Applications & Community Support: Transition Plan

1. Introduction

1. The usage of Grids continues to grow, both in terms of active communities as well as in resource utilization: the current European Grid infrastructure EGEE consists of more than 270 sites distributed in 54 countries. More than 11400 CPUs and 20PB of disk space serve the research purposes of more than 15 communities. Today there are more than 200 VOs and about 16000 active users;
2. The HEP community, which has been running in continuous production mode for several years and is the leading user of the European Grid infrastructure, recently carried out a Scale Testing for the Experiment Programme'09 (STEP'09) exercise. As part of this programme the ATLAS experiment alone ran close to 1 million analysis jobs and generated 6 GB/s of sustained Grid traffic. The exercise was prepared with the experiments during a period of about two months and executed in two intensive weeks of 24x7 operations;
3. The EGEE Grid (as well as additional Grid resources used by WLCG) has already achieved a high level of scalability and reliability in terms of services, software and operations, with service targets for sites being regular met or even exceeded (99% service availability being required for the most critical Tier0 and Tier1 services);
4. A stable and reliable infrastructure makes the Grid an attractive environment for both research and business activities and will help attract new communities: the Grid support team at CERN concurrently supports some 5 additional communities including simulation toolkits, United Nations Initiatives, theoretical physics studies, etc;
5. Based on this infrastructure, private initiatives, businesses and United Nations initiatives have shown their interest in Grid technologies since several years;
6. After three successful phases since 2001, the EGEE project will finish in May 2010. The next Grid project EGI foresees significant changes to the current Grid environment that are described elsewhere in this [transition] document. The basic requirement of all communities is to ensure the stability of the current Grid infrastructure during and after the transition. This issue also affects the service level and operations procedures;
7. In this document we will address the most important aspects of a smooth transition from the point of view of HEP and related communities. The operations procedures and the required level of the services will be also detailed;

2. Areas of Key Impact

8. The expected results of this work are a marked increase in the number of Grid users, as usage expands from the data processing activities that have dominated until now into the realm of data analysis, scientific discovery and

publication. This will be accompanied by wider inter-disciplinary collaboration, both through science (i.e. related disciplines) and technology (e.g. grid tools);

9. This can only be achieved by a significant simplification of user interaction with the Grid, through further adoption of existing tools such as those described in detail below, and by a flexible and scalable end-user support model;
10. This includes the establishment of community support, whereby the communities are encouraged and enabled to be largely self-supporting, with expert guidance to establish and optimize the support structures and associated tools. This is essential not only to deal with the large expansion in terms of number of users but also for long-term sustainability;
11. These activities will help to ensure Europe's leadership role in the areas of grid design, deployment and efficient exploitation.

3. HEP SSC: Motivation and Plans

12. High Energy Physics (HEP) is the leading user of the EGEE infrastructure (80% during the past year), as well other major Grids [references];
13. Along with production usage throughout the year, two specific programmes aimed at demonstrating readiness at the scale required for data taking, processing and analysis at the Large Hadron Collider (LHC) at CERN have taken Grid computing to new heights in terms of service availability, number of jobs per day, number of unique users, inter-site data transfer rates, as well as functionality and usability;
14. As a direct result of this work, a variety of other disciplines – linked either by the science that they perform, or else via the (Grid) technology that they use – have benefited measurably;
15. These activities have been supported during the past years through co-funding from a variety of sources, including national Grid projects and funding agencies, international projects supported through the EU and other agencies, as well as specific institutes such as CERN, which have allowed a relatively small "Grid support" team to empower a number of (very) large distributed communities;
16. To fully realize the value of this investment, this "Grid support" function must continue and it is foreseen that this be done via a "Specialised Support Centre" (SSC) for these communities;
17. We describe below the key functions and benefits of this SSC, together with a concrete transition plan and timeline;
18. The impact will include closer collaboration between scientific disciplines – HEP, astro-particle physics, fusion, light source or "photon" science – as well as numerous international communities, using user-friendly high-level interfaces to underlying Grid middleware services that simplify significantly the use of powerful e-Science infrastructures developed for the above domains;

19. Although the intent is to continue to support a variety of disciplines, we will refer to this activity as "the HEP SSC" in the remainder of this document.

4. HEP Transition Plan

20. One of the key users of the services from the SSC that this section describes will clearly be the LHC experiments and the Worldwide LHC Computing Grid (WLCG) project;
21. Given the schedule for the restart of the LHC in Q3 2009, with a data taking run continuing more or less uninterrupted until late 2010, the transition must start prior to any FP7 and/or related funding being available;
22. This has been commenced through the standard CERN fellowship programme, through which two post-doctoral short-term positions have been opened as of summer 2009. However these positions only partially compensate for staff loss through end of contract / (external) funding and other departures. Further significant reductions will occur naturally as EGEE III draws to a close;
23. These reductions and indeed any further loss of effort represent a major risk not only to the communities supported but also to the use and adoption of grids in general. It is felt essential not only for these reasons but also to retain Europe's leading role in e-Infrastructures that this concern be addressed with priority;
24. Notwithstanding the above, the existing effort is being refocused on the needs of the imminent startup of the LHC, including increased support for data analysis, where "community support" – whereby the large distributed community is trained and enabled to be largely self-supporting, backed up by a team of experts – is proving to be an attractive model;
25. For the purpose of this document we describe the proposed activities in terms of "work packages" that focus on specific communities, as well as "horizontal activities", including training, dissemination and common tools;
26. Although expected to be (largely) performed outside the strict context of an SSC, we also enumerate our requirements in terms of middleware (release and service issues) as well as operations and user support.

4.1. Operations and User Support

The WLCG requirements in terms of Operations and User Support are largely covered in the document "WLCG Operations for LHC Data Taking". We reproduce here the conclusions:

27. WLCG Operations has reached stability and sustainability. For relatively low but non-zero cost it can be maintained and enhanced;
28. Manpower to support the key operations tools and to fill needed operations roles – both WLCG and experiment-specific – is required. It is expected that this be jointly funded to allow the successful exploitation of the world-class grid that has been built up over many years, together with international partners that make this a truly global enterprise;
29. We believe that WLCG operations experience and procedures can have significant value to other communities and are keen to share this knowledge.

4.2. Middleware Requirements

The primary requirements are related to the release and updates of middleware versions and Grid service requirements. Given the move from deployment to production, the primary requirement is for middleware that is designed with robust service deployment in mind: this requirement must be taken into account from the early design stage and must be reflected in issues such as consistency and clarify of error messages and logging, the provision of the necessary hooks for monitoring and eventual debugging, as well as design for robust deployment (scalability and failover). In addition to these basic requirements, the schedule of large communities such as WLCG needs to be considered. In this specific case they are driven by that of the LHC machine, which will typically operate from Spring until Autumn annually. During this period, major middleware or service changes cannot realistically be deployed in production (whereas minor updates and / or bug fixes can). In an ideal situation, larger changes would be made available such that they can be fully tested and debugged to allow production deployment several months prior to the annual accelerator startup. Changes which miss this time window are unlikely to be deployed prior to the following shutdown period.

Although the schedule is expected to vary with discipline and associated scientific machine, similar constraints can be foreseen for other large communities.

4.3. Europe’s Leadership in grid

Europe’s leadership not only of the development and deployment of grid systems but more importantly in terms of their successful utilization for enabling science and other disciplines is well known and attested to by the numbers listed above. This leadership translates into a competitive advantage both for these communities as well as the clear potential for direct and in-direct spin-off.

5. Use case 1: High Energy Physics (HEP)

HEP is the largest community of the EGEE infrastructure, both in terms of number of users and resource utilisation. HEP also has correspondingly large requirements in the number of concurrent and completed jobs per day, I/O and bulk data transfer rates, custodial storage and so forth.

It is the community that often performs (pre-) production testing of new services and / or releases – benefiting not only HEP but also many other Grid communities.

The complex computing models of the LHC experiments have implications on the operations procedures and Grid support infrastructures. The policies and procedures established for the HEP communities – e.g. within WLCG – have frequently been adopted as part of the standard EGEE procedures.

HEP can therefore be considered as a reference for other scientific communities. Collaborative activities between HEP and other communities have been supported during the past years through co-funding from a variety of resources, including national Grid projects and funding agencies, international projects supported through the EU and other agencies, as well as specific institutes such as CERN, which have allowed a relatively small “Grid support” team to empower a number of (very) large distributed communities.

A smooth transition from EGEE to EGI is essential for the HEP community. A transition that satisfies this community can also be successfully applied to others. Given the schedule for the restart of the LHC in Q3 2009, with the data taking run continuing more or less uninterrupted until late 2010, the transition must start prior to any FP7 and/or related funding being available.

Reasonably detailed programme of work to be agreed with community, e.g. LHC experiments and others.

Impact

This will allow the HEP community to successfully exploit the scientific potential of the world's largest scientific machine, potentially leading to new levels of understanding of the Universe around us, in particular in areas not fully or adequately explained by the so-called "Standard Model".

Effort

8 FTE support personnel for the 4 main LHC experiments (2 FTE per experiment).

Technical consultants for liaison with operations and middleware (2 x 0.5 FTE)

Expected VO-specific operations effort provided by the T0, T1 and other sites to cover the LHC experiments' needs (estimated at 1 FTE per VO per site).

Deliverables:

PM12: to be agreed with community

PM24: as above

PM36: as above

5.1. Analysis Support for HEP

Reasonably detailed programme of work to be agreed with community, e.g. LHC experiments and others.

Impact

Effort

Deliverables:

PM12: to be agreed with community

PM24: as above

PM36: as above

5.2. Dashboard

The Experiment Dashboard [1] monitoring system was developed in the framework of the EGEE NA4/HEP activity. The goal of the project is to provide transparent monitoring of the computing activities of the LHC VOs across several middleware platforms: gLite, OSG, ARC.

Status:

Currently the Experiment Dashboard covers the full range of the LHC computing activities: job processing, data transfer and site commissioning. It is used by all 4 LHC experiments, in particular by two largest, namely ATLAS and CMS. Generic functionality, such as job monitoring, is provided by Dashboard server to all VOs which submit jobs via gLite WMS.

The Experiment Dashboard provides monitoring to various categories of users:

- Computing teams of the LHC VOs
- VO and WLCG management
- Site administrators and VO support at the sites
- Physicists running their analysis tasks on the EGEE infrastructure

To allow for the development and building of components of Dashboard Monitoring applications, a Dashboard framework was designed. This is currently used by other projects and development teams and not exclusively for the development of the monitoring tools. The Dashboard framework is used as well for the construction of the high level monitoring system which provides a global view of LHC computing activities across all LHC experiments, both at the level of the distributed infrastructure in general as well as on the scope of a single site.

Web monitoring shows heavy use of the Dashboard servers, for example the dashboard of the CMS VO serves 2300-2500 unique visitors per month with about 30K pages accessed daily. These numbers are growing steadily.

Future Evolution:

The future evolution of the project is driven by the requirements of the LHC community which is preparing for LHC data taking at the end of 2009.

The main strategy is to concentrate effort on common applications which are shared by multiple LHC VOs but can also be used outside the LHC and HEP scope.

Impact:

Reliable monitoring is a necessary condition for production quality of the distributed infrastructure. Monitoring of the computing activities of the main communities using this infrastructure in addition provides the best estimation of its reliability and performance.

The importance of flexible monitoring tools focusing on the applications has been demonstrated to be essential not only for "power-users" but also for single users.

For the power users (such as managers of key activities like large simulation campaigns in HEP or drug searches in BioMed) a very important feature is to be able to monitor the resource behaviour to detect the origin of failures and optimise their system. They also benefit from the possibility to "measure" efficiency and evaluate the quality of service provided by the infrastructure. Single users are typically scientists using the Grid for analysis data, verifying hypothesis on data sets they could not have available on other computing platform. In this case the monitor is a guide to understand the progress of their activity, identify and solve problems connected to their application.

This is essential to allow efficient user support by "empowering the users" 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).

Effort

100% FTE

Deliverables

PM12: Status report

PM24: Status report

PM36: Status report

6. Use case 2: Collaboration with International Entities

In EGEE we observed that the best way to promote Grid computing was to identify established communities with immediate requirements. Building collaboration by successfully attacking a concrete problem has been shown to be an excellent way to get visibility and to promote broader usage of Grid technology and of the EGEE infrastructure in particular. Target communities of interest include those already operating at a large scale but not yet using large distributed computing infrastructures as a permanent component of their normal operational model, such as large scientific or technical communities and / or enterprises. The interest is that one can effectively reach large scientific and technical communities.

In the recent past CERN established a collaboration with ITU (International Telecommunication Union) during the 2006 Radio Communication Conference. The Grid was used to complement the ITU computing system to perform heavy optimisation studies (200,000 jobs per optimisation round to be executed in less than 12 hours). The results became eventually part of an international treaty. The usage of the Grid was presented during the same conference in a plenary meeting (185 participating countries, delegations with technical and diplomatic staff).

More recently CERN helped the launch of an FP7 project lead by UNEP (EnviroGrids: Black sea catchment studies – 27 partners most of them from the Black Sea coastline countries). The contribution of CERN (Grid support) is relatively small (X% of the project budget) but makes possible an essential part of the project's programme of work (systematic study of the hydrological model and data comparison). This is a quite effective way to ramp up such an ambitious project, allowing the project management to concentrate in building the collaboration and the participating scientists to share their diverse competences to face the project scientific challenges.

The goal of formalising the existing activities is to improve and stabilize contacts with international organisations such as WMO, WHO, ITU. Of particular interest would be to establish contacts with EU agencies like the European Centre for Disease Prevention and Control (ECDC), the European Chemicals Agency (ECHA), the European Environment Agency (EEA) and possibly others.

These large entities are essential gateways to the corresponding scientific communities. Several of these institutions have privileged contacts with strategic economical players (e.g. ITU and the world of the Telecoms, WHO and the pharmaceutical industry, etc...). In this fact resides the strategic interest of this activity. In addition these organisations can reach communities (especially small or developing countries) which are often otherwise excluded from innovation by digital and social divides.

This activity should be lead by CERN because of its established role in attracting and collaborating with other activities (scientific and technical) in Grid technologies. Equally important is the experience matured by CERN as centre of excellence in computing for the HEP community (hence one of the largest and probably the most organised international research community).

Effort:

Activity leadership (CERN): 50% of a FTE

Technical consultant (CERN) 100% of a FTE

50 kEuro travel budget

50 kEuro the organisation of events

Deliverables:

PM6: Establish a cross-fertilisation plan with one of the large partners (to be used as a template for the others)

PM24: Report on the collaboration with at least 2 partners

PM36: Final report of the collaboration including prototype activity and a plan to maintain the established collaborations

7. Use case 3: Photon Science

The community of users in photon sciences has many similarities to HEP-users which makes it worth to use the synergies. But as well there are some major differences. The user community itself is much more heterogeneous and thus the applications and the basic ICT-knowledge varies a lot. This requires more effort in training, operation and administration of grid resources in comparison to the HEP community. Furthermore the typical length of an experiment last in contrast to the large HEP-Experiments only day's or weeks and the experiment groups are much smaller. This leads to a different operation model which can be seen in a much more "burst mode" fashion. Some of the experiments do have large data volumes, which occur during an experiment. Those data have to be made accessible through grid methods, where again in contrast to HEP security issues can play an important role.

In the photon community there are already many ongoing grid activities, in particular at large Labs like DESY and in the EGEE context. With the ESFRI-Projects EuroFEL, XFEL – both at DESY and the ESRF-upgrade, major new facilities are set up, which will give a boost to this research field. Therefore a distributed computing infrastructure is needed, which rely on the tremendous experience and development, achieved in EGEE and in the HEP community. Nevertheless this has been adapted to the needs of photon science. In particular, the proposed tasks are in continuation and broadening of the current EGEE activities:

1) User administration and access through Grid methods
2) Large grid enabled Data storage
3) Data access and distribution through Grid methods
4) Operation of Grid resources with common EGEE middleware

5) Training and education of Users
6) Analysis in Grid environments

Effort

The photon community will propose a SSC, which will be driven by 4 major "light sources" in Europe, which are as well involved in the relevant ESFRI-Projects.

These Centres will be:

- DESY with the ESFRI-Projects EuroFel and XFEL
- PSI also with the ESFRI-Projects EuroFel
- Elettra (tbc)
- ESRF with the ESFRI-Project ESRF-upgrade (tbc)

It is expected to have 2 persons at each site, one unfunded and one funded by this project.

For the placement of this effort and according to the proposed scope of the project, each partner has the following role:

- DESY: grid operation, storage elements, Analysis
- PSI: Operations, user interface, storage elements
- Elettra: middleware, training, operation
- ESRF: operations, user interface, training

Impact (tbc)
Deliverables (tbc)

DESY and HEP

Grid technology is seen as a major compute paradigm for the LHC data analysis. Nevertheless Grid technology is heavily in use in other HEP experiments as well like in the data analysis of the HERA experiments, in the detector design for ILC, the theory and in astroparticle physics like Icecube. For all those experiments, DESY is hosting a full set of Grid Services based upon EGEE middleware components.

The DESY role in a HEP SSC is, to

1. operate a basic grid infrastructure for non LHC communities from HEP
2. to further develop the grid enabled storage element dCache
3. to support the scientific users in deploying the grid technology.

Effort

DESY requires 2 FTE funded by the EU and will add 2 FTE from the own staff.

8. Common Tools

The HEP support team has invested significant effort in the creation of tools that ease access to and the management of Grid infrastructures. These tools have been adopted by a large number of communities due their high level of abstraction and the stability. Access to the Grid, optimal resource usage, monitoring of jobs etc. are managed by this set of tools for which continued support is clearly required.

8.1. Ganga

EGEE supported the Ganga project via the EGEE NA4 HEP activity (the application sector of EGEE as part of the effort for High-Energy Physics).

The goal was to provide a flexible and user-friendly tool to enable large user communities and move away from a niche usage of the Grid.

As today we observe several hundred Grid users using Ganga in their daily activity, a number in steady rise over the last years. Interestingly enough several users (not belonging to the HEP community, sometimes even from outside EGEE) “discovered” Ganga and started using them (often in a complete autonomous way using the material available on the web) as shown in Fig.xxx. Note that out of the ~500 users per month about 25% of the users are non-HEP users.

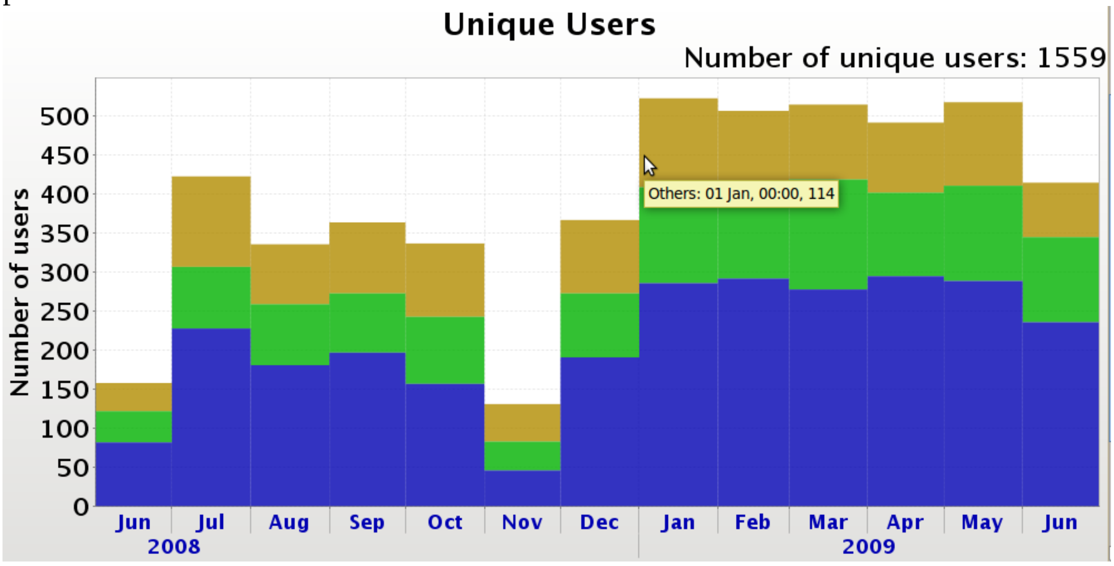

Fig.x: Time evolution of Ganga users per month. Every month, unique users from ATLAS (blue), LHCb (green) are counted alongside with users from other virtual organisation (light brown). The dip in Nov 2008 is due to a monitor failure.

The Ganga system is described in a recent paper [2]. For this document is sufficient to signal its non-intrusive nature. “Non-invasive” means that the application should not be (ideally at all) changed to run on the Grid. Users can seamlessly combine the usage of grid and other resources. Ganga is a natural tool to run on multiple Grids (interoperability) and to protect users from changes in the middleware layer (new APIs).

Ganga does support access to major grid infrastructure (notably EGEE, OSG and NorduGrid) and to all major batch systems.

Ganga is the main tool behind a large and very successful activity of application porting (“gridification”) performed by CERN. More than 20 applications have been ported ranging from Fusion, Material Sciences, Accelerator Studies, Biomedical applications such as dosimetry in oncology and drug searches, large-scale software testing and many others.

Status:

The project is in "maintenance mode" meaning that the main activities are keeping up the natural evolution: for example new execution back-ends, bug fixes, new version of Python etc...).

Impact:

Ganga is one of the most useful elements in our toolbox to support large user communities, enable new ones and collaborate with other infrastructure projects. The return of investment is guaranteed by the large user communities and the existing and future activities in the area of support and outreach of new communities.

Effort:

Engineer: 100% FTE

Deliverables:

PM12: Status report

PM24: Status report

PM36: Final report

8.2. DIANE

DIANE is a tool supported by NA4/EGEE project and listed in the RESPECT program [3].

The tool improves the efficiency of using distributed resources, especially Grid, by providing private, user- or application-specific overlays for handling workloads in Master/Worker paradigm (aka pilots, agents, glide-ins). DIANE may be used to easily create robust production-like systems running autonomously over long periods of time with no human attention required as demonstrated during the LQCD production in 2008, where 20TB of simulation data was produced using on average 750 CPUs (1400CPUs peaks) over a period of 1 year. High granularity splitting of user workloads allows to reduce average execution time and the variance of a set of jobs, as demonstrated during ITU RRC06 activity [4], where 200K short jobs, representing several hundreds CPU-hours where completed within few hours deadline. During the Avian Flu Drug Search in 2006, the efficiency of DIANE-based system was 84% as compared to 46% efficiency using classic Grid job submission in the WISDOM system [5]. DIANE is also routinely used for bi-yearly Geant4 regression testing, where 35K jobs executing in a period of two weeks and corresponding to 4 CPU years, are used for statistical comparison of the implementation of physics simulation algorithms in Geant4.

DIANE is used in conjunction with Ganga [6] to provide an advanced grid-access and workload-management solution for smaller experiments and user communities with zero deployment and zero customization of the infrastructure required, allowing a seamless combination of Grid and local resources. The DIANE/Ganga system is used directly by individuals in several application domains such as medical image analysis including irradiation simulation, MRI cardio-vascular analysis, astrophysics simulations, fusion reactor applications, high-energy physics experiments (COMPASS and Atlas for Athena-based analysis on local clusters). The improvement in workload execution offered by DIANE is perceived by the users as an improvement in the Quality of Service when using the Grid.

DIANE/Ganga system is used as a building block for community-specific services such as Grid portals, as demonstrated with ASGC bioinformatics Grid Portal for genomics and drug design.

DIANE/Ganga also provides an attractive R&D platform (easily pluggable schedulers and parameterization of workload execution) which attracts researchers in various communities, including middleware solutions (e.g. SAGA) and application models (workflow management with DAGMan and Vbrowser/Moteur)

Status:

The project is mainly in maintenance mode: bug fixes, porting to emerging platforms, integration of components contributed by user communities, occasional new features and extensions, updates of documentation etc.

Effort

An effort (approx. 1FTE) is required to maintain the system in the current state and to provide expertise on application porting.

8.3. AMGA

Several applications depend on the Grid as the environment to access and process large and complex data sets. Efficient metadata handling is at the heart of such system. Bookkeeping, annotation, sharing of information about data quality and processing status are essential in any data analysis: on the Grid the importance and the requirements in terms of dependability and performance are vital.

AMGA [7] is a general purpose metadata catalogue and part of the gLite Middleware distribution. It was developed within the HEP support activity in the EGEE project, when it became clear that many Grid applications needed metadata information on files and to organize a work-flow [8]. AMGA is now developed and supported by a collaboration of CERN, INFN in Catania and KISTI in Korea.

AMGA as a metadata services allows users to attach metadata information to files stored on the Grid, where metadata can be any relationally organized data typically stored in a relational database system. In addition, the metadata in AMGA can also be stored independently of any associated files, which allows AMGA to be used as a general access tool to relational databases on the Grid. AMGA features a simple to learn metadata access language in addition to SQL, which has been very useful for the adoption of AMGA in smaller Grid applications, as it lowers the technical hurdle to make use of relational data considerably.

One of the main features of AMGA and unique to it, is the possibility to replicate metadata between different AMGA instances [9] allowing the federation of metadata (e.g. by the Health-e-child[1] project), but also to increase the scalability and improve the access times on a globally deployed Grid (as done by the Wisdom project). In addition to performance, security was another major concern during the design and development of AMGA [10] and AMGA features different authentication and access control methods. These advanced security features have made AMGA the de-factor standard for metadata and relational database access on the Grid for biomedical

[1] An integrated platform for European paediatrics based on a Grid-enabled network of leading clinical centres, http://www.health-e-child.org/

applications (e.g. [11]), within and outside of the EGEE project. Prominent projects making use of AMGA in this field are the Wisdom[2] , and the Health-e-Child projects.

The easy usage and introduction of AMGA in Grid-enabled projects have made AMGA important components in many efforts to Grid-enable applications, like the Unosat project[3] or the SRM-SRB[4] Grid storage project used for digital libraries.

In spring 2009, AMGA featured also the first implementation of the proposed WS-DAIR standard proposed by the OGF for relational data access on the Grid. A thriving collaboration between the EGEE community and the OGF community entailed, aimed at verifying the standards proposal and evolving it towards future needs.

Status:

The project is effectively in maintenance mode, where the effort for fixing bugs and code maintenance is provided by KISTI as part of their contribution to the gLite Consortium.

Impact:

In addition to the code maintenance it will be necessary to provide support for the existing user communities and continue to in supporting the adaptation of new applications to the Grid, both SuperB and the SuperBelle B-Factory experiments have voiced there interest in AMGA. In addition, the requirements of these users need to be represented in the standardization efforts. In addition we advocate the importance to devote a small yet visible effort which should be present as back-office effort for allowing the usage of AMGA in the above-mentioned activities (training, interoperability and gridification).

Effort:

Engineer: 100% FTE

Deliverables:

PM12: Status report

PM24: Status report

PM36: Final report

9. Training and Dissemination

Training and dissemination are activities that will continue in the context of the current related activities. In this document we mention some examples of training events performed and managed by HEP members

Training:

Over last few years an ecosystem of user communities and applications have appeared around Ganga and DIANE tools, and the dissemination and training have played an

[2] An initiative for grid-enabled drug discovery against neglected and emergent diseases http://wisdom.eu-egee.fr/

[3] A UN-CERN collaboration to give field-workers in disaster areas access to satellite imagery and their treatment on the Grid: http://unosat.web.cern.ch/unosat/

[4] http://www.beliefproject.org/news/srm-srb-interface-a-new-milestone-for-grid-interoperation

increasing role in our activities. The long-term vision for training is to off-load the developers and core team at CERN by creating a situation in which user communities may provide a self-sustaining user-support and take, partially, a role of the driving force for the tools. We receive notoriously good feedback for the quality of our tutorials, which provide not only practical introduction to the tools, but also inspiration and more insight to the students than other, often more formal training proposed in EGEE. Here is a list of major training events (outside of LHC context) in last 12 months:

- *Train the trainer*, CERN, 2008
- *How to use the Grid for physics and medical applications*, IEEE NSS 2008, Dresden, Germany
- *EGEE tutorial: Grid introduction*, CSC Helsinki, Finland, 2009 (invited)
- *Application Grid Enabling*, Baltic Grid All Hands Meeting, Riga, Latvia, 2009 (invited)
- *Best practices in putting application on the grid***,** Baltic Grid Summer School, ITPA VU Astronomical Observatory, Lithuania, 2009

The tutorials are organized via different channels, including the contact to the EGEE training group, which also picked up (partially) our training program for generic EGEE tutorials organized by them.

The key of the success is that the training activity involves Grid experts with day-to-day experience in supporting user communities and tools. The expert involvement may be direct or via Train-the-Trainer concept such as the case of BG Summer School tutorial.

<u>Dissemination:</u>

Through EGI actions as:

User Forums

- International events (at least once per year) to bring together Grid users
- Exchange of experiences and techniques
- Scientific program coordinated by EGI which also need to ensure a high qualified technical presence to confront problems from different communities

Topical Workgroups

- Users facing similar problems while using the Grid infrastructure
- Meeting prepared to devise solutions

Discipline Meetings

- Associated to specific disciplines to discuss common solutions related to the use of middleware, techniques, data resources.

10. Liaison with other projects / workpackages

10.1. Summary of support team tasks

- Liaison between the community and services-operations and middleware teams

  - The application might not know many Grid related issues

- Provides Grid expertise for the middleware implementation into the application computing model
  - Software development towards the final implementation
  - Creation of missing components which might be shared by other communities
  - Evaluation of the middleware functionality and scalability tests based on the community needs
- Management of VO-specific services
- User-support if needed (a la GGUS)
- Expertise to solve Grid related issues for running applications
  - To the site and middleware level
  - Deep knowledge of both Grid infrastructure and VO computing model
- Documentation and training (application oriented)
- Liaison with the sites
- Creation of Grid implementation tools
- Creation of the community monitoring systems
- Operations activities if needed
  - Creation and registration of the new VO, eventually VO manager, negotiation of resources and services, etc

10.2. Inter-disciplinary Scientific Collaboration

In the context of a multi community HEP SSC that will include also photon science, astro-particle physics and Fusion applications, the current collaboration which has been created with these large communities in the context of separated clusters in EGEE will be highly promoted. Based on the experiences gained during the lifetime of the EGEE project, the unification of efforts for those communities which can share similar requirements or goals is a fundamental task which will need to be highly promoted in EGI. This will follow to clear benefits from the point of view of the man power effort, middleware creation (middleware duplication should be highly discouraged) and definition of operation procedures.

In spite of the relevance of this issue, it should be treated to the level of SSC definition and setup rather than at the level of the transition phase. In this phase, although the basis of the future SSCs will be defined, the guarantee of a stable and consistence Grid environment will be the major issue that we will have to face.

10.3. Summary of Achievements

1. **CCRC'08**

2. **STEP'09**

3. **Others**

10.3.1. Summary of Future Activities

10.3.2. Summary of Expected Impact

10.3.3. Summary of Manpower Requirements

10.3.4. Summary of Potential Funding Sources

10.3.5. Conclusions

10.3.6. Glossary

11. References

[1] "Dashboard for the LHC experiments", J. Andreeva et al., CERN-IT-NOTE-2007-048, November 2007, in the Proceedings of Conference on Computing in High Energy and Nuclear Physics (CHEP07), Victoria, Canada, September 2007.

[2] "Ganga: a tool for computational-task management and easy access to Grid resources", J. T. Moscicki et al., accepted for publication (May 2009).

[3] http://technical.eu-egee.org/index.php?id=290

[4] J.T.Moscicki et al.: *Dependable Distributed Computing for the International Telecommunication Union Regional Radio Conference RRC06*, http://arxiv.org/abs/0906.2143, submitted to IEEE Computer

[5] H.C.Lee at al: *Grid-enabled high-throughput in silico screening against influenza A neuraminidase,* IEEE transactions on nanobioscience. 01/2007; 5(4):288-95.

[6] J.T.Moscicki et al.: Ganga: a tool for computational-task management and easy access to Grid resources, http://arxiv.org/abs/0902.2685, submitted to Computer Physics Communications

[7] N. Santos and B. Koblitz: *Metadata services on the Grid*, Nucl. Inst. & Meth. in Phys. Res. A 559(2006) pp 53-56.

[8] N. Santos and B. Koblitz, Metadata services on the grid, in Proc. of Advanced Computing and Analysis Techniques (ACAT'05), Zeuthen, Berlin, May 2005.

[9] N. Santos and B. Koblitz, Distributed Metadata with the AMGA Metadata Catalog, in the proceddings of the Workshop on Next-Generation Distributed Data Management HPDC-15, Paris, France, June 2006.

[10] N. Santos and B. Koblitz, *Security in distributed metadata catalogue,* Concurrency and Computation: Practice & Experience 20 (2008), pp 1995-2007.

[11] J. Montagnat, A. Frohner, D. Jouvenot, Chr. Pera, P. Kunszt, B. Koblitz, N Santos, C. Loomis, R. Texier, D. Lingrand, P. Guio, R. Brito Da Rocha, A. Sobreira de Almeida and Z. Farkas: *A Secure Grid Medical Data Manager Interfaced to the gLite Middleware*, J Grid Computing (2008) 6:45–59 DOI 10.1007/s10723-007-9088-2.