware flavours and local job processing systems. Other tools of interest are listed in the RESPECT (Recommended External Software Packages for EGI CommuniTies) programme that can be consulted here (under menu Applications & Tools).

Useful references and in-depth reading: the LCG User Guide.

3 HEAVY USER COMMUNITIES

HUCs are Virtual Research Communities (VRCs)[2] that have been using EGEE and EGI routinely and thus have become more structured and advanced in terms of grid usage. These communities focus on domain specific issues, such as how to access High Energy Physics applications on EGI, how to enable new physics experiments on EGI and so on.

On the one hand these teams are operated by external projects, such as WLCG, but on the other hand have members in the WP6 work package (also called SA3) of EGI-InSPIRE. The effort of the distributed WP6 team of EGI-InSPIRE is targeted towards the provision of shared services that will ease the porting of new applications from these scientific domains to the wider grid by detecting and exploiting commonalities between VOs and driving the implementations to a generic direction.

At the same time inter-VO collaboration results typically not only in more powerful solutions, but also saves significant amounts of manpower in the long run. Such benefits would be unlikely to be achieved with generic support structures, both for individual large communities such as HEP (which could otherwise develop multiple similar solutions to basically common problems), as well across disciplines (e.g. the usage of Dashboards, Ganga and Hammercloud across communities, described in more detail in the Technology Section 6).

In conclusion HUCs can offer benefits not only to new adopters of grid technology but also to each other. This continues to be demonstrated, both by the adoption of tools initially developed for one community spreading to others, as well as at the conceptual level: less optimal but offering a more pragmatic solution for existing communities.

3.1 Who are the Heavy User Communities in EGI-InSPIRE?

The purpose of this section is to offer an overview of the current HUCs in EGI and the capabilities and services they offer (see Table 1). For additional information please refer to the following link: https://wiki.egi.eu/wiki/WP6:_Services_for_the_Heavy_User_Community.

[2] VRCs are the entities that EGI collaborates with. They are defined as groups of researchers, typically residing in more than one country, working together effectively through the use of information and communications technology (ICT) and having an effective coordination model that can articulate prioritized requirements and other needs on behalf of the community.

<table>
<tr><th>Community</th><th>Description, capabilities and services offered</th></tr>
<tr><td>High Energy Physics
[R 4]</td><td>The High Energy Physics (HEP) HUC represents the 4 LHC experiments at CERN, which are fully relying on the use of grid computing for their offline data distribution, processing and analysis. The HEP computing systems are probably the most complex grid-integrated applications currently in production. The services ran by the HEP HUC can be classified into:
Experiment computing systems and services: Software stacks developed by the experiments on top of the WLCG middleware to implement their particular computing models. Nevertheless, successful examples of experiment computing systems reuse exist, e.g. the Linear Collider Detector (see Section 5.3.2)
Middleware services: VO independent, high-level grid services
Data Management: Services for data discovery and data transfer, e.g. LCG File Catalogue, gLite File Transfer Service, WLCG Disk Pool Manager etc. See Section 6.7 for a complete description of each service.
Workload Management: Services that allow users to submit and manage generic batch jobs on grid resources, e.g. Ganga (see Section 6.2), gLite Workload Management System etc.
Persistency: Framework to interface database access for storing and retrieving different types of scientific data, such as event and conditions data. [R 4]
Monitoring: Application and site monitoring to follow the experiment activities and the state of the grid infrastructure respectively, e.g. Dashboards (see Section 6.1), SAM/Nagios monitoring and HammerCloud (see Section 6.6)
EGI InSPIRE WP6-SA3 supports and contributes development effort to some of these services in particular through subtasks TSA 3.2 .1, 3.2.2 and 3.3.</td></tr>
<tr><td>Earth Sciences</td><td>Earth Science (ES) applications cover various disciplines like seismology, atmospheric modelling, meteorological forecasting, flood forecasting and many others. Their presence in SA3 is currently centred in the implementation, deployment and maintenance of the EGDR service to provide access from the grid to resources within the Ground European Network for Earth Science Interoperations - Digital Repositories (GENESI-DR).
The ES HUC includes also researchers and scientists working in the climate change domain. In particular most of them actively participate in the Climate-G use case (see Section 5.5). This use case exploits the GRelC service (see Section 6.3) for distributed metadata management and the Climate-G portal as scientific gateway for this collaboration.</td></tr>
</table>

Community	Description, capabilities and services offered
Life Sciences [R 5]	The Life Science (LS) HUC originates from the use of grid technology in the medical, biomedical and bioinformatics sectors in order to connect worldwide laboratories, share resources and ease the access to data in a secure and confidential way through the health-grids. A non-extensive list of the services supported and / or developed by this HUC is given below: • **Grid database management service:** The GRelC service (see Section 6.3) will be exploited to implement some LS specific use cases. In the coming weeks a questionnaire will be sent to the SA3 mailing list to get feedback about available data sources that need a grid-database access interface (porting in grid of existing/legacy databases). • **Metadata catalogues:** AMGA • **Workflow engines:** MOTEUR2 (see Section 5.5) and Taverna (http://www.taverna.org.uk/) • **File encryption service:** Hydra [R 5] • **Monitoring:** Dashboard and Hudson infrastructure monitoring [R 5]
Astronomy and Astrophysics	Originally, Astronomy and Astrophysics (A&A) institutes usually acquired the necessary resources on local computing facilities and quite often also contracted the access to a pool of resources at supercomputing centres. However, some A&A applications grew in complexity and were considered very suitable to be run on grid infrastructures. At the moment the A&A HUC is devoted to the evaluation of different solutions for the gridification of a rich variety of applications. In particular the work is focused on: • **Visualization tools:** VisIVO (http://visivo.oact.inaf.it/index.php) for the visualization of data collected by different A&A projects • **Databases and catalogues:** Currently the HUC is evaluating the possibility of adopting and tailoring different tools already in production (e.g. AMGA, GRelC, etc.). • **Parametric applications:** Use DIANE (see Section 6.2) to study parametric space applications. Additionally, one of the main goals is the accomplishment of a good level of interactivity among different technologies related to supercomputing, i.e. High Performance Computing and High Throughput Computing, Grid and Cloud.

Table 1 – Overview of the Heavy User Communities in EGI InSPIRE

3.2 "The Knowledge"

Aside from specific details about the services and tools for the Heavy User Communities and how they could be applied elsewhere, perhaps the most tangible benefit that the HUCs can offer is a huge wealth of knowledge and experience. Whilst this work has been extensively documented through EGEE milestones, deliverables and reports, there is much to be gained by a short session of informal consultancy that can address many of the issues described in detail below.

Having read this document, communities interested in going further would be well advised to contact an appropriate member of the HUCs, based on the contact points detailed in MS601 (HUC Contact points and the support model), as well as on the wiki pages at https://wiki.egi.eu/wiki/WP6:_Services_for_the_Heavy_User_Community.

4 USER SUPPORT IN EGI AND EGI-INSPIRE

The growing user demands have provided, and will continue to provide, the necessary push for development and extension of the grid infrastructure. It must be clear that therefore the active support of these communities is a primary concern for the EGI.eu / NGI ecosystem and one of the primary goals in EGI is to provide significant added value for both existing and new user communities.

User support activities in EGI must define, implement and operate services and tools that enable users to use infrastructure services and to develop themselves into self-sustainable grid communities. Self-sustainable communities are capable of some or all of the following:

- port new applications to the infrastructure
- extend the infrastructure according to their members' needs
- expand their user base without depending (heavily) on user support services of EGI

Self-sustainability of user communities guarantees permanent usage of the grid infrastructure, continuous flow of scientific results from applications, and as an overall result, a European-wide interest in sustaining the EGI collaboration.

Partners within the EGI-InSPIRE project, specifically within WP3 (User Community Coordination) and WP6 (Services for Heavy User Communities) are the key stakeholders for defining and providing the user support processes in EGI.

5 CASE STUDIES

5.1 *Case Study: EnviroGRIDS [R 6]*

5.1.1 Introduction

The Black Sea catchment is internationally known as an area of ecologically unsustainable development and inadequate resource management. This has led to severe environmental, social and economic problems. The **EnviroGRIDS @ Black Sea** catchment project addresses these issues by bringing together several emerging information technologies that are revolutionizing the way we are able to observe our planet. The project is developing a system that aims to assist governments and communities to track and respond to environmental trends in the Black Sea catchment.

The Group on Earth Observation Systems of Systems (GEOSS) is building a data-driven view of our planet that feeds into models and scenarios to explore our past, present and future. EnviroGRIDS aims at building the capacity of scientists to assemble such a system in the Black Sea catchment, the capacity of decision-makers to use it and the capacity of the general public to understand the important environmental, social and economic issues at stake. EnviroGRIDS will particularly target the needs of the Black Sea Commission (BSC) and the International Commission for the Protection of the Danube River (ICPDR) in order to help in bridging the gap between science and policy.

CERN is a partner in this project, providing grid expertise and specific support for porting the "SWAT" (see below) application to the grid environment. Recently, the project has asked to have its own VO created by the CERN team and supported by the CERN VOMS infrastructure.

By partnering with a member of the Heavy User Communities, this project benefits from:

- Access to consultancy and expertise that represents many hundreds of years of combined grid and data management knowledge;
- Access to infrastructure services – e.g. the VOMS services at CERN – at a marginal cost with respect to dedicated services for the project;
- Specific support for tools such as Ganga and Dashboards – again available at a marginal cost with respect to bespoke applications for a single community or project.

5.1.2 SWAT

SWAT is the acronym for Soil and Water Assessment Tool, a river basin, or watershed, scale model. It was developed to predict the impact of land management practices on water, sediment and agricultural chemical yields in large complex watersheds with varying soils, land use and management conditions over long periods of time.

5.1.3 Gridification of SWAT

Gridification of SWAT in EnviroGRIDS means creating the possibility for parallel execution on grid nodes. The key points are integration with existing tools and users habits and to control grid instabilities (execution errors due to the infrastructure and minimize / make dependable the latency as seen by scientists. Here we list the main use cases:

- **Single SWAT run:** we can gridify a single SWAT run by splitting the SWAT model based on sub-basins and followed by routing and running each sub-model separately on the grid.

- **LH-OAT sensitivity analysis:** parameter sweep; requires hundreds of iterations; in this use-case we provide a SWAT model with different parameter sets and find out which parameters are the "relevant" for the model under study. This is achieved by observing the effect of parameter variation on the quality of the simulation (e.g. nitrogen transport and transformation in soil). In this case we run on the grid the same SWAT model using different parameter sets.
- **SWAT-CUP calibration and uncertainty analysis tool:** SWAT-CUP is a computer program (ran on scientists' desktops) for the calibration of SWAT models. Calibration and uncertainty analysis also requires running SWAT model simulations hundreds of times (parameter sweep). Gridification here is focused on the sufi2 algorithm (the only SWAT-CUP algorithm that can be parallelized). The overall gridification operating principle is the same as for the LH-OAT algorithm but it is required that Ganga and DIANE software (see Section 6.2) were used to run SWAT and SWAT related software in parallel mode, which greatly reduced the effort needed to port SWAT into the grid. In addition Ganga and DIANE allow to achieve the two main goals already mentioned, namely to have a dependable behaviour of large simulation in terms of execution time and of error control (automatic error recovery and dynamic task distribution to minimize completion waiting time).

5.2 Case Study: PARTNER [R 7] and ULICE [R 8]

These "physics for health" projects – in which CERN is also a partner – are linked to the core business of the laboratory primarily through the use of particle accelerators as hadron sources. As is the case with EnviroGRIDS, what is of much more value to them is access to the considerable knowledge base through the WLCG project and the Experiment Support group in CERN's IT department. Whereas data management is an important issue for these projects, the concept of *Virtual Organisations* as a means of tying together multiple medical facilities into one logical whole is also of key value. Once again, partnering with a member of the Heavy User Communities results in huge resource savings together with access to a large amount of expertise and services at small incremental cost.

The PARTNER project has performed a requirements analysis based on a small number of specific Use Cases. These are described in detail in the following project deliverable document and can be briefly summarized as a list of general requirements from the health-grid perspective:

- User requirements: The software stack should be easy to deploy, to use and to maintain and be robust and fault tolerant
- Technical requirements
    - Compatibility with existing hospital data and image archiving systems (standards like DICOM, HL7)
    - Security should enforce legal policies and ethical guidelines without overhead on end-users
    - Framework for negotiating agreements
    - Framework for ensuring data interoperability
- Domain specific requirements
    - Data interoperability for the hadron therapy field

Although the project is not currently "running in the grid" (the project is evaluating grid services for eventual production usage in a later stage), the above gives a good example of the sort of application specific analysis that needs to be done in evaluating grid services and tools. It once again underlines the need for close cooperation between people expert in the application domain as well as those versed in grid technology – only by close collaboration between the two can a success story be built.

5.3 Case Study: Support for a new VO

Establishing support for a new Virtual Organisation includes a variety of steps, starting with a request that is currently made using the CIC portal.

This, however, is just a first step that needs to be performed, after which (or preferably in parallel, or beforehand) the more complex issue of resource allocation must be addressed. This is a point that is often overlooked – whilst there may be resources available "on loan" for pilot activities, the bulk of grid resources are provisioned for specific purposes, such as the Resource Pledges made for the WLCG project.

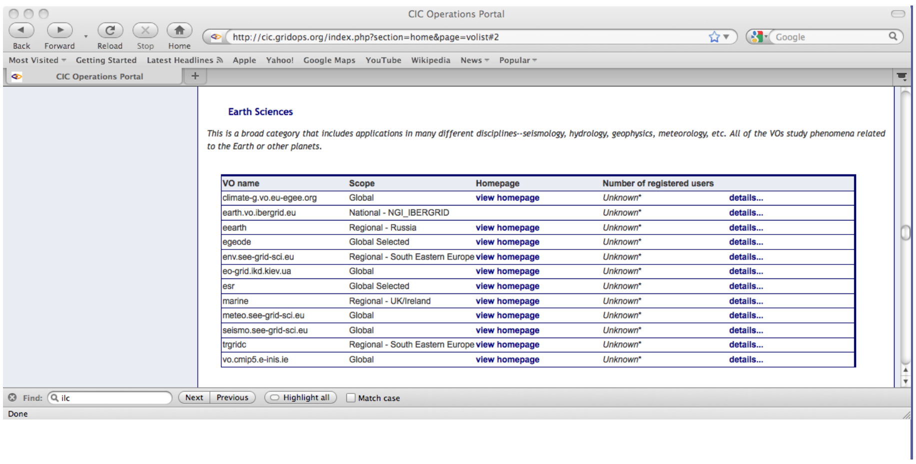
The information listed below is based on the following CERN wiki pages: https://twiki.cern.ch/twiki/bin/view/LCG/EnablingNewVOATCern and has been applied to a number of VOs, including the International Linear Detector [R 9] project – a recognised High Energy Physics project at CERN – as well as the EnviroGRIDS Earth Science project.

Enabling support for a new VO at other grid sites may well differ in detail – what is important to retain is that it is a non-trivial operation that typically has both resource (computer and storage hardware) and support issues.

Consequently, the use of an existing VO might be more appropriate for smaller / shorter projects and / or for test activities.

5.3.1 Enabling a VO without access to resources (EnviroGRIDS)

This boils down to registering the VO and then configuring the agreed Virtual Organization Membership Service (VOMS) infrastructure as required. The following example is for the EnviroGRIDS VO, created in the Earth Science category.



VO name	Scope	Homepage	Number of registered users	
climate-g.vo.eu-egee.org	Global	view homepage	Unknown*	details...
earth.vo.ibergrid.eu	National - NGI_IBERGRID		Unknown*	details...
eearth	Regional - Russia	view homepage	Unknown*	details...
egeode	Global Selected	view homepage	Unknown*	details...
env.see-grid-sci.eu	Regional - South Eastern Europe	view homepage	Unknown*	details...
eo-grid.ikd.kiev.ua	Global	view homepage	Unknown*	details...
esr	Global Selected	view homepage	Unknown*	details...
marine	Regional - UK/Ireland	view homepage	Unknown*	details...
meteo.see-grid-sci.eu	Global	view homepage	Unknown*	details...
seismo.see-grid-sci.eu	Global	view homepage	Unknown*	details...
trgridc	Regional - South Eastern Europe	view homepage	Unknown*	details...
vo.cmip5.e-inis.ie	Global	view homepage	Unknown*	details...

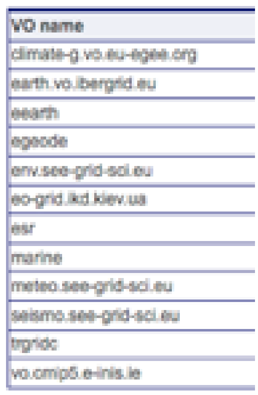


Figure 1 – Starting point: without EnviroGRIDS VO

Having completed the registration form for new VOs on the CIC portal, the new VO shortly appears as below:

Last Update of the VO Card	**29/07/2010 14:56**
Unique VO Name	envirogrids.vo.eu-egee.org
Discipline	Earth Sciences
Status	new
Scope	Global
VO Homepage/link	**http://www.envirogrids.net/**

Figure 2 – EnviroGRIDS appearing as new VO

Once this request has been made it has to be validated by the EGI.eu User Community Support Team (UCST)

As part of the validation process, the VO Manager has to complete the Acceptable Use Policy (AUP) duly compiled in all its parts. The VO Acceptable Use Policy (AUP) is a statement, which, by clearly describing the goals of the VO, defines the expected and acceptable usage of the grid by the members of the VO. By requiring that all members of the VO who participate in the grid agree to act within the constraints of the VO AUP the VO Manager defines a community of responsible users with a common goal. This definition enables Site Managers to decide whether to allow VO members to use their resources.

When the validation process has been completed a broadcast message is issued as shown below.

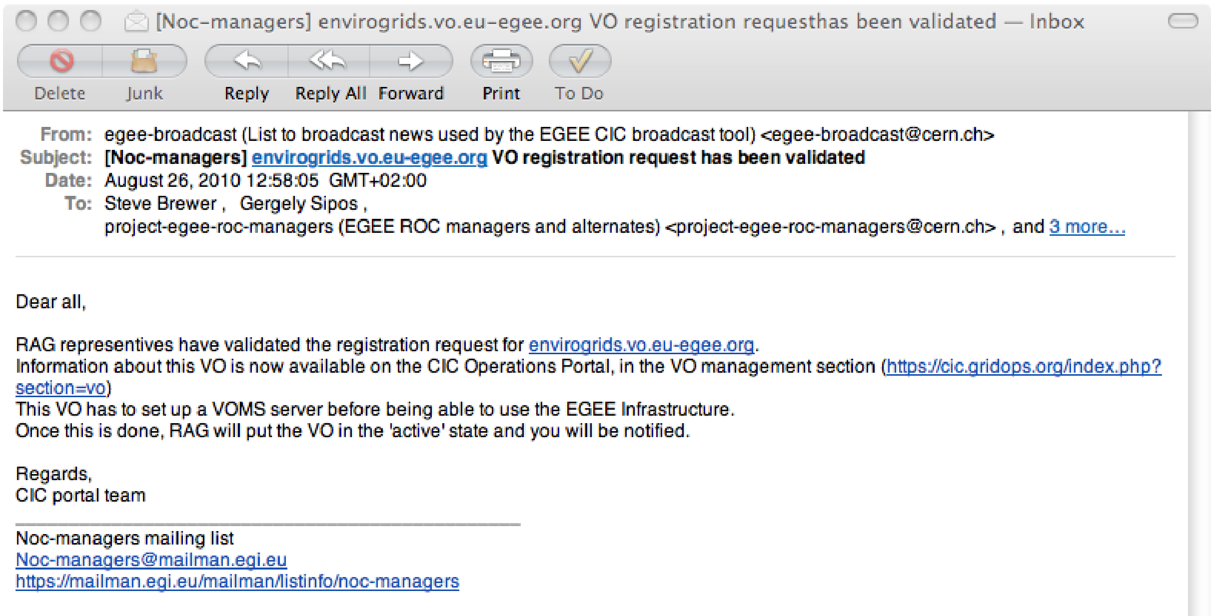


Figure 3 – Confirmation of EnviroGRIDS as new VO

5.3.2 Case Study: Enabling a VO with compute / storage requirements (LCD)

The Linear Collider Detector is a HEP project at CERN that focuses on physics and detector issues for a future e+e- collider at the TeV-scale. The work is carried out in collaboration with several study groups from across the globe, addressing physics and detectors at either the International Linear Collider (ILC) or the Compact Linear Collider (CLIC). ILD, the international large detector concept, is one of several detector concepts that are being studied for the International Linear Collider (ILC). The primary grid-related application for this community is detector studies, based on the DIRAC framework of the LHCb experiment. Their motivation for access to grid resources stems from the need to seamless run simulations at a number of (grid) sites, together with time pressure to produce reports, based on conference schedules, plus the schedule for the production of the related ILC technical design report.

Within WLCG, there is a dedicated body to handle computing resources across the different sites (Tier0, Tier1, Tier2) – the Computing Resources Review Board. At a specific site – in this case CERN – resource management is the responsibility of the WLCG Resource Manager. Once a given project or experiment is approved by the Scientific Management of the laboratory, it is the Resource Manager's task to ensure that the request fits in the overall resource envelope of the site.

In the case of a new approved VO, some initial discussions between the VO representative(s) and the Resource Manager are required to determine whether the use of grid resources is indeed appropriate (as opposed to simply using a local batch), as well as the overall requirements in terms of CPU, disk, tape, network, services, etc. Should the green light be given, tickets are issued to the services concerned (batch, storage services) and a detailed checklist is followed to ensure that all of the necessary setup and configuration is performed. This list is somewhat extensive and clearly only makes sense for projects of a certain size and duration.

Some specific details of the project's resource requests are given below for sizing purposes.

- $51 * 10^9$ kSI2K seconds of CPU time
- 67.5 TB of mass storage, 1.0 TB of Disk and 6 TB of Disk cache
- Enabling the ILC VO to use the CERN resources
- VO box server for the LCD project (with typical LHCb VO box profile and special accounts for MySQL and DIRAC)

For the year 2011, similar needs of CPU time and data storage are expected.

[To be compared with the resource requests for the LHC experiments, where CPU is specified according to the HEP-SPEC06 benchmark results.]

In this specific case, one can expect that support for this project piggybacks on that for LHCb. This is clearly a big potential saving and such re-use of technology between projects is clearly to be encouraged. The project also uses an existing VO – ILC – but one that was not previously enabled to use CERN resources, nor for which resources were provisioned.

It is not the purpose of this document to act as a guide to system administrators / service managers on the steps that need to be taken in order to support a new VO with access to computational and storage resources at a large grid site such as CERN, however the number of steps that need to be followed to configure and verify the setup are based on being followed by a dedicated contact person on the experiment / grid support side, who works in close contact with other groups in IT together with the experiment / project in question.

5.3.3 Grid Support Issues

Historically, all new applications that have been ported to the grid have required both initial as well as on-going support. Much of this was provided via the EGEE NA4 work-package, together with corresponding infrastructure (neither funded nor matching funding) in a number of key sites, such as CERN. The continuation of this activity was expected to be via the ROSCOE/SAFE/TAPAS projects, which have not been funded. Some limited effort for such activities is available through this work-package, e.g. for official CERN-recognised experiments and projects, but this effort of necessity competes with the mainstream activities of this work package. Some complementary effort can also be expected from the EGI TNA3.4, VO Services activity, which helps new communities to support and operate their VOs.

Nevertheless, full support for a new VO – as witnessed by the above checklist – is a non-trivial operation and requires the involvement of a number of groups and services, plus preferably a dedicated contact person on both the CERN and VO sides, at least during the duration of the setup and initial testing. On-going support, for both integration and operations purposes, is clearly needed by the larger VOs, such as the LHC experiments.

5.4 Case Study: Fusion

The fusion community makes use of top computational environment with a wide range of applications, tools and developments. Several different studies and use cases can be established by means of these applications and tools. These studies can be focused on various aspects of fusion research, like NBI heating, transport, turbulences or MHD equilibrium. Considering these different areas and the possible relations among them, some application workflows could be also considered in order to carry out more complex simulations. Thus, the impact of the heating in the evolution of the trajectories of particles in the confined plasma arises as an important use case.

As stated, there is a wide range of applications. Several of them are Monte Carlo codes that are perfectly suitable for the grid. There are also Message Passing Interface (MPI) applications that can be used. In order to perform these studies, the different applications must be coupled. In this coupling Kepler (see Section 6.4) plays an important role in order to assure this procedure is carried out transparently to the user. This tool allows the exchange of data among the applications involved in the study, the management of the execution of the different components and a user-friendly environment to use the computational resources.

5.5 Case Study: Earth Sciences and Environmental domains

In the Earth Science and Environmental domains, the Climate-G testbed represents an important case study. Climate-G is a multidisciplinary collaboration involving both climate and computer scientists and currently includes several partners such as: Centro Euro-Mediterraneo per i Cambiamenti Climatici (CMCC), Institut Pierre-Simon Laplace (IPSL), Fraunhofer Institut fur Algorithmen und Wissenschaftliches Rechnen (SCAI), National Center for Atmospheric Research (NCAR), University of Reading, University of Salento and the Italian Southern Partnership for Advanced Computational Infrastructures (SPACI).

The Climate-G testbed provides an experimental large-scale data environment for climate change addressing challenging data and metadata management issues. The main scope of Climate-G is to allow scientists to carry out geographical and cross-institutional climate data search & discovery, access, visualization and sharing.

The GRelC software (see Section 6.3) is exploited in this use case to perform metadata management, harvesting functionalities and data search & discovery. From the metadata management point of view, the testbed involves relational and non-relational (XML-based) data sources hosting Climate-G projects and experiments metadata information.

A new version of the GRelC service is now tested to manage climate metadata. Also a new version of the Climate-G portal has been deployed (http://grelc.unile.it:8080/ClimateG-DDC-v2.0/) and tested in July. The new version fixes some bugs related to the search and discovery functionalities as well as to the metadata extraction and presentation. New requirements and scenarios will be gathered and defined by talking with the Earth Sciences representatives involved in the EGI-InSPIRE project. A virtual organization (named climate-g.vo.eu-egee.org and defined with global scope during the EGEE-III Project) will support distributed testing activities on the Climate-G testbed.

5.6 Case Study: The Computational Chemistry, Molecular Science and Technology Community

5.6.1 Introduction

The target of the Computational Chemistry, Molecular Science and Technology community is the study of properties of the matter referable to the characteristics of the nuclei considered as stable (though structured) entities and as (from weak to strong) aggregates of atoms and molecules. The purpose of such study is to invest the knowledge developed in this way to build new science, innovation, technologies and services for social and economic growth.

Accordingly, the engagement of the community in computation is addressed to a series of research activities, computer codes (either commercial or in-house implemented), database construction and management, data rendering and virtual reality handling. These activities are fragmented into a large number of laboratories that concur to carry out advanced modelling and simulations based on multi-scale and multi-physics approaches to reality starting, whenever is appropriate, from the nano-scale level. This means that there are several important applications and a large number of in-house developed methods and software packages with a high innovative potential that benefit from the efforts of a large number of researchers and may serve an even larger number of users.

The success of the community is visible in the implementation of highly successful quantum chemistry packages (like Gamess, Gaussian, etc.), which are at present solid foundations of any determination of molecular structures and properties. Their applications range from materials design to photo-assisted processes rationalization and spectroscopic analytic studies. With the advent of the grid more and more of these studies are becoming "flagship applications" although the natural computational platforms are supercomputers.

5.6.2 The present situation

The Computational Chemistry, Molecular Science and Technology community is presently composed of two pan-European grid VOs constituted during the final years of the EGEE project (UNIPG and CYFRONET) as can be seen at http://cic.gridops.org/index.php?section=home&page=volist:

1. **Gaussian:** mainly devoted to implement and maintain codes and to provide support on the grid to computational chemists carrying out their research using commercial software (as is the popular Gaussian package) for predicting in an ab initio fashion the structure and the electronic properties of molecules and molecular aggregates. Related calculations are of invaluable help in

applicative fields as spectroscopy, chemical analysis, materials, biochemistry, drug design, energy conversion and storage, etc.

2. **CompChem:** mainly devoted to implement codes and provide support on the grid to computational chemists carrying out their research by combining structural, dynamical and kinetics studies of chemical processes using components of ad hoc workflows linking ab initio calculations to the experimental signal. The process was accompanied by the cooperation activities of some working groups of the Actions D23 and D37 funded by the European initiative COST.

To the same Computational Chemistry (CC) cluster of EGEE III belong three national VOs that consist of more than 150 national researchers. The two most important of these national VOs are associated with the Computational Chemistry and the Molecular Science communities of Turkey. The researchers who belong to these VOs are mainly running chemistry and molecular dynamic packages such as VASP, NAMD, Espresso, Gaussian'03, Gaussian'09, Charmm, CPMD, Gamess, Gromos, Gromacs and Amber.

At the same time other geographically dispersed minor groups have expressed the intention to join the grid activities of the community though they were not active in EGEE. Altogether this makes a block of more than 500 people that refer to EGI as a coordination body for distributed computing. Progress of this community in grid computing was relying on three different proposals (ROSCOE, HIPEG, SCHEMA respectively presented by C. Loomis, A. Laganà and E. Rossi) that failed to succeed at the INFRA 2010 call.

However, despite the failure in getting funded, some basic activities of the community are being carried out albeit at slower pace and on a best effort regime. In particular:

- The activities of the VOs are being continued
- The hardware patrimony has been saved and consolidated
- The program library has been kept alive and updated
- The support to the users has been continued
- Some tools facilitating the use of the grid by the members of the community were further developed.

The intention of the mentioned CC VOs is therefore to act at present as a hub for its community, coordinating computing activities on heterogeneous platforms, providing support for its members, safeguarding its knowledge as well as preserving and maintaining a set of services on the best effort basis. The activities of the VOs of the community have already produced in the final period of EGEE III a significant volume of activity (see Figure 4) that would be a pity to disperse (and that has to be added up to the traditional high share of computing resources traditionally used by the community on large scale facilities that is usually allocated by research projects).

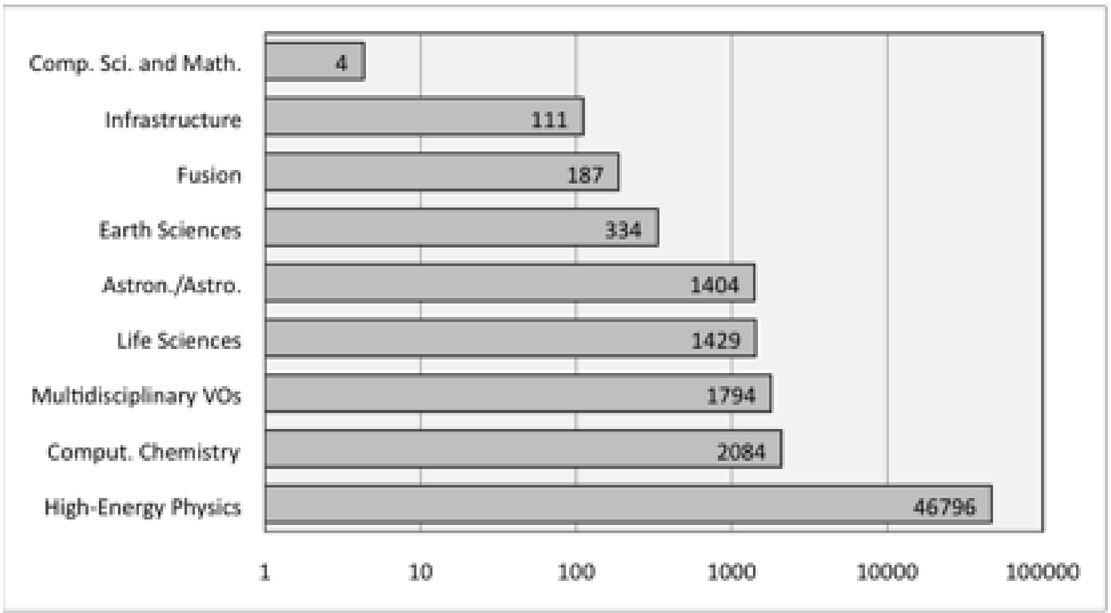


Figure 4 – Approximate number of CPU's used by different communities on EGEE (9/2008-8/2009) (from EGI-InSPIRE proposal)

5.6.3 The work plan

The originally proposed contribution of the CC community is given in the footnote[3], where the community that was intended to be structured as a Specialized Support Centre (SSC) is identified as CMST (Chemistry and Molecular Science and Technology). The final effort assigned to the community within EGI-InSPIRE, is the support to the use of MPI within its members. This is part of the effort made in porting to the grid the above-mentioned packages and in-house developed programs. Detailed report of this activity will be given as part of the progress report. The plans are to test the use of MPI and develop MPI versions of examples belonging to three categories of products: linear algebra routines, in-house developed codes and licensed packages. However, as a best effort voluntary contribution of the community to EGI progress will be made to deploy the InsilicoLab portal developed by CYFRONET as an example of user centric work environment etc.

The Gaussian VO is going to become the standard testbed for the InsilicoLab environment. All the work for Gaussian VO is done on best effort basis. The main engagement of recent times has been the updating of the package Gaussian to version 09. Work will also be spent to enrich the offer of products by enabling the use of Turbomole, NWChem (freely available) and ADF packages. A rescue operation of the users of Turbomole (due to the planned closure of the Turbomole VO) is being designed by incorporating them into the Gaussian VO.

The best effort contribution of the CompChem VO will consist in extending and consolidating the library of programs implemented on the grid. In particular, efforts will be devoted to implement a grid version of some of the programs on the GPUs (typically the bench program used for that is either ABC or RWAVEPR) in order to enrich the articulation of the GEMS simulator. A second important line of best effort commitment of the CompChem VO will be that of developing a facilitator for the

[3] **CMST contribution to EGI proposal (Version 0.2 August 14)**

Antonio Lagana, Mariusz Sterzel

gridification of the programs on the grid that, moving along the line of evaluating QoS and QoU, does not need any knowledge of the grid to submit jobs on the grid.

5.6.4 Services for other communities

Currently Computational Chemistry VOs are open for all communities looking for access to molecular dynamics and ab initio/first principle methods. Available software portfolio, workflow engines and tools are accessible on a standard basis via Computation Chemistry VOs to any person with a valid certificate. Also, it is highly recommended for beginners looking for vast amount of computational power in a grid environment to start using the InsilicoLab portal. Currently, aside from chemists, the CC resources are mainly utilized by users from biology and related fields, as well as by users from the earth science community.

6 TECHNOLOGY

In the following sections we consider some of the shared technologies that are available today, followed by an explicit discussion of data management. This list is not exhaustive but covers some of the main tools and services that are supported through this activity.

6.1 Dashboards

6.1.1 Introduction

In order to perform production and analysis tasks across a highly distributed system crossing multiple management domains, powerful and flexible monitoring systems are clearly needed. The Experiment Dashboard monitoring system was originally developed to support the four main LHC experiments (LHCb, CMS, ATLAS, ALICE). This framework not only supports multiple grids / middleware stacks, including gLite (EGEE), VDT (OSG) and ARC (NDGF), but is also sufficiently generic to address the needs of multiple user communities including, but not limited to HUCs.

Furthermore, the system covers the full range of the experiments' computing activities (job monitoring, data transfer, site commissioning etc.) and addresses the needs of different categories of users (computing teams of the LHC VOs, VO and WLCG management, site administrators and VO support at the sites, users running their computational tasks on the grid infrastructure etc.)

Experiment Dashboard applications are widely used by the LHC experiments for their everyday work and receive over a thousand unique visitors every day and this number is steadily growing.

6.1.2 The Dashboard framework

All Experiment Dashboard applications are developed using the Dashboard framework, which is implemented in Python. The Dashboard framework defines the structure of Dashboard modules. It provides the build system and a common implementation of the main components of monitoring applications and services. The typical structure of a Dashboard application consists of information collectors, a data repository, normally implemented using Oracle database technology, and user interfaces. Information from data sources can be transmitted to the Dashboard application via various protocols. In most cases, the Dashboard application uses asynchronous communication between the data sources and data repository. Dashboard agents are the components that have to perform regular tasks. Examples of such tasks are: collecting monitoring data, aggregating statistics, analyzing statistics, sending alarms, etc. The framework provides all necessary tools to manage and monitor Dashboard agents.

6.1.3 Site monitoring

LHC virtual organizations represent one of the main user communities of the grid infrastructure and therefore monitoring LHC computing activities provides the best estimation of grid reliability and performance. Constant improvement of the quality of the grid infrastructure is achieved due to LHC site commissioning activity and LHC computing shifts. One of the most important functions of the Experiment Dashboard is to provide necessary information for LHC computing shifts and LHC site commissioning. LHC computing shifts follow up the status of the distributed sites and services and their capability to perform the tasks of the LHC virtual organizations. The result of the visible improvement quality of the distributed sites is measured and shown in the Dashboard Site Usability Monitor (see Figure 5).

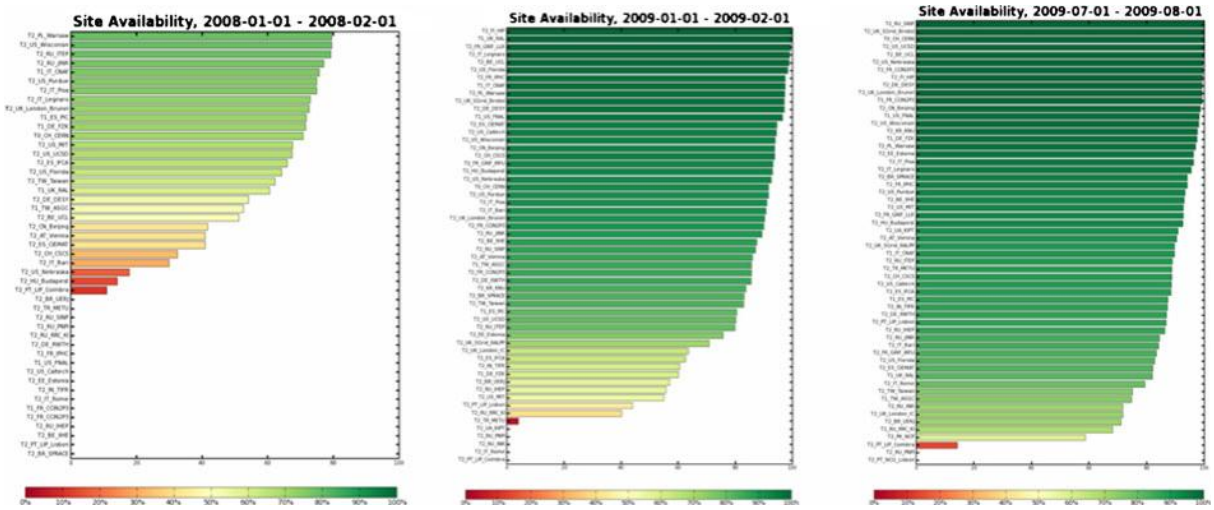


Figure 5 –Shows monthly availability of the sites used by CMS. Every bar corresponds to a particular site – availability is in the range 0-100%. *Green* corresponds to high availability and *red* to low availability

The Site Status Board (SSB) is another example of a Dashboard application that was developed for monitoring the distributed sites and services from the perspective of a particular virtual organization. SSB provides a very flexible framework enabling a customizable metric store and user interface exposing runtime and historical monitoring data.

For power users (e.g. managers of key activities such as large simulation campaigns in HEP or drug searches in LS communities) a very important feature is to be able to monitor this resource behaviour to detect the origin of failures and optimize their system. They also benefit from the possibility to measure efficiency and evaluate the quality of the service provided by the infrastructure.

6.1.4 Application monitoring

The importance of flexible monitoring tools focusing on applications has been demonstrated to be essential not only for power-users, but also for single users. Single users are typically scientists using the grid for data analysis, and verifying hypotheses on data sets that could not be achieved on other computing platforms. In this case, monitoring, provided by Experiment Dashboards, is a guide to understanding the progress of their activity, and identifying and solving problems connected to their application.

It is essential to allow efficient user support by empowering the user in such a way that only non-trivial issues are escalated to support teams (for example, jobs on hold due to scheduled site maintenance can be identified as such and the user can decide to wait or to resubmit). The Task monitoring application was developed in order to facilitate user analysis on the distributed infrastructure by providing necessary monitoring information to analysis users and to analysis support teams.

6.1.5 Building generic Dashboards

The Experiment Dashboard applications obtaining information from sources that are not VO-specific can be used by VOs outside the LHC and HEP scope. Among those applications are Site Status Board and Site Usability interface. A generic job monitoring application that will use job status information from the Messaging System for the Grids (MSG) is now being prototyped. Job status information can be reported to MSG either from the new version of the Logging and Bookkeeping system or from VO-specific frameworks. The Dashboard team provides generic methods enabling instrumentation of VO-specific frameworks for MSG reporting.

Future work will concentrate effort on common applications which are shared by multiple LHC VOs, but which could also be used outside the scope of LHC and HEP.

References:

Experiment Dashboard for monitoring computing activities of the LHC virtual organizations

Journal of Grid Computing, Volume 8, Number 2, 323-339, DOI: 10.1007/s10723-010-9148-

6.2 *Ganga & DIANE [R 10 - 14]*

Ganga is an easy-to-use frontend for job definition and management that provides a uniform interface across multiple distributed computing systems. Ganga has been developed to meet the needs of the ATLAS and LHCb user communities and is heavily used and supported by these users. It includes built-in support for configuring and running applications based on the Gaudi / Athena framework which is common to the two experiments. However, it also provides functionalities for running of a wider range of applications, including arbitrary executables. Ganga was designed in a way that makes it very easy to use for parameter-sweep applications and to realize more complex application use cases. It allows trivial switching between testing on a local batch system and large-scale processing on grid resources and effectively provides a simple but powerful abstraction layer for distributed computing environments. As it is based on a plugin system, it is readily extended and customised to meet the needs of different user communities. It has also been used in EGEE tutorials to introduce new users to the grid.

Functionality provided by Ganga includes a command line interface for job configuration, submission and management; automatic configuration of complex applications; automatic handling of large datasets as job inputs and outputs; splitting of jobs for parallel processing and subsequent merging of outputs. The supported backends for general use are following: Portable Batch System (PBS), Load Sharing Facility (LSF), Sun Grid Engine (SGE), Condor, gLite WMS/CREAM-CE, ARC, Globus/GRAM, GridWay and the SAGA API standard.

The usage of Ganga has been increasing in the last years and is presented in the Figure 6. On average, there are 500 users every month using the tool for their everyday work.

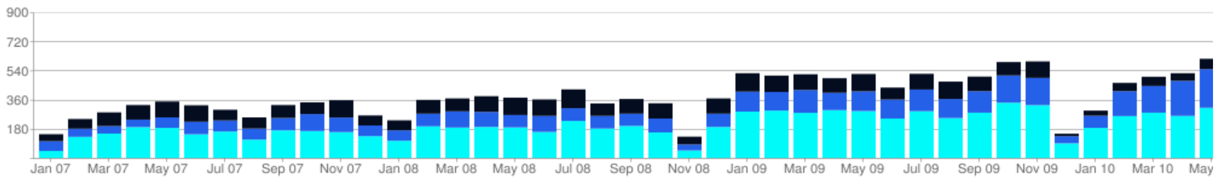


Figure 6 – Monthly number of Ganga users in 2007-2010

DIANE is a lightweight task processing framework which allows more efficient and robust execution of large numbers of computational tasks in unreliable and heterogeneous computing infrastructures. It exploits the late-binding method (also known as pilot jobs or worker agents) and provides an application-aware scheduler that may be extended by a system of plugins to support master/worker workloads such as task farms and bag of tasks. Plugins for Direct Acyclic Graph (DAG) and data-oriented workflows have been implemented as third-party contributions by a number of interested user communities. The framework also supports customized, application-specific processing methods and failure-management strategies. DIANE worker agents are often submitted using Ganga. This integration of the tools allows for a dynamic combination of resources across several distributed computing infrastructures, including High Performance Computing (HPC) resources.

As an example, a Ganga / DIANE overlay was interface to the MOTEUR workflow engine (Section 6.5) and is run as a service for the biomed community in the context of GATE medical imaging applications. Between July 2009 and August 2010 more than 360 DIANE master instances were activated in the service backbone that handled 58000 worker agent jobs completing more than 113000 simulation tasks.

Besides the LHC communities, the use of Ganga/DIANE tools has been reported for UNOSAT applications, Geant4 medical and space simulations and grid-enabled regression testing, AvianFlu Drug Search, ITU digital broadcasting planning, LatticeQCD simulations, Fusion, Image Processing and Classification, EnviroGrids and Simulation of Gaseous Detectors (GARFIELD).

Both Ganga and DIANE include Dashboard plugins that provide the user communities with on-line monitoring of executed tasks. Both tools are part of RESPECT.

6.3 Grid Relational Catalog - GRelC

The Grid Relational Catalog Project (GRelC) aims at providing a set of advanced data grid services to manage databases on the grid in a transparent, user-friendly, efficient and secure manner.

At the moment the GRelC Data Access and Integration Service (GRelC DAIS) allows users to access and interact with different DBMSs, both relational (PostgreSQL, MySQL, SQLite, etc) and non-relational (eXist, XIndice, XML flat files, etc), providing a uniform access interface to heterogeneous data sources.

It works well in concert with the gLite software by expanding the functionality of the grid infrastructure and provides 2 different kinds of clients: the Command Line Interface (CLI), available as a single rpm, and the GRelC Portal. The GRelC Portal (available from the "GRelC Portal" link on the GRelC website – http://www.grelc.unile.it) offers all the functionalities provided by the CLI in a direct way:

- submit queries
- manage the enterprise grid
- manage users and their privileges
- define virtual spaces and their properties
- manage VOs
- manage several GRelC DAISs at the same time
- view metadata information concerning the schema of your database/tables

The GRelC DAIS is now part of the IGI (Italian Grid Infrastructure) release and since 2008 has been included in the RESPECT Program.

The GRelC software has been exploited in several grid research projects to support bioinformatics experiments on distributed data banks as well as metadata management in the Earth Sciences and Environmental domains.

The GRelC service is currently adopted as the grid metadata management service in the Climate-G testbed to enable geographical data sharing, search and discovery activities. Moreover it is currently used at the Euro-Mediterranean Centre for Climate Change to manage climate metadata across the CMCC data grid infrastructure through the CMCC Data Distribution Centre portal.

Finally, a wide set of tutorials are available for end users to learn more about the GRelC Portal, the CLI and the JDK. Useful links can be found here:
https://grid.ct.infn.it/twiki/bin/view/GILDA/ GrelCProject.

6.4 Kepler

Kepler is a free and open source, scientific workflow application. Kepler is designed to help scientists and developers create, execute, and share models and analyses across a broad range of scientific and engineering disciplines. Kepler ships with a searchable library containing over 350 ready-to-use processing components ('actors') that can be easily customized, connected and then run from a desktop environment to perform an analysis, automate data management and integrate applications efficiently. Kepler workflows can be nested, allowing complex tasks to be composed from simpler components, and enabling workflow designers to build re-usable, modular sub-workflows that can be saved and used for many different applications.

EUFORIA (EU Fusion for ITER applications) project has extended the Kepler components repository with the actors enabling support for multiple grids / middleware stacks, including gLite (EGEE), UNICORE (access to HPC resources) and i2g (int.eu.grid). The extension covers the activities like: job submission, monitoring, data handling and interactive access to running jobs.

All actors are generic and could be reused by multiple other communities. There are also a number of different predefined use cases that could be easily customized and reused by different applications. As an example use case in terms of the Euforia project there are a number of mixed workflows running fusion codes, using grid (EGI/EUFORIA) and HPC (DEISA/EUFORIA) resources in the same time.

6.5 MOTEUR

Grid-enabled workflow designers and workflow enactment engines are application production environments delivering an abstract framework for the description of application logics and shielding the non-expert users from direct interaction with the grid environment. The MOTEUR workflow manager was developed to cope with the needs for describing and enacting data-intensive medical image analysis pipelines on the EGEE grid infrastructure. This tool is not specific to this application area though and may be reused in a broad category of applications covering the needs for many scientific areas using the grid for its High Throughput Computing capability.

MOTEUR workflows are described through a graphical editor. Their internal representation is a workflow-specific language based on array programming principles that makes it well adapted to represent parallel data flows and complex data manipulation constructs which are common in data-driven scientific applications. This language is particularly compact and it makes representation parallelism implicit and is well suited for non-expert users. In addition, it has more expressive power than the Scufl language underlying the Taverna workbench and is backward compatible with it. It can represent all forms of data parallel constructs such as parameter sweep studies, map-reduce operations and much more complex use cases.

The MOTEUR workflow enactor is an independent fully asynchronous gLite-interfaced engine capable of exploiting maximum data and service parallelism while exploiting grid resources to harness the computation load. It was designed for performance in a grid execution context. It can accommodate the heavy loads required for large-scale grid applications. Furthermore, it includes a command line application wrapper tool that makes easy the process of packaging and interfacing a command line interface-based tool to the gLite middleware.

MOTEUR is an open-source workflow engine available from:
http://modalis.i3s.unice.fr/softwares/moteur/start.

6.6 HammerCloud

HammerCloud is a distributed analysis stress testing system built around Ganga. It was motivated by a requirement from the ATLAS collaboration for site- and central-managers to easily test a set of grid sites with an arbitrarily large number of real analysis jobs. These tests are useful during site commissioning to validate and tune site configurations, and also during normal site operations to periodically benchmark the site performance. HammerCloud generates a test report including metrics such as the event processing rate, the mean CPU utilization, and timings related to various stages of the user analysis jobs. The report is presented in a web-interface that makes it simple to compare sites and observe trends over time. The system has been used by the ATLAS experiment to run greater than 200,000 CPU-days of test analyses. HammerCloud is implemented as a Django web application, with state maintained in a MySQL database and job management built around Ganga in Python. Jobs can be submitted to WLCG sites using the gLite UI and to all ATLAS sites using PanDA.

Prototype plugins for the CMS and LHCb experiments are in development.

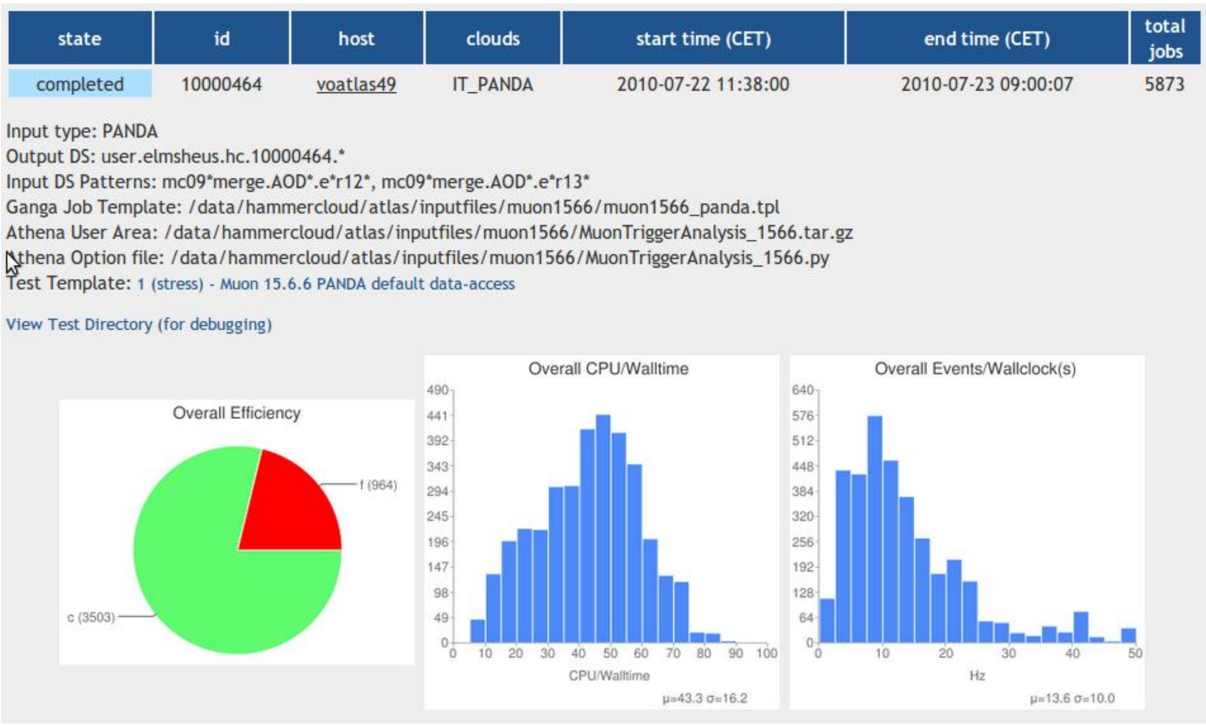


Figure 7 – The HammerCloud web interface to the results of distributed analysis stress tests

6.7 Data Management

6.7.1 DPM

The WLCG Disk Pool Manager (DPM) – based on the CERN CASTOR storage system – is a lightweight but fully functional Storage Resource Management (SRM)-enabled storage element (SE) that is currently used by the LHC VOs and some 60 others. As mentioned DPM offers SRMv2 as management interface, while files can be accessed both via the GridFTP (ideal for wide are network transfers) and RFIO (for local access) protocols. Both protocols support GSI authentication via X509 certificates. The DPM namespace offers flexible and configurable authentication and authorization capabilities, such as support for VOMS groups and roles, mappings to virtual IDs and ACLs at the directory and file level. Whilst DPM has been initially developed for storages on the order of 100TBs, we have examples of successful deployments at large sites at the PB scale, while one should remember its primary advantages as ease of deployment and management.

6.7.2 LFC

The LCG File Catalog (LFC) is the replacement for previously existing European Data Grid file (pre-EGEE) catalogues, providing more features and improving the performance and scalability. Sharing a common code base with the DPM, it also offers ease of deployment and management.

Some of the features it provides are the use of hierarchical namespace and namespace operations, built-in security and the possibility to use bulk methods to avoid long round-trip times. LFCs are used in over 60 grid sites and by some tens of VOs, including two of the LHC experiments, ATLAS and LHCb. The deployment model between these two experiments is somewhat different: ATLAS currently deploys LFCs at Tier1 sites to catalog data at the Tier1 and associated Tier2s, whereas LHCb deploy a global catalog with readonly replicas at the Tier0 and Tier1s. Synchronisation is maintained using the Oracle Streams product. To get an idea of the scales managed by LFC, in the case of ATLAS over 118 million files are registered between 11 LFC instances as of today.

6.7.3 FTS

The gLite File Transfer Service (FTS) is a point-to-point scheduler for data transfers over the wide area network. The grid Storage Element endpoints are paired in channels, resembling the network topology between the sites and ensuring balancing and throttling of data movement requests over the infrastructure. FTS channels can be configured in order to guarantee a minimal share of bandwidth to each Virtual Organization enabled in the channel. The FTS service can negotiate with the storage endpoints via the SRM protocol, while it utilizes gridFTP in third party mode for the actual byte streaming. A web service frontend allows the user to submit transfer jobs (consisting of many files), to poll their status and cancel jobs. The states of all transfers (both the current ones and the history) are stored in an oracle database, together with the monitoring information. In WLCG, the deployment of FTS is distributed across many sites (the Tier0 and the Tier1s). The FTS server at CERN regulates the traffic from CERN to all Tier1s (and vice-versa), while each FTS server at a given Tier1 regulates the incoming traffic to that Tier1 (apart from the one coming from the Tier0) and to all Tier2s associated to the Tier1.

6.7.4 Future Data Management Developments

Following first experience with real LHC data taking, production and analysis it has become clear that some of the assumptions behind the current model are not strictly valid. For example, the “pre-placement” of data from Tier1 sites to Tier2s for analysis has shown that a large fraction of the data is never accessed. Alternative models, such as dynamic caching of data and / or remote access, have recently been proposed for evaluation. As with most developments in this area, such strategies could be equally beneficial to other communities and could result in better resource utilization (potentially lower network traffic as only the needed data is transferred and improved storage management through smaller numbers of data copies). Other issues currently being studied include the potential use of industry-standard components for storage management and / or data access – again of likely interest to all communities.

6.8 Capabilities of the WLCG Service

The WLCG service can be considered the "high end" of grid computing. In terms of number of unique users, number of jobs per day, data volumes and access rates, number of cores provisioned, volume and growth of data it well exceeds any other community in its use of – as well as commitment to – grid computing. WLCG chose to align itself as closely as possible with EGEE from the very early days and this commitment continues with EGI with the use of key operations and support tools, together with a middleware stack that is primarily based on gLite (WLCG sites also exist in the Nordic region, running ARC plus some gLite components such as the LFC and FTS, and the US – based on OSG but again with selected gLite components).

Although subject to a Memorandum of Understanding, WLCG works primarily as a collaborative effort (between sites, service providers and the experiments) – much in the way that the experiments themselves do. WLCG has developed a small number of operational procedures, some of which have already been adopted by EGEE (and hence by default also EGI). Other procedures are fully generic and can be easily be adopted by different communities, as and when their production needs grow. This includes strategies for deploying and running robust and resilient services, as well as policies for dealing with prolonged site or infrastructure (e.g. GGUS) outages.

Large new communities may wish to adopt some of these strategies from the beginning – all are extensively documented and further details can be obtained from the MS601 contact points.

The goal of the WLCG project is to enable the exploitation of the physics potential of the LHC machine itself. This has born fruit, as witnessed at the 2010 International Conference on High Energy Physics (ICHEP), as witnessed by the following slides from the ATLAS and CMS status reports.

In particular, Figure 8 shows the daily average throughput rate (in MB/s) for worldwide ATLAS data transfers over the last seven months.

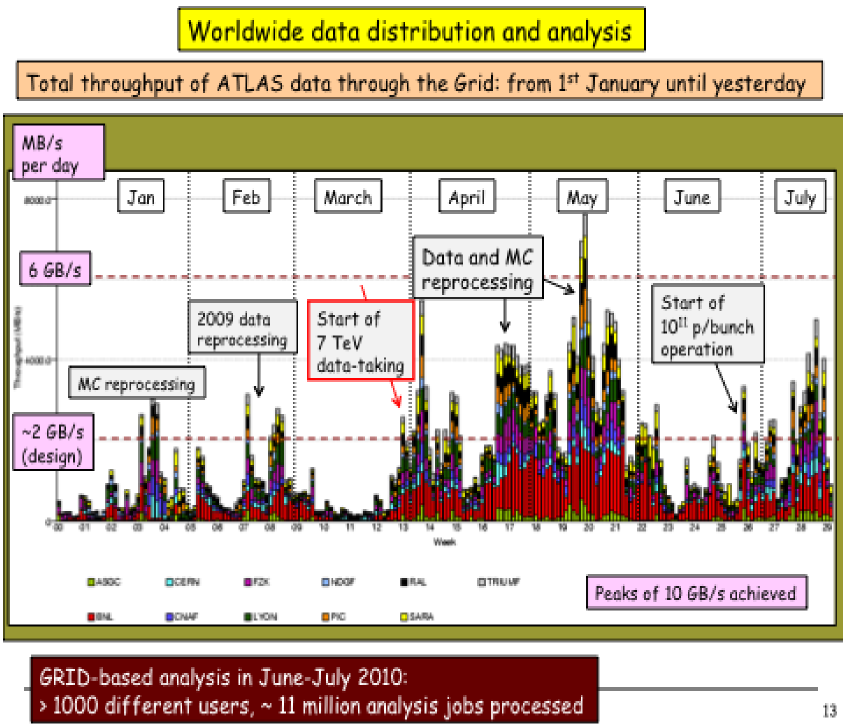


Figure 8 – ATLAS Grid status at ICHEP 2010

One can notice that while the mean throughput is consistent with the design value of 2GB/s, there are large fluctuations in order to deliver promptly newly reprocessed data to the physicists. Such fluctuations, peaking at 10GB/s are properly absorbed by the system that on no occasion has been seen to suffer of high load. At the same time, the infrastructure demonstrated to be able to cope with multiple experiment activities at the same time: in Figure 9 one can see how an increase of ATLAS activity (in the period from April to June) largely above the design model has no implication on the CMS data transfer capability with no evidence of saturation, up to the point where more than 700TB have been dispatched worldwide in a single day by the two experiments.

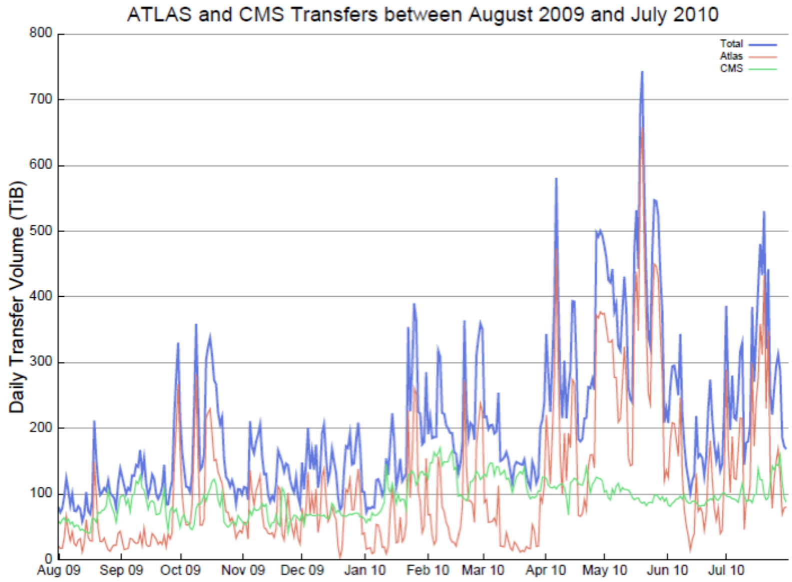


Figure 9 – Daily transfer volume of ATlAS and CMS virtual organizations

Figure 10 shows the increase of CMS analysis jobs over the last many months, in conjunction with events like important conferences.

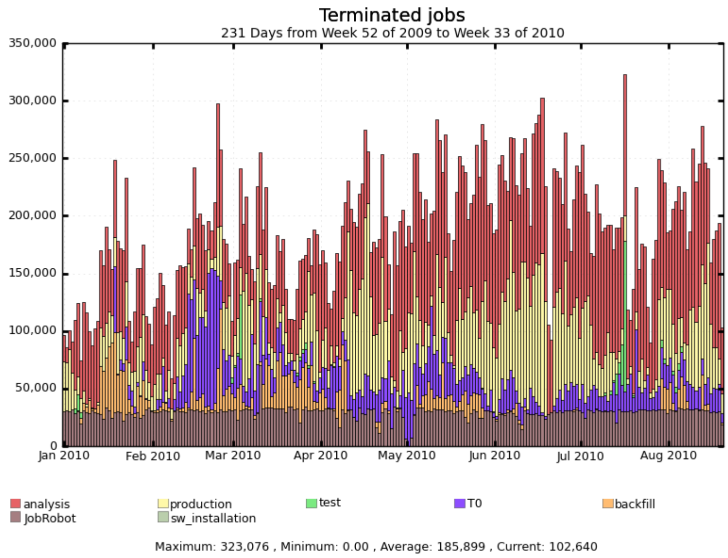


Figure 10 – Number of jobs processed worldwide per day by the CMS VO

Again, while routinely the CMS system runs the order of about 200K jobs per day, it is capable to absorb peaks of activity up to 1.5 times larger than the steady rate. The load comes of course both from scheduled productions and single user analysis. In Figure 11 one should notice how the number of single users submitting jobs to the grid in CMS has exceeded 500 within the same week, while the same number was approximately 100 before the beginning of data taking in 2010.

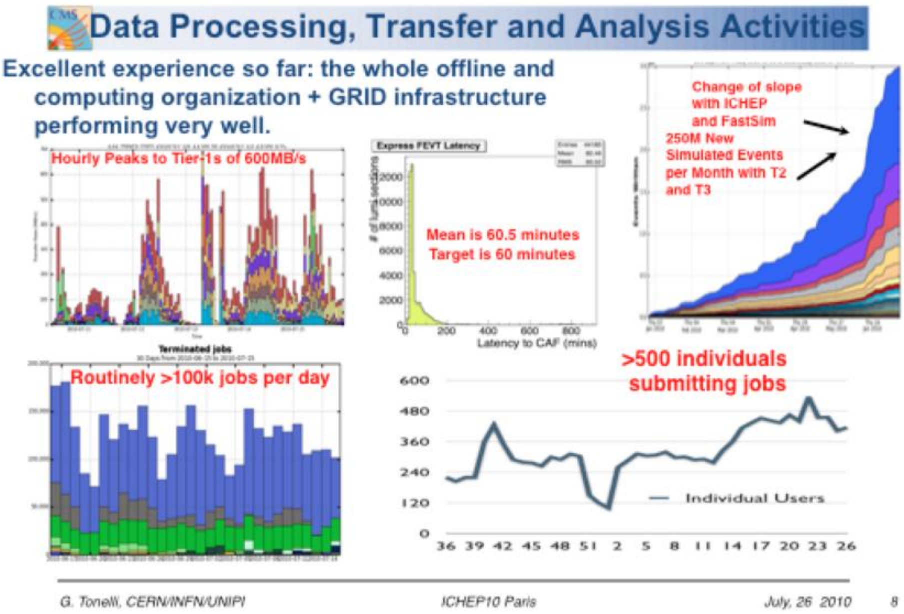


Figure 11 – CMS Grid status at ICHEP 2010

The increase of analysis activity is even more visible in Figure 12(right), showing the number of analysis ATLAS jobs running simultaneously over the WLCG infrastructure on a daily basis. While this number progressively increases over the last 7 months, one can easily see a burst of activity in the middle of March after the first 7TeV LHC run, in May after the major Monte Carlo and Detector Data reprocessing campaign and in July, prior to ICHEP 2010. To this, one should add the continuous (and more regular load) of the production activity in Figure 12(left), adding up to a maximum of more that 60K jobs running concurrently over the grid computing resources

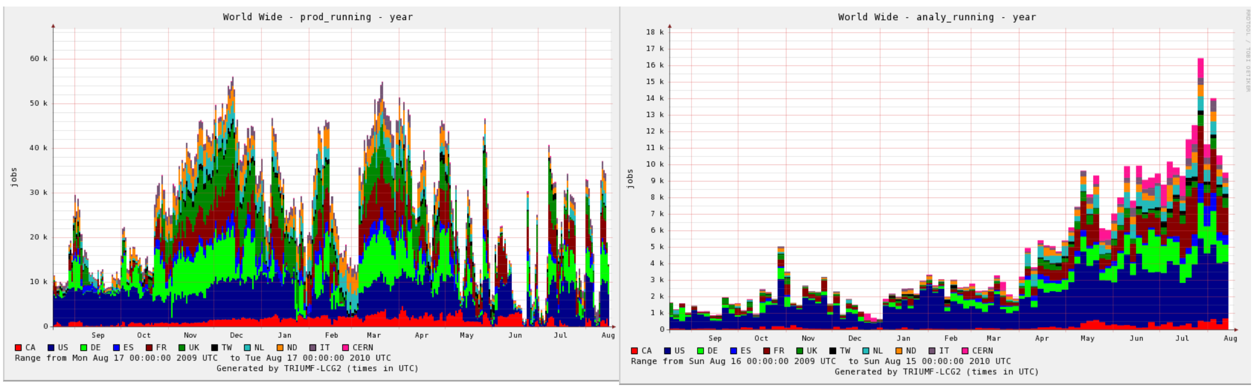


Figure 12 – Number of ATLAS production (left) and analysis (right) jobs running in parallel

The screenshot of the WLCG Google Earth display in Figure 13 shows a runtime picture of the LHC jobs processing and data transfer activities on the worldwide distributed infrastructure.

Figure 13 – WLCG Google Earth display shows the LHC computing activity on the distributed infrastructure

7 CONCLUSIONS

Through a series of case studies we have described a variety of application domains and the concrete steps required to use grid services and resources from first steps to petascale production. Grid computing is not a panacea: for suitable computing models – such as those that can be readily decomposed into a set of parallel tasks – it offers a number of well documented advantages over alternative solutions. In particular, it is significantly cheaper in terms of hardware and support costs than super-computer based solutions – especially in the case of long-lived projects where the application development cycle is quasi-continuous. It also allows funding to be invested locally, potentially breathing life into local economies and societies via a vibrant local research and education infrastructure, something that would be less obvious (at least in the former case) than current commercial cloud-based solutions. Most importantly, it has been proven to work. Until alternative models have been demonstrated on the production scale required by a project such as the LHC, this is an undeniable advantage.

8 REFERENCES

R 1	EGI-InSPIRE Description of Work (DOW)
R 2	MS601 – HUC Contact points and the support model: https://documents.egi.eu/document/91
R 3	MS602 – HUC Software Roadmap: https://documents.egi.eu/document/230
R 4	MS603 – Services for High Energy Physics: https://documents.egi.eu/document/160
R 5	MS604 – Services for the Life Science Community: https://documents.egi.eu/document/236
R 6	EnviroGRIDS project home page: http://www.envirogrids.net/
R 7	PARTNER project home page: http://partner.web.cern.ch/
R 8	ULICE project home page: http://ulice.web.cern.ch/
R 9	Linear collider detector home page: http://lcd.web.cern.ch/lcd/
R 10	Ganga: a tool for computational-task management and easy access to Grid resources; *Computer Physics Communications, Volume 180, Issue 11, (2009)*
R 11	Lattice QCD Thermodynamics on the Grid; *Computer Physics Communications 181 (2010), pp. 1715-1726*
R 12	Grid-Enabled High-Throughput Screening Against Influenza A Neuraminidase; *IEEE Trans. Nanobioscience, Volume 5 (2006)*
R 13	Dynamic partitioning of GATE Monte-Carlo simulations on EGEE; *J. Grid Computing, Volume 8 (2):241-259, (2010)*
R 14	Scheduling for responsive Grids; *J. Grid Computing, Volume 6 (15-27), 2008.*