CERN COMPUTER NEWSLETTER

Volume 43, issue 3   July–September 2008

Contents

IOP Publishing

IT groups focus on physics computing

Following changes to the IT Department's physics computing groups at the start of the year, we interviewed each of the group leaders concerned during the final moments before the LHC start-up.

Ian Bird, LHC Computing Grid project leader
What is the reason for the changes to the IT physics computing groups?
The idea is to focus more on the really important things that we need to do in the next few years as the accelerator starts up. For example, large-scale data management, mass-storage systems and the associated Grid tools that are the foundation of LHC computing.

What are the main changes?
A new Data Management (DM) group has been formed, led by Alberto Pace, to bring together all of the relevant expertise that was previously spread across several groups. This will better enable us to have a coherent data-management strategy, and to be able to cross-train development staff on all of the systems so that we don't have key components relying on one or two people.

At the same time, when Markus Schulz took over as group leader of the Grid Deployment (GD) group, it was agreed that the entire production Grid services would move to the Fabric Infrastructure and Operations (FIO) group so that they could be managed with the same teams and procedures used for other computing services. This leaves GD to focus on the broader Grid services not specific to CERN, which includes running the EGEE Grid operations and providing the Grid middleware building and testing services.

Another new group was also formed, led by Jamie Shiers. This is the Grid Support (GS) group, which is responsible for steering the overall WLCG service in all of its aspects, as well as dealing with Grid support for the experiments.

What does the LCG group do?
It manages day-to-day project operation, ensuring that everything runs smoothly.
● Catharine Noble has just taken over from

Fabienne Baud-Lavigne to look after the reporting of accounting and reliability for WLCG, managing the website and running the LCG office.
● Sue Foffano took over from Chris Eck as resource manager, responsible for coordinating and reporting on human and computing resources for the project, as well as managing the budgets of the physics computing groups.
● Alberto Aimar is responsible for overall project planning, including the follow up of our high-level milestones and quarterly reporting. He also produces the *LCG Bulletin* and spends part of his time as an activity leader in the ETICS project.
● Bernd Panzer-Steindel takes care of the resource and capacity planning for the Tier 0 and CERN analysis facility, and he tracks technology developments for WLCG.

The LHC Grid Fest – the official inauguration of the LHC Computing Grid service – will take place at CERN on 3 October. This event, which is by invitation only, will celebrate the launch of the WLCG e-infrastructure and will include video links to various Tier 1 centres around the world.

Alberto Pace describes the evolution of DM
Although featured in the last issue of *CNL*, the DM group has since evolved to have an additional section called Data Access to ensure that the future data-analysis requirements of the LHC experiments will be met successfully.

A new working group led by Bernd Panzer-Steindel has been created to look into the details of analysis activities at CERN, with a special focus on the storage aspects. It is expected to analyse the experiments' requirements thoroughly for future analysis.

In parallel, the DM group has launched several activities to consolidate and improve software data-management components currently deployed in the CERN Computer Centre. This is to ensure that all of the necessary building blocks will be available by the end of 2009, and to adapt quickly and implement new requirements of the experiments.

Editorial

Tony Cass talks about the changes in FIO

How has the FIO group been affected by the reorganization?
In terms of organization, the most noticeable change is to the Fabric Development section, which is now much smaller without the CASTOR developers. It is led by Véronique Lefébure, who switches from being a user of quattor, Lemon and the rest of the ELFms toolkit to being responsible for future developments.

The Fabric Services section has also greatly changed. Olof Bärring, the section leader, has had to reorganize to cope with the departure of Véronique and the arrival of new staff and responsibilities from GD. The FIO group is now responsible for the day-to-day operation of Grid-level services, such as the LHC File Catalogue and the File Transfer Service, as these move out of development and into routine-production mode.

Elsewhere, the Linux and AFS section led by Jan Iven, the Technology and Storage Infrastructure section led by Tim Bell, and the System Administration and Operations section led by Vincent Doré remain unchanged, keeping their focus on supporting Scientific Linux, AFS, CERN's tape robotics complex and the day-to-day operations of the Computer Centre.

What are the main challenges for FIO in the coming year?
LHC data. Despite all of the preparation and testing, we don't know what will happen when real data arrive and real analysis starts. Our services have performed well during the Combined Computing Readiness Challenges, but I am sure that the start of LHC operations will bring something new.

So the two key points that we want to address this year are:
● the migration of services to routine production – with installation, monitoring and operation integrated with other FIO services and understood by the team rather than one or two experts;
● improvements to our automation to reduce the impact of hardware failures and to allow us to react dynamically and automatically to changing workloads.

Markus Schulz explains how the GD group is organized
The GD group covers several areas related to Grid computing.
● The Operations section, led by Maite Barroso Lopez, is coordinating the operation of the EGEE infrastructure. "Grid operations" is a very diverse term and has wide-ranging activities. For example, there are 255 sites providing resources for about 180 000 jobs each day. The partners are located in 48 countries, representing a total of 228 FTEs. Site monitoring and Grid security coordination are also important activities in the section.
● The Integration Testing and Releases
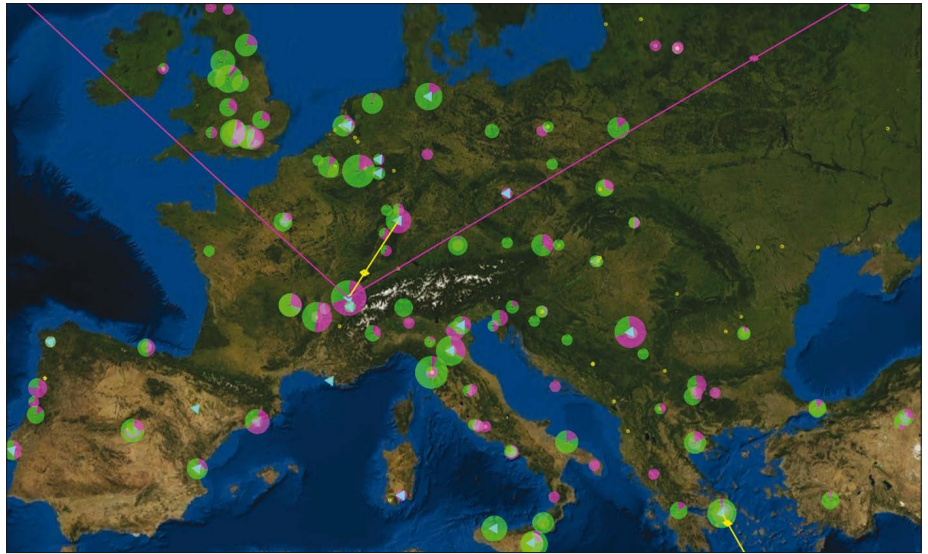
The GridPP real-time monitor provides a view of Grid activity on a map of the Earth.

section is led by Oliver Keeble, who has also taken up the role of EGEE-SA3 activity manager. As the name suggests, this section builds and tests the weekly updated releases of more than 250 packages in the gLite distribution. The team coordinates the work of 17 contributing institutes in 13 countries. All Grid components are scrutinized by the section, which has amassed an enormous amount of knowledge that proves invaluable for support.
● The Software Lifecycle Tools section is headed by Alberto Di Meglio, who has also been the ETICS project director since it started in 2006. ETICS, an EU-funded project, provides integrated services for managing the software lifecycle of complex projects such as gLite. This includes build services for a large number of platforms, automated testing and managed repositories. ETICS is used by 37 projects with 263 active developers. The system maintains close to 9000 software configurations and about 180 build and test jobs are scheduled daily on dedicated clusters.

How do you see the LHC start-up for GD?
The different GD activities operated smoothly during this year's Common Computing Readiness Challenges. While confident, we are aware that the start of the LHC will bring up some interesting challenges for our teams.

The increased scale and criticality of the infrastructure usage will undoubtedly uncover some surprising issues. This will mean that swift action from the operations and release preparation teams might be required at any time. In particular, this will be a test for these activities, which have recently been confronted with accelerated staff rotation.

In the long term, GD strives for increased efficiency by automation and the

progressive devolution of responsibilities throughout the infrastructure.

Jamie Shiers looks at the Grid Support group
Although now in the final ramp-up to first data from collisions in the LHC, for many people – including me – LHC computing started in the early 1990s. Discussions about future programming languages and methodologies, tools, databases, data management and so forth led to a somewhat turbulent era in the mid-1990s, but eventually gave way to a clearer picture by the late 1990s with the emergence of the MONARC model and finally the Grid computing paradigm in 2000.

Leading the Grid Support group now, after many years of preparation and hardening of Grid services by a huge worldwide collaboration, is not only a challenge but also a privilege. Being so close to the LHC experiments and the sites that form the WLCG collaboration is surely one of the most exciting places to be in physics computing at this time.

How is the group organized?
The Grid Support group has three sections:
● Experiment Integration Support section, led by Harry Renshall;
● Monitoring 'n' Dashboards, led by Julia Andreeva;
● (distributed) Data Management and Analysis, led by Massimo Lamanna.

I hope that the roles of these sections, and hence the group, are clear. In addition, the group is involved in the coordination of the overall WLCG Service – including the organization of WLCG Collaboration workshops – as well as coordination of this year's Common Computing Readiness Challenges. This involves constant interaction with other physics and infrastructure groups in the department, as well as with the experiments, sites and members of the Grid communities.

Laptop registration process speeds up

Since its deployment, the laptop registration for a CERN registered user required that their human resources (HR) data (including office and phone numbers, group and department units) had been fully validated in the HR database.

This meant that many new users had to wait for one or two days before they could register their laptop, causing unnecessary delays. This will no longer be the case, thanks to improvements implemented by IT-AIS (Administrative Information Services) and IT-CS (Communication Systems) experts.

The registration of all portable computers connected to CERN's network was enforced in December 2003. Since then, only registered computers are allowed network access on wired and wireless connections.

The central network database (LANDB) is managed by a number of different software tools, among which is the well known public interface Network Registration Form, which is available at http://network.cern.ch. The use of this interface is mandatory to register any computer at CERN, and to keep the registered information up to date.

LANDB and the Network Registration Form apply a precise set of rules to decide if a person is authorized to be responsible for, or a user of, a network resource. These rules are as follows:
● The person's CERN account is neither blocked nor expired.
● They have a valid registered e-mail address.
● The person is at CERN.
● They have signed Operational Circular no. 5 (OC5).
● They have an active (not terminated) affiliation/contract with CERN.
● They have a department assigned.

This information is stored in the central HR and Computing Resource Administration (CRA) databases. Part of this information was transferred into LANDB each night, so LANDB was using a delayed copy of this information. This prevented a newly registered person from being recognized as valid in LANDB on the same day, usually displaying a message informing the user that the information had not yet been synchronized. Consequently, they had to wait until the next day, when LANDB found the updated information and authorized the laptop to be registered.

Thanks to the joint effort between IT-AIS and IT-CS, a direct connection now exists between LANDB and Foundation, removing the one-day delay in registration.

This enhancement was deployed in mid-June and will make life easier for laptop users, as well as reducing the number of questions directed to the departmental and experimental secretaries, the Users' Office, the Computing Helpdesk staff and the Network Operation team.

This modification does not affect the registration process for short-term visitor's laptops (users with no CERN account).

However, the computer visitor registration in the Network Request Form has been extended so that all people considered "members of the personnel" as defined by the latest edition of the "Staff rules and regulations" can sign and become responsible for a visitor. This means that associates, users and students, in addition to staff members, fellows and apprentices, can accept the requests.

In addition, a service account attached to a mailing list can be specified as "responsible" for requesting visitor access. Any authorized members (as defined above) of this mailing list can accept the request and then become the responsible contact for that particular visitor.

Useful links
Operational Circular no. 5 (OC5): http://cern.ch/ComputingRules
Network registration: http://network.cern.ch
Computer registration: http://cern.ch/it-dep/registration
Catherine Delamare, Jose Carlos Luna Duran and Wim Van Leersum, IT Department

Accounts converge into single credential pair

Over the past few years the IT Department has been streamlining CERN users' access to all of the central computing services. For each user the long-term goal is to converge on a single CERN account with a unique credential pair (username and password). This strategy will make IT services more coherent and much easier for users to understand. It will also simplify account maintenance and provide a central point of control where security measures can be applied.

As the next step of this process, by 1 July your CERN, PLUS and AFS accounts will have converged into a single CERN account. From now on the account names will be unique and universal.

The passwords will become unique and universal after the first password change, which users are encouraged to make at their earliest convenience. Until then the existing passwords will remain valid on each individual service, but afterwards the new credentials will become truly common to all three services.

Thus, changing the password for PLUS or AFS services must now be done exclusively through the web interface at http://cern.ch/cernaccount. Any password problem should be addressed to the Helpdesk (tel +41 22 767 88 88).
IT Department

IT strengthens reset password procedure

From 19 August the security of the CERN account password reset procedure has been strengthened. As a result, users requesting to have their password reset by the Computing Helpdesk will be asked to provide some private information. This will have to occur before the Helpdesk can reset the password.

Please note that all relevant information about the CERN account and password can be found at http://cern.ch/it-dep/AccountsandpasswordsatCERN.htm.

Thank you in advance for your co-operation in this process.
IT Department

Web attacks target instant messaging

Be cautious of any unexpected messages containing weblinks, even if they appear to come from known contacts. If you click on a link and your permission is requested to run or install software, always decline it.

Several computers at CERN have been broken into by attackers who have tricked users of instant messaging applications (e.g. MSN, Yahoo Messenger) into clicking on weblinks that seemed to come from known contacts. The links appeared to be photos from "friends" and requested software to be installed. However, this was attacker software. In the past fake messages were sent mainly by e-mail but now a wider range of applications are being targeted, including instant messaging.

Cybercriminals use fake messages to try to trick you into clicking on weblinks, which will help them to install malicious software on your computer. Anti-virus software cannot be relied on for detecting all cases. Your vigilance is needed. If you have any questions, contact helpdesk@cern.ch.
IT Department

Grid news

Europe e-infrastructure ideas guarantee equal participation

Soon after the launch of the first phase of the EC-funded EGEE project (http://eu-egee.org/) in May 2004, many other regions followed suit and set up their own e-infrastructure projects. The first and most notable was the South-East European/SEE Grid e-Infrastructure Development (SEE-GRID) project.

The highlights of the first two phases of SEE-GRID include:
● creating a powerful regional e-infrastructure using EGEE gLite as Grid middleware;
● successfully attracting a number of SEE applications from diverse fields and deploying them in the SEE e-infrastructure;
● establishing a strong, highly skilled human network in the area of Grid computing;
● encouraging and supporting national Grid initiatives in south-east Europe.

The timescale linking with EGEE continues and, at the same time as EGEE-III was launched, the third phase of SEE-GRID started – SEE-GRID for e-Infrastructure for Regional e-Science (SEE-GRID-SCI). This phase aims to build on previous results and in particular:
● to enable new scientific collaborations;
● to enlarge its e-infrastructure further with additional resources;
● to continue supporting the consolidation of national Grid actions in SEE.

Regional scientific collaborations will be promoted in three user communities: meteorology, seismology and environmental protection. These are communities that currently do not directly benefit from the SEE e-infrastructure and/or are not yet engaged in cross-border collaboration.

Currently the e-infrastructure has 35 sites offering in excess of 2200 CPUs and 57 TB of storage and these resources will increase substantially over the course
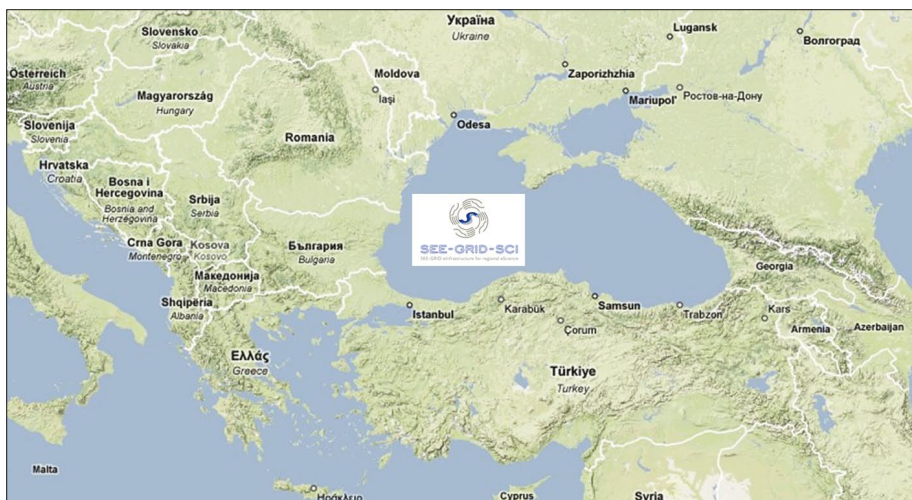
A map of the geographical area of the SEE-GRID-SCI e-infrastructure development project.

of the project to meet the demands of its growing user base. Many of the resources are shared with the EGEE production e-infrastructure, offering services to high-energy physics, biomedical and earth-science communities.

In this third phase, the e-infrastructure scope will expand to include two new countries, meaning that the project consortium will be composed of 15 funded institutions: 14 from the SEE area (Albania, Armenia, Bosnia and Herzegovina, Bulgaria, Croatia, Former Yugoslav Republic of Macedonia, Georgia, Greece, Hungary, Moldova, Montenegro, Romania, Serbia and Turkey) and CERN. GRNET (Greece) continues as project coordinator for this two-year project, which started on 1 May and is co-funded by the EC under the 7th Framework Programme (http://cordis.europa.eu/fp7/home_en.html).

SEE-GRID-SCI is organized into six activities: management, national Grid initiatives, dissemination and training, user support, Grid operations, and the development of application-level services and tools.

CERN will have four main activities within the SEE-GRID-SCI project:
● **Support** the deployment and monitoring of core gLite services in the SEE-GRID-SCI region.
● **Analysis** of existing operations and user support tools from EGEE, and investigation of their usability and areas of improvement or extension to meet the requirements of SEE applications.
● **Liaise** between the operations and development teams of SEE-GRID-SCI and EGEE.
● **Promote** further collaboration with other Grid projects.

Useful link
SEE-GRID-SCI project website: www.see-grid-sci.eu/.

Maria Barroso Lopez, Florida Estrella and Frédéric Hemmer, IT Department

The deadline for submissions to the next issue of CNL is

24 October

Please e-mail your contributions to cnl.editor@cern.ch

AFS revisited: controlling access

These "AFS revisited" articles are intended as short reminders looking at standard AFS commands, common pitfalls and some tips and tricks to get more out of AFS at CERN.

Data access in AFS is controlled by per directory access control lists (ACLs), which map user or group identities to a set of access rights. The *listacl* and *setacl* subcommands of the *fs* command suite are the interface to list and change the ACLs of a directory. Here is an example of a listing:

```
$ fs listacl $HOME/alliance
Access list for /afs/cern.ch/user/p/pam/
alliance is
Normal rights:
        pam rlidwka
        jim rlidwk
```

To understand what type of rights have been granted to users Pam and Jim, this shows the seven rights provided by AFS to control access to a directory and to the files that the directory contains:

● **r = read** allows the named user to read the file content and to query file status.
● **l = lookup** allows the user to list the files and directories, to examine the ACL and to access the subdirectories.
● **i = insert** allows the user to add new files and directories.
● **d = delete** allows the user to remove files from a directory and to remove subdirectories, for which they have insert right.
● **w = write** allows the user to change the file content (and also to change the UNIX permission bits, see below).
● **k = lock** allows the user to use full-file advisory locks. (Note that there is no byte-range locking available in AFS.)
● **a = administer** allows the user to change the ACL of a directory.

An additional constraint to make use of these privileges is that the user has at least the right to list (access right "*l*") all parent directories.

To make life easier when setting ACLs via *fs setacl*, AFS provides some shortcuts to assign certain sets of rights:

● *read* corresponds to *rl* and provides read access;
● *write* corresponds to *rlidwk* and grants read and write access;
● *all* corresponds to *rlidwka* and gives full access;
● *none* removes all permissions for a user.

Assigning permission *none* removes the rights of a user but does not prevent access in all cases because the user may be a member of a group that still has the right to access a directory. To prevent access to directories explicitly, ACLs in AFS support negative rights. By adding the negative option when setting ACLs, you can deny users rights that they would otherwise have by virtue of their group membership.

```
$ fs setacl -dir $HOME/alliance -acl dwight
all -negative
$ fs listacl $HOME/alliance
...
Negative rights:
        dwight rlidwka
```

User dwight has now been denied all rights on the alliance directory, even if a group where he is a member is granted access later on. Caveat: if anonymous groups like system:anyuser are on the ACLs, the assignment of negative rights will not have the desired effect.

A common pitfall

By far the most confusing parameter that comes with the *fs setacl* command is the *-clear* option. This does not clear all of the rights of a given user or group but rather acts as a two-step command that wipes out the whole ACL and then assigns the rights given.

Taking our example from above,

```
$ fs setacl -dir $HOME/alliance -acl dwight
all -clear
```

would not remove all permissions for user dwight but remove all entries from the ACL and then assign all permissions to dwight, which is contrary to the original intention.

How do ACLs and UNIX permission bits interact?

For directories, AFS completely ignores all of the UNIX permission bits – only the AFS ACL rights concerning directories (i.e. lida) are taken into account.

For files, the situation is a little different: AFS also ignores the group and world permissions for files, but the user permission bits act as an additional level of access control after the ACL permission checking has taken place.
● The *r* bit allows anyone with *rl* on his ACL to read the file. If the *r* bit is not set, no one can read that file (not even the owner).
● The *w* bit allows anyone with *wl* on his ACL to write to the file. If the *w* bit is not set, no-one can write that file (not even the owner).
● The *x* bit allows anyone with *rl* on his ACL to execute the file. If the *x* bit is not set, no one can execute that file (not even the owner). This also implies that you cannot have a file in AFS that a user can execute but not read.

So, by using the combination of ACLs and UNIX permission bits it is possible to define something close to a per file access control, even though it does not provide the full flexibility of AFS' per directory ACLs. There is currently a discussion in the AFS community to extend AFS to support fully fledged per file ACLs.

It should be emphasized that AFS does not really care about group membership or user ID (although there have been applications in the past that tried to outsmart the system and take decisions based on that). As a consequence, you should not be worried if an *ls -l* displays unknown owners or groups, because really only the UNIX owner bits matter for AFS.

Further information
The openAFS website provides information about all AFS commands (http://www.openafs.org/manpages). The CERN *AFS User Guide* is a more comprehensive manual and is available at http://cern.ch/service/afs.
Arne Wiebalck, IT-FIO

Subversion-based system replaces central CVS service

The CERN CVS Service (Concurrent Versions System, see http://cern.ch/cvs/) is currently hosting more than 300 software projects providing a version control (VC) infrastructure to more than 3000 developers both at CERN and worldwide. This service holds more than 90 GB of source code, which covers most of the code that has been written at CERN during the last few years. The full list of software projects is available at http://cern.ch/cvs/howto.php#help.

The IT Department is now preparing to replace its central CVS service with a more modern Subversion (SVN)-based service. SVN should provide better performance and many other new features, such as read access control, atomic commits and offline operations. This article describes the motivation and the milestones for a future CVS to SVN migration.

Software version control at CERN
A VC system (also known as revision control system) provides the means to control versions of files. It is a key tool for software engineering, particularly on large software projects where several developers are working on the same source code files and need to be able to merge all of their modifications as well as track revisions.

CERN's VC system is based on CVS, a system that was created in 1986 to help the open-source software community, where many developers contribute to each project. CERN's IT Department put in place a central CVS service in 2000. This provides features such as high availability, automatic load balancing, web interface to repositories, remote repository replication and automatic building of usage statistics. Originally this service was designed to host a few tens of software repositories and a few hundred developers. It has expanded well beyond that target, reaching a peak of more than 100 000 commit (write) operations per month (figure 1).

Reasons for change
CVS deserves to be honoured with a special award in the history of software engineering, but it has many limitations as a VC system (table 1). The most constraining limitation is that CVS transactions are not atomic. Some transactions may be interrupted halfway through, leaving corrupted source code files as well as locked files behind. This is intrinsically linked with the fact that CVS
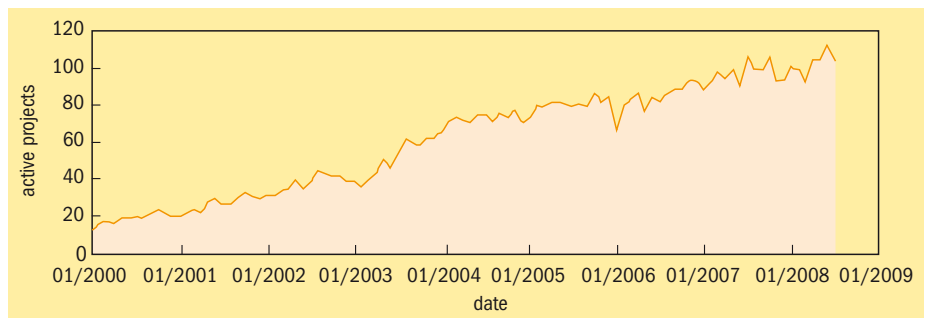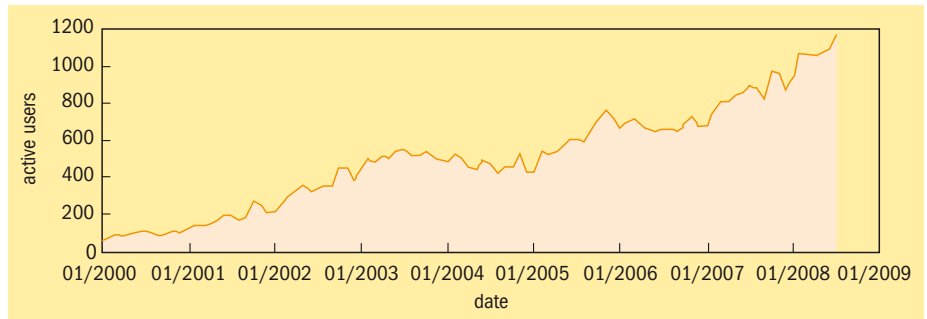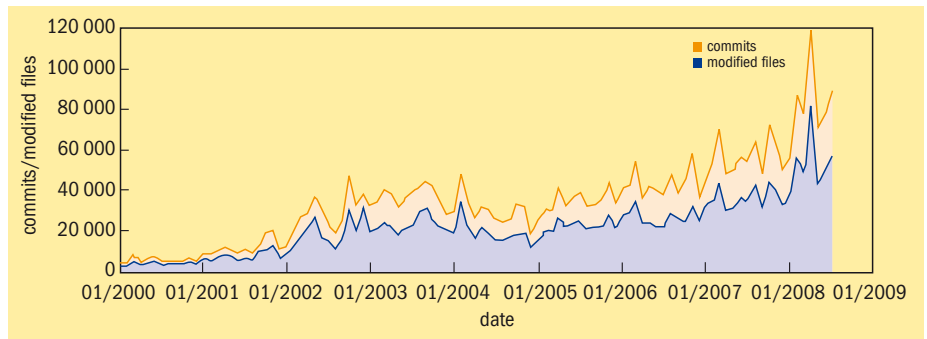
Fig. 1. CVS service activity has reached a peak of more than 100 000 operations a month.

does not use a transactional database but rather relies on the file system where the repository resides.

Another important feature that is missing, which has often been requested by CERN CVS users, is read access control. Being designed to be used within open-source projects, the last thing that CVS authors had in mind was implementing any kind of read access control mechanism because all source code was supposed to be publicly available.

There are many commercial and open-source VC systems, such as GIT, Bazaar, Darcs, GNU arch, Mercurial and Monotone, but the most popular seems to be SVN. It was initially released in 2000. Since then it has been adopted by many

large open-source as well as commercial software projects. It is used by software projects such as Apache, GNOME, KDE, FreeBSD, GNU CC compilers and Python.

SVN is also a widely used VC system in the physics user community. It is used in institutes such as IN2P3 and Fermilab, and in many projects at CERN (e.g. ROOT and Totem-DCS). SVN overcomes most CVS limitations mentioned above and it has often been requested by CERN's CVS user community. It is also worth mentioning that there is a wide choice of SVN clients for both Windows and Linux/Unix platforms, which is not the case for other modern VC systems, such as GIT.

Due to this increasing popularity, more locally managed SVN servers have been

observed across CERN in the past two years. This is neither optimal in terms of CERN's resource usage nor a good basis to standardize on robust security and back-up practices.

Preparing for CVS to SVN migration

A plan (figure 2) has been put in place to allow, in the long term, full replacement of CVS by SVN as the central VC system at CERN. In a preliminary study, administrators of already existing SVN servers at CERN and in other physics institutes were surveyed to identify their SVN set-up and implementation choices. CVS librarians were requested to answer a questionnaire about their SVN preferences in terms of web interface, access method, etc. This led to the setup of an SVN pilot project, which was made available to CERN users in July at http://cern.ch/svn/.

The SVN pilot project aims to deliver all of the features currently offered by the central CVS service while adding new functionalities. To make provision for automatic load balancing and fail over, a cluster set-up using a distributed file system where all SVN repositories reside has been chosen (figure 3). There are two instances of this architecture to split SVN read/write access operations from web read-only access to SVN repositories. This way, the SVN facility will not be overloaded with web access requests. These will be handled on a separate cluster.

Regarding access methods to SVN pilot servers, SSH/svnserve access has been chosen in favour of HTTPS access. Performance tests during the preliminary study showed that HTTPS is noticeably slower than SSH/svnserve. This is due to the fact that HTTP is a stateless protocol and requires more turnarounds. The choice of web interfaces to SVN is quite large. For the pilot, WebSVN and Trac were selected, based on the CVS librarians' preferences.

Repository migration

One of the most critical issues when migrating to SVN is converting already existing CVS repositories into SVN format. For this reason, the plan is to keep both the central CVS service and the future SVN service running in parallel for some time while encouraging the migration of all CVS repositories to SVN. Additionally, a copy of all CVS repositories will be kept.

There are a few tools that can help librarians performing this repository conversion. The most popular is CVS2SVN (http://cvs2svn.tigris.org/), which does the job pretty well, but it is up to each librarian to decide which options to use. There are several choices depending on each project's needs, varying between full conversion (keeping all tags, branches and history) to the so-called "top-skimming", which is light and saves significant disk space but preserves no tags, branches

CVS vs SVN features

Feature	CVS	SVN
atomic commits	no	yes, uses transactional DB
file and directory moves	no	yes
file and directory copies	no	yes
remote repository replication	no	yes, used for branching
propagating changes to parent repositories	no	yes, using svn-push
read access control	no	yes
access protocols	SSH, pserver, kserver	SSH, https
handling of binary files	limited, no real difference	yes
web interfaces	ViewVC and CVSweb	ViewVC, Trac and many others
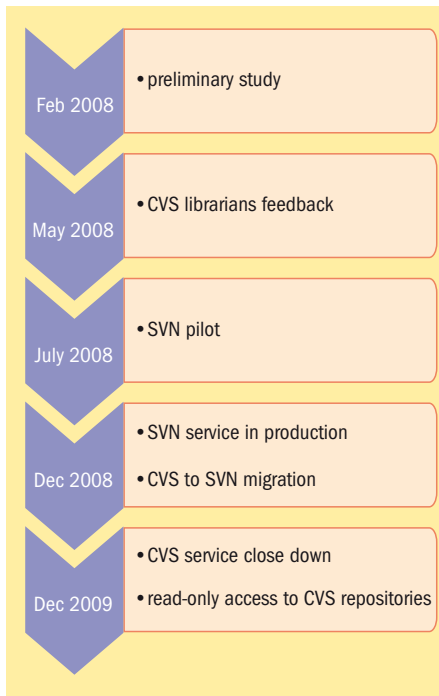
Fig. 2. CVS to SVN migration milestones.

- Feb 2008 — • preliminary study
- May 2008 — • CVS librarians feedback
- July 2008 — • SVN pilot
- Dec 2008 — • SVN service in production • CVS to SVN migration
- Dec 2009 — • CVS service close down • read-only access to CVS repositories

Fig. 3. SVN cluster service architecture.

repository on a distributed file system
servers
DNS load-balancing
client

or history. There are also various other CVS2SVN options for symbol handling, encoding, revision number, etc.

More details about how to use CVS2SVN are available on the SVN Pilot webpages (http://cern.ch/svn). CVS librarians can start testing the look and feel of their converted CVS repositories by creating a new SVN project.

Conclusion

The goal of the pilot is to deliver a production-level, modern, feature-rich VC service, but migrating to a new VC system is not painless and requires modifying development frameworks as well as changing some developers' habits. Users of the pilot need to accept that this is not yet a production service and it may suffer some disturbances in its early stages. However, their feedback will be a key factor that will help us to improve and stabilize the future production SVN service from December.

CVS librarians should keep in mind that

our target is for SVN to replace CVS as the main VC system at CERN by the end of 2009. We believe that the convenience brought by SVN is worth the efforts of rewriting tools and moving the current CVS infrastructure to SVN. Therefore, all CVS users (and particularly librarians) are encouraged to participate in this SVN pilot project.

Useful links

Concurrent Versions System: www.nongnu.org/cvs/
CERN's CVS pages: http://cern.ch/cvs/
CERN's SVN Pilot webpages: http://cern.ch/svn
Subversion SVN: http://subversion.tigris.org/
IN2P3 SVN site: https://svn.lal.in2p3.fr/
Fermilab SVN site: www-numi.fnal.gov/offline_software/srt_public_context/WebDocs/subversion.html
WebSVN: http://websvn.tigris.org/
Trac: http://trac.edgewall.org/

Hugo Hugosson and Manuel Guijarro, IT-DES

Technical brief

CERN develops Enterprise GSM monitoring tool

In the last decade the development in mobile phone services (GSM) has led many organizations to rely on this new communication technology. As for all IT services, the necessity of monitoring GSM services has become critical in order to be able to react appropriately in an incident. Mobile operators have their own monitoring system, but they do not take into account corporate constraints and it is necessary to evaluate an operators' performance against a service-level agreement independently. After an extensive market survey showed that commercial solutions were not yet mature, the telecom team decided to develop its own monitoring system.

Since 1991, CERN has subcontracted the installation and maintenance of a GSM network covering all of its sites to national mobile operators. In addition, the telecom team has installed more than 50 km of leaky feeder cable to propagate GSM frequencies in all underground facilities (experiments, LHC tunnel, PS, SPS, etc). This cable infrastructure is also used to propagate VHF radio signals for the CERN fire brigade. At CERN's request, the mobile operator implemented a private network offering advanced services to CERN users, such as short dialling plan, different roaming classes, flat-rate subscriptions, mobile data services (GPRS/EDGE/3G) and e-mail to SMS services. Today, with more than 4000 GSM users (CERN staff, external institutes, subcontractors, etc), GSM has become one of the most effective ways of communicating at CERN.

The need for CERN GSM network monitoring

Even if our mobile operator is monitoring the active equipment of the GSM network, it is not possible to guarantee the availability of GSM services where signals are carried by the leaky feeder cables. A portion of cable can be damaged for any reason (e.g. cut, pinch, faulty FQ combiner, faulty connector), degrading the propagation of GSM signals. We have yet to gain experience of how many years this infrastructure will survive during LHC operation and how the service will degrade. Given that the fire brigade VHF network also relies on the GSM network, it is clear that GSM monitoring is essential to be able to inform CERN users in case of an incident.

CERN asked an electronic manufacturer to build GSM probes and develop a server to collect information recorded by these probes. They are located in "strategic" locations and monitor the level of GSM frequencies. Basically, a probe is placed closed to each GSM emitter to monitor local frequencies, and another one is placed at the far end of a leaky feeder cable segment to monitor the same frequencies. If the

Leaky feeder cable-monitoring concept for mobile phone services in underground areas.

VHF relay and GSM base station at BA2.

second probe detects a frequency problem, but the first one doesn't, we can deduce that the segment of leaky feeder is not working. As VHF frequencies are much less sensitive to the variations of leaky feeder cable propagation, we can guarantee that these signals are working if the GSM frequencies are working.

The monitoring server collects data from the remote GSM probes via the CERN GPRS network. For each probe, a list of GSM frequencies to monitor is configured with respective thresholds. If a threshold is reached, the server activates a cabled alarm to the CERN Control Centre (CCC), by opening a relay contact attached to its serial port. The CCC can then alert the telecom piquet service, which will undertake appropriate actions.

In case of problems with the GPRS network, the server generates an alarm and continues to collect data from probes via an exchange of SMS messages. A master probe attached to the server sends/receives the

SMS messages. This master probe can also generate an alarm if the connection with the server is lost.

A webpage shows the status of all GSM probes at CERN. From here it is possible to add, configure and delete probes. Finally, the server and probes are also monitored in the network management server (SNMP platform) to relay alarms via IP to the computer centre.

The server will store data and logs over many years. Thus, it is possible to analyse the evolution of propagations over time and to plan proactive maintenance. The probe software can be remotely flashed to avoid manual intervention when an update is required. Five GSM numbers are configured in the probes with special rights, which allows technicians on site to query and configure the probes via SMS messages.

The GSM monitoring system has been operational in the LHC, PS and SPS for six months. It can monitor the availability of GSM frequencies, GPRS services and VHF signals (indirectly) in these underground areas. New probes will be installed on the surface to monitor the whole CERN GSM network. This innovative solution is a real success and is now being proposed by Swisscom to its clients who face the same constraints as CERN.

Frédéric Chapron, IT-CS

Software factories inspire process control

For the CERN experiments, IT-CO is developing process-control applications based on UNICOS, a CERN framework constructed from reusable PLC (industrial computers) and SCADA (e.g. PVSS) software components.

To allow the extension of the supported target platform and the easy integration of new components in UNICOS-based control applications, IT-CO, in collaboration with AB-CO, is in the process of upgrading its code-generation tools towards a software factory. This factory is in charge of assembling project, domain and technical information seamlessly into deployable PLC and SCADA components.

To generate a UNICOS process-control application, three types of information are typically required:

● **The raw data**, which characterizes the participating elements to the process-control system, such as the list of sensors and actuators to read temperatures and switch on pumps.

● **The logic** to describe how to use these elements to fulfil the functionality of the process-control application. This knowledge is rather independent of the final platform.

● **The technical knowledge** to express the raw data and the logic of the system in the vendor-specific language of the PLC and SCADA target platforms.

With flexibility and extensibility in mind, our new code-generation tool, the so-called UAB (UNICOS Application Builder), is designed to enable faster and cheaper code generation in the context of frequently changing user requirements. By abstracting the technical details, it allows the process-control experts to focus on the functionality and behaviour of their process-control systems.

The raw data and their grammar

The raw data are expressed in XML and constrained by XML schemas, presented here as the "Grammar check" packet in figure 1. These XML schemas are designed as an extensible asset to be shared across process-control projects. The internal representation of the raw data is managed by the JAXB library (Java Architecture for XML Binding). Thanks to JAXB, structural modification of the raw data can be achieved on the fly and the code-generation rules can navigate and manipulate the raw data in a natural fashion.

The logic and the code-generation rules

The logic is described in terms of code-generation rules. Just like the conductor of an orchestra, the rules simply dictate what the code-generation plug-in will do with the data through a
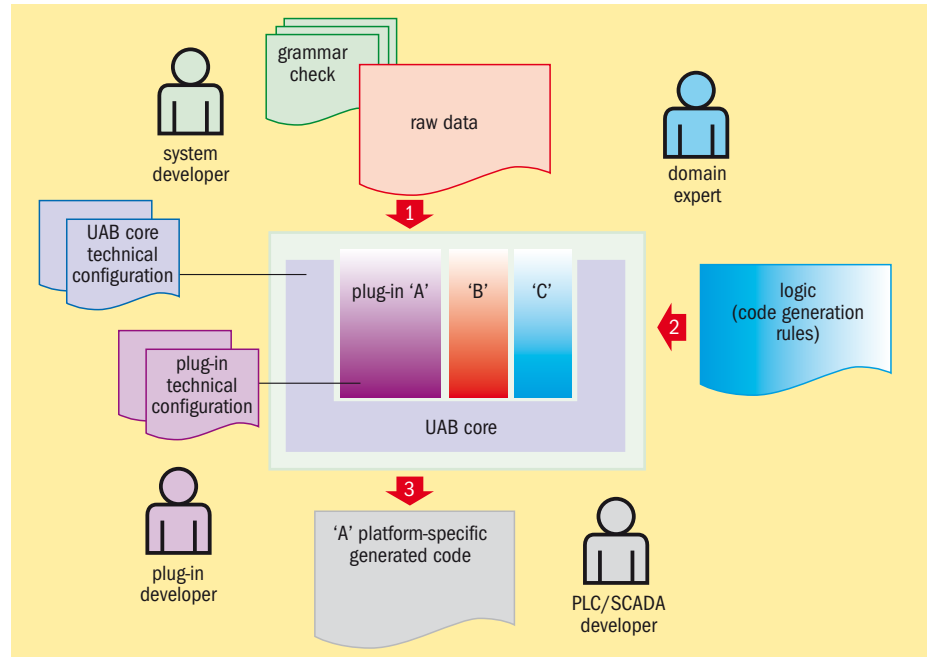

Fig. 1. UNICOS Application Builder packets and stakeholders: interactions and roles.

set of abstract services. The rules have been designed to enable platform syntax abstraction, a step further in helping to avoid error-prone syntaxes.

The code-generation rules consist of Jython script files (Python for Java). Using a genuine scripting language as Jython rather than a flat properties file has several benefits: it is standard, allows powerful constructions and integrates with Java to provide bidirectional communication.

Using code-generation rules is well suited to incremental development where the elements composing the final system are known gradually.

The technical knowledge

The UAB tool is actually a container for target-specific code generation plug-ins. Java has been chosen for the development of the UAB tool and its plug-ins. It permits high coding productivity and abstraction mechanisms such as introspection and run-time class loading, which are used efficiently in the UAB tool. To minimize maintenance, the UAB Core follows the broker design pattern and provides the plug-ins with an extensive set of high-level interfaces to connect the raw data to the logic, and to manage the traditional aspects of standalone software applications, such as user interface, command line support and online logging.

Each UAB plug-in focuses on pure code-generation aspects and simply "knows" how to transform the abstract requests from the code-generation rules into proper vendor-specific source

code. For example, one of our plug-ins is responsible for the Schneider UNICOS PLC code generation and only knows how to instate PLC objects and map them in PLC memory.

All plug-ins are built into the same model to make it easy to develop and integrate new ones, even with little programming experience. This approach not only prevents typing errors in vendor-specific code, but also allows the same raw data and logic to be reused by simultaneously generating the same process-control application on different platforms. This is useful when starting with a lab prototype, then reusing its assets to produce the system on the final platform.

● The UAB tool allows the developers to focus on the expected result rather than on the means to produce it. Mixing static configuration, auto-adaptive mechanisms and abstraction, it is a powerful, yet simple, user-driven code-generation tool. Initially designed for the code generation of process-control applications, the UAB concepts have been validated on PVSS and Schneider PLC platforms. Beyond process control, the UAB is particularly well suited to execute any versatile and powerful offline data transformations.

Useful links
UAB: http://itcofe/Projects/UAB/
UNICOS: http://cern.ch/ab-project-unicos
The GlassFish community (JAXB): http://java.sun.com/javaee/community/glassfish/
The Jython Project: www.jython.org
Mathias Dutour, IT-CO

Review motivates EGEE-II: project ready for next phase

In early July, project members of Enabling Grids for E-sciencE (EGEE) met in the CERN council chambers for the EGEE-II project's final review. Five reviewers – specialists in information technology nominated by the European Commission – attended the event. Project reviews are carried out annually to assess the project's performance and progress, and to give the consortium specific commendation and recommendations.

By all accounts the review was a success. Reviewer John Martin described it as "outstanding." He says that all of the reviewers were particularly impressed by the openness, transparency and responsiveness of the project. They commended it on consistently meeting and exceeding its targets. The reviewers noted the progress on application support and say that the degree to which the project has been able to contact and understand the needs of the user community is remarkable. They also note the risks in taking certain decisions, such as to restructure the gLite

middleware halfway through the project's lifetime and using ETICS as the sole build system, and they say that these risks have been vindicated by the outcome.

In the reviewer's final report the project earned the highest marks in every category. However, says Anna Cook, head of administration for EGEE, this was not a surprise: "We expected it to go well. We'd worked very hard in our preparations. The secret to this success was our hard work during the past year and our follow-up on their recommendations from the previous review. It's nice to have both praise and areas for development."

One of the suggestions from the previous review (June 2007) had been to strike a balance between stable operations and innovation. Striving for this, the project focused on stabilizing the infrastructure and moving applications from prototype phase to use in day-to-day research.

What effect has Cook seen this most recent review have on the project? "The need for summer holidays," she said,

laughing. "What was nice about the review was that we received endorsement from our reviewers for our future plans. Being told that we're on the right track has given us all some extra enthusiasm."

Now in its third phase, for the next two years EGEE will focus on expanding and optimizing the Grid infrastructure through support for more user communities, and adding more computational and data resources. The grand challenge will be preparing for the migration of the existing production European Grid from a project-based model to a sustainable federated infrastructure based on National Grid Initiatives. The project hopes for another "outstanding" final review in two year's time.

At CERN, 77 people currently work on EGEE. The project is a fundamental part of the LHC Computing Grid. The LCG uses EGEE's gLite middleware and many EGEE-managed resources.

Danielle Venton, IT-EGE
http://eu-egee.org

openlab's summer programme for students promotes high tech in a multicultural setting

Since its creation in 2003, CERN openlab – a framework for partnership with industry – has welcomed students in computer science or physics every summer to work on cutting-edge Grid technology projects and other advanced openlab-related topics. This year the programme accepted 13 students from 12 countries (China, Colombia, Croatia, France, Greece, Pakistan, Poland, Romania, Russia, South Africa, Spain and the USA) for two months, during the period June to September. The students were funded jointly by CERN openlab, two of its partners (Oracle and Intel) and the students' home universities.

The students worked on very diverse topics and were fully integrated into the openlab-related projects. The summer student programme is valued by both the students, who are given a chance to work in a highly demanding and leading-edge environment, and the CERN teams, who consider the students' work to be an asset to the technical part of their projects.

Sverre Jarp, openlab CTO responsible for the summer student programme, tailored a dedicated and enriching schedule including

The openlab summer students from around the world, in the CERN Computer Centre.

a series of nine lectures, given by CERN and external experts. The key topics covered were Grid overview and gLite details, server hardware, virtualization, compilers, secure software creation, computer architecture and performance optimization, LHC Computing Grid, networking and Oracle database architecture. Sverre explained: "In addition to their work in the various

CERN groups, we aim to give the students an overview of the current technologies that we use and also to introduce them to other activities at CERN."

In line with this spirit, visits were organized to the CERN Control Centre, ATLAS, LINAC 2 and the Anti-matter Factory. There was also a one-day trip to EPFL in Lausanne, where the students were able to attend presentations on information technology projects, such as biocomputing and environmental monitoring.

At the end of the summer, participants submit a written report, which can result in wider publication. However, for some this first experience at CERN is just a prelude to a longer collaboration. Although this is not its main purpose, the openlab summer student programme has also proved to be successful at convincing a few of these talented students to come back. Two of the current openlab team members first came to CERN as openlab summer students.

Useful link
http://cern.ch/openlab
Mélissa Le Jeune, IT-DI (openlab)

EGI Geneva workshop discusses blueprint for sustainable grids

The third European Grid Initiative (EGI) workshop was organized at CERN on 30 June, and was followed by the EGI Policy Board meeting on 1 July. More than 100 attendees took part in the CERN workshop and discussed the EGI blueprint proposal prepared by the EGI Design Study.

The purpose of this EGI blueprint draft is to assess a possible model for the future sustainable Grid infrastructure in Europe and to get feedback leading to the final blueprint that should be produced by September this year.

The director-general of CERN, Dr Robert Aymar, welcomed the participants, emphasizing the increasingly important role of Grids for science in Europe in many domains and in particular for CERN and its community. He underlined the importance of the ongoing and exceptional support given by the European Commission to European Grid projects. He wished success to the EGI project in building on these achievements by establishing a sustainable Grid infrastructure.

The organizer of the workshop, Jürgen Knobloch (CERN), expected feedback and critical questions from the national representatives and from Grid users, leading to a final blueprint in time for the next workshop in September. This will take place during the EGEE'08 conference in Istanbul, Turkey.

The EGI blueprint draft was presented to the audience by different members of the EGI Design Study team. Dieter Kranzlmüller (GUP) stressed in his overview the need for a sustainable European Grid infrastructure. Grid users need to be assured that their investment in Grids and long-term perspectives in the field will be protected. The EGI Grid infrastructure should be a large-scale production infrastructure built on national Grids that interoperate seamlessly at many levels. The infrastructure should offer reliable and predictable services to a wide range of applications.

Kranzlmüller also named the key players. The EGI will include both the EGI organization and National Grid Initiatives (NGIs), which are recognized national bodies that ensure operations of the Grid infrastructure in each country. Representation will include that the requirements of the scientific community and the resource providers are met.

The role of the EGI organization will be to facilitate interaction and collaboration between the NGIs and to provide a

The participants enjoy a break at the EGI Geneva workshop, which was held at CERN.
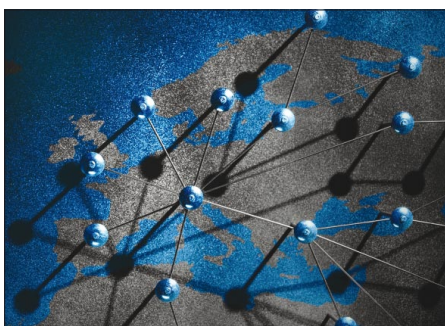
The main reach of the EGI Design Study.

common managerial framework for the pan-European Grid infrastructure. The EGI Council would be the sole governing and decision-making body of the EGI.

The EGI operations, middleware and user support questions, awoke vivid discussions. Tiziana Ferrari (INFN) introduced the EGI operations. They will include a set of services, such as the coordination of resource allocation, central repositories and ticketing, security and middleware roll-out. These services are needed to ensure the optimal functionality of the pan-European Grid infrastructure and overall seamless effective interoperation of national and regional Grids.

Ludek Matyska (CESNET) concentrated on the middleware evolution. He stated that the middleware is considered an essential part of EGI, and its existence and further development are of utmost importance for the EGI Grid.

The EGI blueprint proposes a common middleware solution based on the Universal Middleware Toolkit (UMT), with common policies, rules, quality and standard compliance criteria for UMT components. Patricia Mendez Lorenzo (CERN) introduced the EGI user-support section of the blueprint. She highlighted

the need to ensure a smooth transition to an NGI support infrastructure for users. The goal is to ensure that all current communities will continue to be supported and that the infrastructure will be ready to admit new communities.

Legal aspects and NGI guidelines for EGI were expanded by Anne-Claire Blanchard (CNRS). She presented plans for the location bidding process, which will be launched soon. Michael Wilson (STFC) made a presentation of the plans for EGI resources and finances.

It became clear that for the initial period, and to sustain innovation, co-funding by European sources will be crucial.

Bob Jones, the EGEE project director, presented some key issues for the transition from EGEE (Enabling Grids for E-sciencE) to EGI, seen from the EGEE side. He emphasized the need for co-operation with other infrastructure projects. EGI should also take into account the experiences and knowledge gained from the EGEE project.

The closed Policy Board (PB) meeting, chaired by Gaspar Barreira, followed the workshop, where a great number of NGIs were represented. Kyriakos Baxevanidis presented the European Commission's view on the future of e-infrastructures in Europe. Otherwise, the EGI PB discussion concentrated on the EGI blueprint proposal, focusing on EGI functionality, NGI responsibilities and the EGI funding model.

The next EGI workshop will be organized during the EGEE'08 conference in Istanbul, Turkey, on 22 September. The workshop will concentrate on planning the transition to the future EGI model (EGI Organisation and NGIs) and will be followed by an EGI Policy Board meeting.

Katja Rauhansalo, EGI-DS
www.eu-egi.eu

Information corner

Colloquia present future trends

CERN Computing Colloquia present future trends in computing and information technology that are of broad interest to the physics and computing community at CERN. Speakers are experts in their fields, whether from industry or academia. The colloquia complement the more technical IT seminars, which are also organized by the IT Department.

The list of future scheduled colloquia can be found on the InDiCo website by following Seminars and Courses/Seminars/CERN Computing Colloquium. This site also lists previous colloquia, along with slides and video recordings of the presentations, where these are available.

The next Computer Colloquium will be held on 26 September and will be presented by Rudi Noser, chairman of Noser Management AG and vice-president of the Swiss Liberal Free Democratic Party.

The talk is entitled "The challenges facing European software companies".

The colloquia are open to everyone and are free of charge. Those who are not CERN staff or users and would like to participate should contact the colloquium organizer, David Myers, to arrange access (e-mail: David.Myers@cern.ch?subject=re%20 Computing%20Colloquia).

Please contact the colloquium organizer if you have any ideas for future speakers.

Useful links
Indico Computing Colloquia category: http://indico.cern.ch/categoryDisplay. py?categId=167
Computing Colloquia website: http://cern. ch/it-dep/colloquia
IT Seminars website: http://cern.ch/ computing-seminars
David Myers, IT-DI

Bookshop plans another fair

Following the success of the first CERN Book Fair, the Bookshop is organizing another one on 8–10 October to give interested publishers and booksellers the opportunity to introduce themselves to CERN's resident and guest scientists.

The theme of this exhibition and the Book Fair will be "Books for leading scientists". The topics that will be covered include physics, mathematics, engineering and computing science.

During the Book Fair, exhibitors will also organize events relating to the celebration of scientific books during each of the three days, from 9.00 a.m. to 6.00 p.m. These might take the form of, for example:
● literature in focus event (whereby

CERN-connected authors present their book and their experience as a writer);
● author book-signing session;
● presentation of products;
● information about how to become an editor or author;
● meet-and-greet aperitifs.

Cambridge University Press, E-books Corporation, EPFL, Elsevier, Oxford University Press, Princeton University Press, Springer, Wiley and World Scientific are among the publishers who are expected to participate.

The fair will take place in the main building (Building 500). We look forward to your visit.
Eva Papp, CERN Bookshop

Calendar

September
22–26 **EGEE '08**
Istanbul, Turkey
http://egee08.eu-egee.org/

29 September – 1 October **Grid 2008**
Tsukuba, Japan
http://grid2008.org/

October
13–15 **8th Cracow Grid Workshop**
Krakow, Poland
www.cyfronet.pl/cgw08/

18–24 **HEPiX Fall Meeting 2008**
Taipei, Taiwan
http://indico.twgrid.org/
conferenceDisplay.py?confId=471

18–25 **2008 IEEE Nuclear Science Symposium NSS 2008**
Dresden, Germany
www.nss-mic.org/2008/

21–22 **e-IRG Workshop**
Paris, France
http://e-irg.eu/

22–24 **eChallenges e-2008**
Stockholm, Sweden
http://echallenges.org/e2008/

November
3–14 **Advanced School in High Performance and Grid Computing**
Trieste, Italy
http://agenda.ictp.it/smr.php71967

15–21 **SuperComputing08**
Austin, TX, US
http://sc08.supercomputing.org/

Look out for the October issue of CERN Courier, celebrating the start-up of the Large Hadron Collider.