

# CERN COMPUTER NEWSLETTER

Volume 43, issue 4 October–December 2008

## Contents

### Editorial

EDMS portal delivers better functionality 1  
CERN openlab welcomes Siemens as latest partner 2

### Announcements & news

New mailing list manager will take over from Simba 4  
User Registration Service closes 4  
Change your CERN passwords once a year 4  
CERN provides Vista training opportunity 4

### Grid news

LHC computing grid celebrates going operational 5  
US LHC supporters toast grid success via satellite 5

### Technical brief

AFS revisited: volume snapshots enable the retrieval of lost data 6  
Parallel coordinates technique implemented in ROOT framework 7  
J2EE Public Service implements security and usability upgrade 9

### Conference and event reports

CSC2008 revisits Norway 10  
Istanbul hosts EGEE conference 11  
EGEE to EGI: we must not be afraid of the future 11

### Information corner

Phoney messages target CERN 12  
Next generation of CERN Linux 12  
Calendar 12

**Editor** Natalie Pocock, CERN IT Department, 1211 Geneva 23, Switzerland. E-mail [cnl.editor@cern.ch](mailto:cnl.editor@cern.ch). Fax +41 (22) 766 8500. Web [cerncourier.com/articles/cnl](http://cerncourier.com/articles/cnl).

**Advisory board** Wolfgang von Rueden (head of IT Department), Alberto Pace (group leader, Data Management), Christine Sutton (*CERN Courier* editor), Tim Smith (group leader, User and Document Services).

**Produced for CERN by IOP Publishing** Dirac House, Temple Back, Bristol BS1 6BE, UK. Tel +44 (0)117 929 7481. E-mail [jo.nicholas@iop.org](mailto:jo.nicholas@iop.org). Fax +44 (0)117 930 0733. Web [iop.org](http://iop.org).

**Published by CERN IT Department**

©2008 CERN

The contents of this newsletter do not necessarily represent the views of CERN management.

**IOP Publishing**



## EDMS portal delivers better functionality



The EDMS team has redesigned its home page and improved search functionality.

The CERN Engineering & Equipment Data Management Service (EDMS) was launched in 1997 with the goal of providing the LHC project with an advanced information system together with industrial best practice, for managing engineering information throughout the project's entire life cycle.

Since 1997 the service has grown significantly in terms of scope, offered functionality and number of users; today it serves the whole organization including experiments, projects and collaborations. Currently there are more than 5700 registered users and the system manages more than 900 000 documents and drawings as well as information about almost 700 000 installed components.

This growth prompted the EDMS team to redesign the EDMS home page, creating a new service portal to facilitate access to the system's different modules and the increasing volume of information stored in the EDMS system.

In October the EDMS team launched a new EDMS portal offering EDMS users a faster and improved search functionality for any object, quicker navigation, a "My EDMS top 10" and categorized menus for access to other modules.

The new search functionality permits a variety of search options such as searching on objects (documents, items, equipment

etc.) using a combination of text and status information, e.g. you can search on "your name released". The contextual category menu can be used to search and access other EDMS modules such as the Manufacturing and Test Folder (MTF) and CERN Drawing Directory (CDD).

The "My EDMS top 10" is a new addition to the EDMS portal. This feature is a subtle combination of a user's personal, most recently and most frequently accessed objects. Rather like bookmarks, this functionality allows users to easily access documents, structures and equipment that they have been working on at that particular moment in time.

Users will also note that fewer clicks are required for searching and for logging in: The cursor is positioned directly in the "Username" field when you reach the portal, allowing for immediate login.

The team is also launching the SharePoint-based EDMS website <http://cern.ch/edms-services/>, where users can find useful information about the service, training sessions as well as FAQs and tutorials related to EDMS. In conjunction with the introduction of the EDMS portal, we invite users to consult the EDMS portal FAQs, and in particular: <http://cern.ch/edms-services/faq/EDMS Portal/>.

**Rachel Bray and David Widegren, EDMS team, TS Department**

# CERN openlab welcomes Siemens as latest partner

CERN has been joining forces with industrial partners for six years in a unique framework for development and prototyping activities, as well as for the evaluation of potential computing solutions for the LHC scientific community.

This initiative, CERN openlab, has been organized into successive three-year phases. In openlab-I (2003–2005) the focus was on the development of an advanced prototype called opencluster, and openlab-II (2006–2008) addressed a range of domains from platforms, databases and interoperability to security and networking. We are now entering the third phase – openlab-III (2009–2011) – which will not only capitalize on and extend the successful work carried out in openlab-II, but it will also tackle new crucial areas.

Siemens is an industrial partner that will collaborate with CERN in openlab-III in one of these new areas: the domain of industrial controls. Some European companies have previously been members of CERN openlab as contributors, but Siemens is the first European company to join as a full openlab partner. In addition to in-kind contributions, Siemens will fund three fellows and one staff position.

## Control systems at CERN

At CERN, control systems are deployed for accelerators, experiments and infrastructure. The role of the experiments' control systems is to bring experiments to, and maintain them at the correct conditions to take physics data. The accelerator control systems are used to provide beams with the expected and required luminosity for the experiments.

As illustrated in figure 1, a CERN control system is typically made up of several layers. Various technologies and tools are available to implement these layers. In

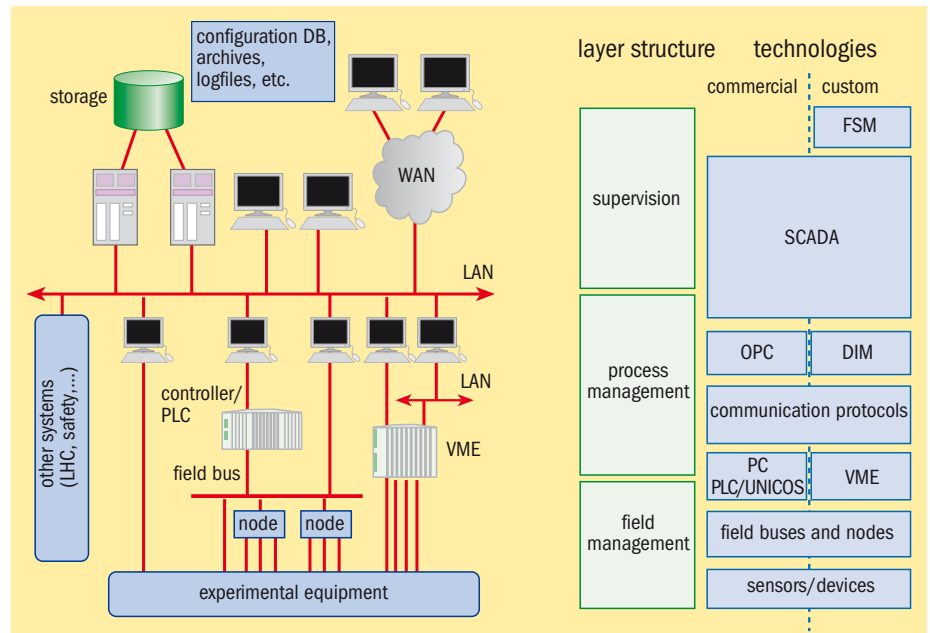


Fig. 1. The architecture of a typical CERN control system is made of several layers. Siemens components are mainly in two of them: SCADA systems in the supervision layer and Programmable Logic Controllers (PLCs) in the process management layer.

particular, Siemens provides components for the process layer – Programmable Logic Controllers (PLCs) – and for the supervision layer – Supervisory Control And Data Acquisition (SCADA) systems.

CERN has selected PVSS, the SCADA system from ETM (which is a Siemens company), for its control system. The PVSS architecture is described in figure 2.

Like any SCADA system, PVSS provides facilities to acquire, display and archive control data, and to handle alarms and events. PVSS also has some additional benefits that make it attractive to CERN. Unlike most SCADA systems, PVSS is highly scalable; a PVSS system is composed of

managers that can be distributed over many computers when the control applications grow, and, if necessary, several PVSS systems can be federated to implement one large control application. About 150 PVSS systems are required to implement a typical large LHC experiment control system.

PVSS is a particularly open SCADA system; functionalities can easily be added to PVSS either by using the script language or the API to integrate externally developed components. Finally, PVSS applications can be deployed on both Windows and Linux machines.

PVSS applications are currently being used for the LHC machine, LHC experiments

The deadline for submissions to the next issue of CNL is

# 20 February

Please e-mail your contributions to [cnl.editor@cern.ch](mailto:cnl.editor@cern.ch)

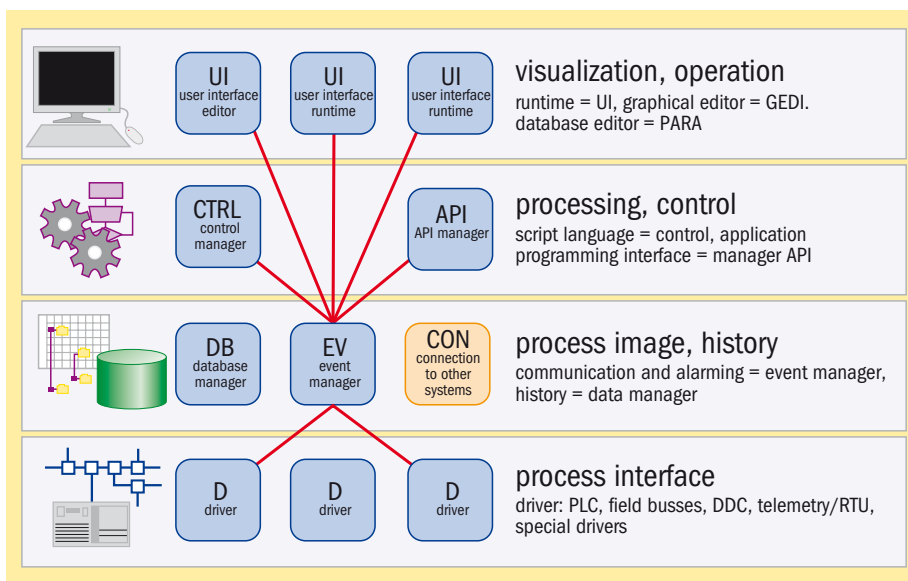


Fig. 2. This schematic illustrates the PVSS architecture. PVSS provides facilities to acquire, display and archive control data, and to handle alarms and events.

as well as fixed targets experiments. About 500 licences have been generated.

Siemens is also one of the two PLC manufacturers selected for use at CERN. PLCs are dedicated diskless computers for control. These computers are robust and programmed with specific standardized programming languages. They are used to implement input/output accesses, interlocks, close control loops and automatic action-taking systems (e.g. by means of Finite State Machine programs). About 500 Siemens PLCs (S7-400, S7-300 and S7-200 models) are deployed across CERN in accelerators, experiments and infrastructure control applications.

#### openlab and Siemens collaboration

Siemens and CERN will collaborate on three main topics: security; opening of automation tools towards software engineering; and handling of large environments.

#### Security

Due to the growing usage of Ethernet and TCP/IP in automation devices, and because of the move away from proprietary or dedicated networks, the automation devices and control applications have

to become resistant to the common threats on Ethernet cable. These threats can be deliberate (attackers), collateral (viruses and worms), or accidental (misconfiguration, such as an error in the IP address or broken devices flooding the network).

Siemens and CERN have an inherent interest that automation devices (e.g. PLCs) survive these kinds of attacks and they will be investigating the resistance of the devices. More specifically, resistance to malicious network traffic will be analysed through robustness tests, proving that a device can withstand for example ARP and BOOTP floods, and vulnerability tests, proving that a device is not affected by common viruses, worms or known vulnerabilities of the services, protocols and TCP/IP stacks used in these devices. There is a similar interest as far as PVSS-based control applications are concerned.

#### Software engineering

CERN control systems (like many others in industry) will be operated and maintained for a very long period of time (about 15 years). Thus, the PVSS-based control

applications and the user programs running in the corresponding PLCs have to follow the same cycle. During this time they have to be managed, maintained, adapted and extended – not necessarily by the same developer(s). In addition, the increasing complexity of the control systems sometimes leads to the production of the PVSS and PLC code with the help of external tools (e.g. model-driven ones).

Siemens and CERN are interested in following the convergence trends between the automation and the information technology worlds. For example, it would be beneficial if the Siemens tools were able to use external (open source or third party) tools to handle proper source code management (e.g. CVS, subversion), code production and analysis (tools highlighting the difference between two versions of code written in graphical language).

#### Large environment

PVSS is used at CERN for large distributed control systems, some with more than 150 computers. PVSS and the controls applications have to be deployed on these computers. This requires an initial installation as well as regular upgrades and patches for both PVSS software and control applications. PVSS does not offer any native facility to deal with this type of environment.

The tools developed at CERN for software deployment – CERN Management Framework (CMF) and Linux for Controls – are only appropriate for initial PVSS installation. The installation of the applications (software + configuration) has to be handled manually. This is also true for the PLC layer. CERN and Siemens will collaborate to build adequate deployment solutions for the supervision and process control layers of large control applications.

#### Useful links

CERN openlab: <http://cern.ch/openlab>

PVSS at CERN: <http://cern.ch/itcobe/Services/Pvss/>

Siemens at CERN: <http://cern.ch/ab-dep-co-is/Siemens/>

Introduction to PLC: <http://cern.ch/itcofe/Services/PLC/WhatIsPLC/>

**Renaud Barillère, IT-CO**

**If you want to be informed by e-mail when a new CNL is available, subscribe to the mailing list [cern-cn1-info](mailto:cern-cn1-info).**

**You can do this from the CERN CNL website at <http://cern.ch/cnl>.**

# New mailing list manager will take over from Simba

Starting at the end of November a new interface called “e-groups”, which manages mailing lists, will be introduced progressively, eventually replacing the current “Simba” application.

Last year e-groups replaced Adressage for the management of dynamic mailing lists, i.e. mailing lists dynamically populated from a set of criteria such as age, nationality or membership of a group, section or experiment. However, owners of dynamic lists still had to use Simba to manage some of the properties such as posting

restrictions. From the end of November, all the properties of dynamic lists will be available from the e-groups application, so Simba won't be used for these lists anymore.

In addition, while providing the same base functionality, e-groups will introduce a number of improvements over Simba:

- all changes will be applied immediately; in particular, when a new list is created, it can be used immediately;
- a new archiving system that organizes messages in discussion threads;
- significantly improved search and

navigation features in the archives.

Attachments will also be searchable;

- programmatic interfaces to create and manage lists.

As a result, all existing Simba mailing lists (including non-dynamic lists) will be transferred over time to the e-groups application. Eventually, Simba will be completely replaced by e-groups.

Information about the migration schedule and documentation is available from the Simba website <http://simba.cern.ch>.

**CERN Mail Services**

## User Registration Service closes

Words like Wylbur, MVS, CernVM, Cray X-MP, CernSP2 and Shift might not mean much to some of our younger colleagues at CERN, but to many they are evocative of the recent history of CERN computing: Wylbur and MVS in the early 1980s (when the current group codes were invented) running on IBM360s; CernVM, the first real interactive service on the IBM3090s; followed by the move to Unix with the Shift service for the LEP experiments on SGI, HP machines and others; and the CernSP2, AIX service for the rest of CERN users. Not to be forgotten is the short passage of the then star of the computer centre, the Cray X-MP.

All these services needed users registered and accounts created. In the early days this was done using flat files, then with the arrival of Oracle the Computer Centre DataBase (CCDB) was created and managed the

accounts for many years. CCDB was followed by Computing Resources Administration (CRA) for resource management.

For a number of years the dedicated User Registration Service behind CCDB and CRA has been manned by Marie-Claire Perler. Always polite, efficient and competent with users, Perler's incomparable ability to find bugs in both CCDB and CRA was much appreciated by developers too. As anyone who has worked with Marie-Claire can confirm, she was always available to solve the most intricate problems with accounts, e-mail addresses and all the other information that overwhelmed even the most experienced group administrator. In fact, being so efficient and helpful sometimes worked against Perler as users and group administrators knew that the answer was at the other end of the phone, which was quicker than having to search for it themselves!

In October Marie-Claire will be leaving CERN and we would like to thank her and wish her all the best for her retirement. Her departure marks the end of the dedicated User Registration Service.

Like many services, from banking to airline ticket reservations, users and group administrators will be expected to manage more of the user details themselves. Steps have been taken to ensure services still run smoothly: group administrators received a guide to CRA, which is available online via the group admin menu in CRA, and specific group administrator training was organized.

From 21 October, user registration e-mails have been redirected to the help desk, which will answer a number of questions, but more importantly assist users by explaining how to do things themselves and where to find the necessary information. Requests that need special privileges will be dealt with by the AIS 2nd level support and the IT Manager on Duty. However, please note these are not dedicated lines of support, so users and group administrators are advised to make full use of the CRA guide and training available.

### Useful links

CRA: <https://cra.cern.ch>; CRA FAQ: [http://cern.ch/ais/projs/cra/tutorial/faq\\_e.htm](http://cern.ch/ais/projs/cra/tutorial/faq_e.htm).

**Nick Ziogas, IT-UDS**

## Change your CERN passwords once a year

CERN users are reminded that they MUST change their CERN password once a year. This is important and easy to do with password synchronization, which gives access to most of the central services.

Several reminders are sent to the mailbox attached to the address “@cern.ch” (which can be a CERN or external mailbox). The mailbox can also be linked to a service account, whose password should also be changed once a year by the account owner.

It seems that some users don't pay enough attention to the mails inviting

them to change their password, or possibly never access their CERN mailbox. However, if your password is not changed in time your account will be blocked (which can occur over the weekend). So to avoid this, remember to change your password at least once a year at: <https://cern.ch/cernaccount>.

Users who have a CERN mailbox but don't regularly access it should ask their group administrator to delete it and set their PEM (physical e-Mail address) to a mailbox they do use regularly.

**Nicole Crémel, IT-UDS**

## CERN provides Vista training opportunity

Are you running the Vista operating system on your PC, or planning to install it? The Technical Training Service is organizing half-day sessions on the new features of Vista. The course introduces the new interfaces and functionalities, and is designed for people experienced in using Windows XP. The next bilingual sessions of this course will be on 12 December and 30 January.

Register using our catalogue at <http://cta.cern.ch/cta2/f?p=110:9>, or e-mail us with any questions or comments at [technical.training@cern.ch](mailto:technical.training@cern.ch).

## LHC computing grid celebrates going operational

On 3 October, grid enthusiasts gathered – physically and virtually – from around the world to celebrate the success of the Worldwide Large Hadron Collider (LHC) Computing Grid.

“After the very successful start of the Large Hadron Collider on 10 September, it is now the turn of the Worldwide LHC Computing Grid to celebrate the transition from a development and deployment phase into operation,” said Wolfgang von Rüden, head of CERN’s Information Technology department, in his welcome address to the 250 grid enthusiasts gathered in CERN’s globe building. “I am very happy to tell you that we are ready to handle the unprecedented flood of data that will be generated by the LHC.”

In his address entitled “CERN and LHC: their place in global science” CERN’s director-general Robert Aymar, highlighted the necessity of computing for studying particle physics: “This is just the first step in a long voyage. The LHC is a discovery machine, and its research programme stretches out two decades into the future,” he said. “There are three needed tools to study physics. You need accelerators,



*Wolfgang von Rüden unveils the ceremonial globe sculpture before an audience in CERN’s globe-shaped building. Image courtesy of GridFest.*

detectors and computing. It is the computing that turns the flow of data into useful information.”

While physics was the star scientific discipline of the day, attendees also witnessed the importance of a range of applications from a host of disciplines. “The great thing about grids is that they have many uses,” said Bob Jones, project director of Enabling Grids for E-science, in his talk “The grid beyond physics.” He discussed applications in seismology, atmospheric research, astronomy, fusion and the life sciences.

On-site demonstrations, held throughout the day, included the Worldwide LHC Computing Grid, the ALICE experiment, the ATLAS experiment, the CMS experiment, the LHCb experiment, the Health-e-Child project (pediatrics), the ITER project (fusion energy), Open Science Grid and the WISDOM project (drug discovery).

You can see many of these demonstrations on *YouTube*, posted by the GridTalk project, and at <http://cern.ch/LCG/lhcgridfest>.

**Danielle Venton, EGEE**

• This article was published online in *iSGTW* on 8 October.

## US LHC supporters toast grid success via satellite

Three weeks after the US Large Hadron Collider (LHC) community celebrated in the small hours as the first beam circulated in the LHC across the Atlantic, staff and supporters gathered again in Fermilab’s signature high-rise. Via satellite, they co-celebrated the official start up of the World-wide LHC Computing Grid (WLCG) with their colleagues in the US and abroad.

As the US Tier-1 site for the CMS experiment, Fermilab has contributed significantly to the grid development for the experiment, and is an important Open Science Grid (OSG) site, in terms of resources and expertise. The same is true of Brookhaven National Lab with respect to the ATLAS experiment. The OSG supports the WLCG computing needs in the US.

Patricia McBride, deputy head of Fermilab’s computing division, recognized the contributions of the scientists and computing professionals who worked on the grid: “A decade ago, a global computing centre for data analysis by scientists collaborating on the four LHC experiments was an idea, a dream. To all of you who shared in this dream and worked hard to deploy (and debug) FermiGrid, Open Science Grid, and the World-wide LHC Computing Grid, I say congratulations and thank you,” she said. FermiGrid refers to



*Speakers were invited to a post-GridFest tour of Fermilab’s grid computing centre. L to r: Kevin Thompson, Patricia McBride, Lothar Bauerdick, head of US-CMS software and computing, and Don Petravick. Image courtesy of Reidar Hahn, Fermilab.*

Fermilab’s campus grid, which provides the lab’s interface to the OSG.

Michael Ernst, director of the RHIC and ATLAS computing facility at Brookhaven, acknowledged the years of work behind this achievement and the “lively collaboration among some of the best universities and institutions in the US” that made it possible.

Miron Livny, professor of computer science at the University of Wisconsin in Madison and principal investigator of the Open Science Grid, reasserted the OSG’s commitment to the WLCG to providing know-how as well as resources. The OSG, he said, will advance both software tools

and distributed computing understanding, and will have a significant impact on other sciences as well.

The OSG is already serving other sciences. Andrew Schultz, a chemical and biological engineering researcher from SUNY in Buffalo, spoke about his high-accuracy simulations of molecular interactions in gases run on the OSG. The OSG accelerated his research by about a factor of 10 in the last few months, he said.

Steve Cotter of ESNNet discussed the networking advancements that have enabled the WLCG. Kevin Thompson of the NSF’s Office of Cyberinfrastructure acknowledged in particular the US LHC Tier-2 centres, and praised the WLCG as an exemplar of the office’s vision of international collaboration.

Susan Turnbull, program manager, ASCR, Department of Energy (DOE), thanked the local and remote audiences on behalf of the DOE’s Office of Science for their “steadfast courage, vision, and leadership in conceiving and building the World-wide LHC Computing Grid.”

“While you’re blazing a path forward with your built-to-share scientific discovery infrastructure, the world is also watching and learning from you how global collaboration with expansive purpose can span traditional boundaries and multiply the creativity of our differences,” she said.

**Anne Heavey, iSGTW**

• This article was published online in *iSGTW* on 8 October.

## AFS revisited: volume snapshots enable the retrieval of lost data

These “AFS revisited” articles are intended to act as short reminders. Here we look at standard AFS commands, common pitfalls and some tips and tricks to get more out of AFS at CERN.

AFS stores all files and directories in logical containers called volumes, which are mounted into the AFS file hierarchy. From the service’s manageability point of view, this additional internal abstraction layer is the key feature that allows required service interventions to be completely transparent to the user.

A direct benefit for the user is AFS’s ability to create snapshots of volumes, i.e. consistent images of a volume’s data at a certain moment in time. Slightly misnamed, these snapshot volumes are referred to as back-up volumes. Strictly speaking however, these read-only containers do not provide real back-ups, but resemble hard links: only pointers to the real data are copied when a back-up volume is created. Changes to the original volume trigger a copy-on-write. Nonetheless, back-up volumes can be useful in retrieving data that have been changed or deleted in the original volume.

### User home volumes

At CERN, back-up snapshots of the users’ home volumes are taken daily in the late afternoon and are then available via `/afs/cern.ch/ubackup/<first letter of login name>/<login name>`. Our example user Pam would hence find her back-up volume mounted under:

```
/afs/cern.ch/ubackup/p/pam
```

Using this path is a convenient and quick way to access files that have been accidentally deleted or modified since the last snapshot was taken. To find out when the last snapshot of your home volume was taken run:

```
$ cd $HOME
$ /usr/bin/fs examine
```

The output will contain a line that includes the name of your home volume. These are typically named `user.<login name>`. Examine that volume with

```
$ /usr/sbin/vos examine <volume name>
```

Look for a line starting with *Backup*. This line contains the desired timestamp.

The data as contained in the back-up

volumes are also the basis for the real back-ups which are stored on tape. In case one needs to recover data that were deleted or modified before the last snapshot, the data can only be recovered from tape by means of the `afs_admin` command:

```
$ afs_admin recover <path to restore>
```

Depending on the activity on the tape system, restoring a file this way can be very slow. So, some patience may be required here. Once the data are recovered, the caller is put into a shell where the current working directory is the top level directory of the back-up volume. Hence, to copy a lost file called *alliance/operations.txt* back to a user’s home directory one has to type

```
$ cp -pf ./alliance/operations.txt $HOME/
alliance/operations.txt
```

`afs_admin` creates a log file in `/tmp` which should be sent to `afs.support@cern.ch` in case there are problems with the recovery.

Note that back-ups are stored on tape for approximately one year.

### Project and scratch volumes

As well as the user home volumes, there are two other main classes of volumes: project volumes and scratch volumes. Project volumes are backed up in a similar way to user home volumes. However, there is no automatic mount point creation to access their back-up volume. By default, data are recovered using the very same `afs_admin` command as above.

When the data to be recovered have been deleted or modified since the last snapshot was taken, the following recipe describes a shortcut to save some time by bypassing a recovery from tape:

- Find out if the lost data are backed up, do so by:

```
$ /usr/bin/fs examine <path of lost data>
```

This will return the name of the underlying volume in a line similar to

```
Volume status for vid = 537163838 named
p.alliance
```

Check if this volume is backed up by:

```
$ /usr/sbin/vos examine p.alliance
```

The output of this command should contain

a line like:

```
Backup Mon Sep 29 18:52:58 2008
```

If the *Backup* line contains *Never* instead of a date, the data has not been backed up and you can stop here.

- Mount the back-up volume into your AFS space, do so by:

```
$ cd $HOME
$ /usr/bin/fs mkmount -dir <directory>
-vol <volume name>.backup
```

In our example above, the second line could read:

```
$ /usr/bin/fs mkmount -dir alliance.old
-vol p.alliance.backup
```

Note that we added the extension `.backup` to the volume name. By convention the back-up volumes have the same name as their master copies with this extension.

- Copy your files.
- Remove the mount point again, do so by:

```
$ /usr/bin/fs rmmount -dir <directory>
```

In contrast with project volumes, scratch volumes are not backed up at all (a *vos examine* will show that immediately). Hence, there is no way to restore a file that has been deleted from scratch space!

### A common pitfall

If the lost data are not stored in your home volume, but in a volume only mounted into your home volume, the simple approach of checking `/afs/cern.ch/ubackup/...` does not work. A typical example for such a situation would be a workspace volume mounted under `$HOME/w0`. The underlying reason is that the mount point in the back-up volume does not point to the back-up version of your workspace volume, but to the original. To access the back-up volume of your workspace volume, you have to explicitly mount it, which can be done by the same recipe as outlined for project volumes above.

### Additional documentation

The openAFS website provides information on all AFS commands: <http://www.openafs.org/manpages>.

The CERN AFS User Guide is a much more comprehensive manual, available at: <http://cern.ch/service/afs>.

**Arne Wiebalck, IT-FIO**

# Parallel coordinates technique implemented in ROOT framework

The ROOT's trees provide many tools to handle multiple variables data sets: statistical analysis, cuts, variable combinations etc., but it was poor on the visualization side where only a maximum of four variables could be represented simultaneously on the same plot (a 3D scatter plot with the 4th variable mapped on a colour palette). Various techniques exist to visualize many variables and several have recently been implemented in the ROOT context. This article presents the parallel coordinates plots.

The parallel coordinates plots are a common way of studying and visualizing multiple variables data sets. They were proposed by A Inselberg in 1981 as a way to represent multidimensional information.

In traditional Cartesian coordinates, axes are mutually perpendicular. In parallel coordinates, all axes are parallel which means data can be represented in more than three dimensions.

To show a set of points in parallel coordinates, a set of parallel lines is drawn, typically vertical and equally spaced. A point in n-dimensional space is represented as a polyline with vertices on the parallel axes. The position of the vertex on the *i*th axis corresponds to the *i*th co-ordinate of the point. Figures 1 and 2 show some simple examples.

The parallel coordinates technique is useful for spotting irregular events, seeing a data trend, and finding correlations and clusters. Its main weakness is output cluttering: as each "point" in the multidimensional space is represented as a line, the output quickly becomes opaque and therefore it can be very difficult to see data clusters.

Most of the work done on parallel coordinates centres on finding techniques to reduce the output cluttering.

The parallel coordinates plots in ROOT have been implemented as a new plotting option *PARA* in the *TTree::Draw()* method. To demonstrate how parallel coordinates works in ROOT we will use the tree produced by the "pseudo C++" code shown in the box below right.

The data set generated has:

- 9 variables: *x, y, z, u, v, w, a, b, c*
- $3000 \times 8 = 24\,000$  events
- three sets of random points distributed on spheres: *s1, s2, s3*
- random values (noise): *rnd*
- The variables *a, b, c* are almost completely random. The variables *a* and *c* are correlated via the 1st and 3rd coordinates of the 3rd "sphere" *s3*.

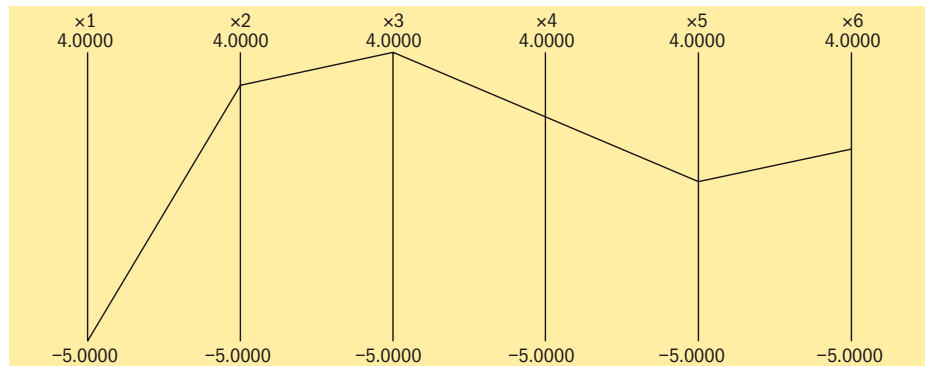


Fig. 1. A representation of a six-dimensional point (-5,3,4,2,0,1) using parallel coordinates.

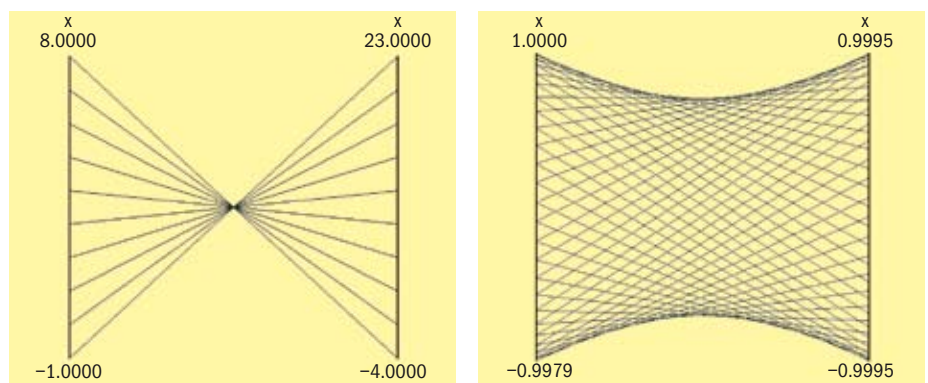


Fig. 2. A representation of  $y = -3x + 20$  (left) and a circle (right) using parallel coordinates.

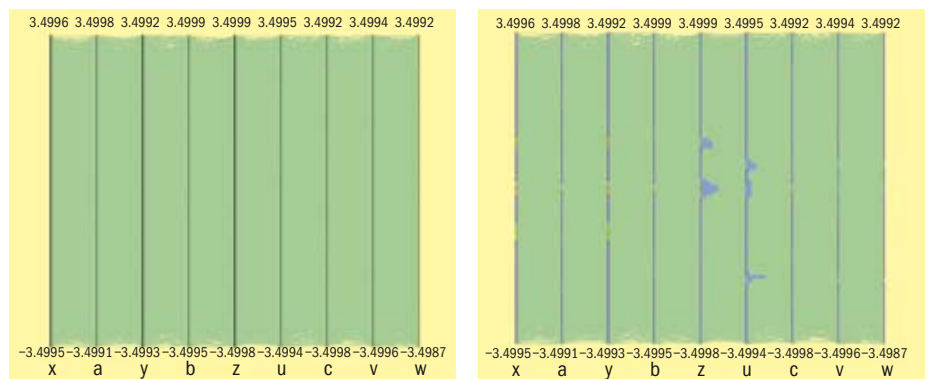


Fig. 3. A cluttered output is produced when all of the tree events are plotted.

Fig. 4. The axis on the histogram above can be represented using colour or a bar chart.

```
TNtuple *nt = new TNtuple("nt","Demo ntuple","x:y:z:u:v:w:a:b:c");
for (Int_t i=0; i<3000; i++) {
    nt->Fill( rnd, rnd, rnd, rnd, rnd, rnd, rnd );
    nt->Fill( s1x, s1y, s1z, s2x, s2y, s2z, rnd, rnd, rnd );
    nt->Fill( rnd, rnd, rnd, rnd, rnd, rnd, rnd, s3y, rnd );
    nt->Fill( s2x-1, s2y-1, s2z, s1x+.5, s1y+.5, s1z+.5, rnd, rnd, rnd );
    nt->Fill( rnd, rnd, rnd, rnd, rnd, rnd, rnd, rnd, rnd );
    nt->Fill( s1x+1, s1y+1, s1z+1, s3x-2, s3y-2, s3z-2, rnd, rnd, rnd );
    nt->Fill( rnd, rnd, rnd, rnd, rnd, rnd, s3x, rnd, s3z );
    nt->Fill( rnd, rnd, rnd, rnd, rnd, rnd, rnd, rnd, rnd );}
```

# Technical brief

The command used to produce the parallel coordinates plot is:  
`nt->Draw("x:a:y:b:z:u:c:v:w", "", "PARA");`  
If the 24000 events are plotted as solid lines and no special techniques are used to clarify the picture, the result can be cluttered and useless (figure 3). To improve the readability of the parallel coordinates output and to explore interactively the data set, many techniques are available. We have implemented a few in ROOT.

First of all, to show better where the clusters on the various axes are, a 1D histogram is associated with each axis. These histograms (one per axis) are filled according to the number of lines passing through the bins.

These histograms can be represented with colours (from a palette according to the bin contents) or as bar charts. Both representations can be compiled on the same plot (figure 4). This technique shows where the clusters on an individual axis are but it does not give any hint about the correlations between the axes.

We have implemented a simple technique to make the clusters stand out: instead of painting solid lines, we draw dotted lines. The cluttering of each individual line is reduced and the clusters can clearly be seen (figure 5). The spacing between the dots is a parameter that can be adjusted to get the best results.

Interactivity is an important aspect of parallel coordinates plots. To really explore the data set it is essential to manipulate events and axes directly. For example, changing the order of axes may show clusters that were not visible before (figure 6). We can see that many more clusters appear and all the "random spheres" we put in the data set are now clearly visible. Having moved the variables *u*, *v*, and *w* after the variables *x*, *y* and *z* the correlation between these two sets of variables is also clear.

To enable further exploration of the data sets, we have implemented the possibility of defining selections interactively. A selection is a set of ranges combined together. Within a selection, ranges along the same axis are combined with logical OR, and ranges on different axes with logical AND. A selection is then displayed on top of the complete data set using its own colour and only the events fulfilling the selection criteria (ranges) are displayed. Ranges are defined interactively using cursors, like on the first axis of figure 7. Several selections can be defined at the same time, each selection having its own colour. On figure 8 each cluster is now clearly visible and the zone with crossing clusters is now understandable.

Interactive selections on parallel coordinates are a powerful tool because they can be defined graphically on many variables (graphical cuts in ROOT can be defined on two variables only), giving

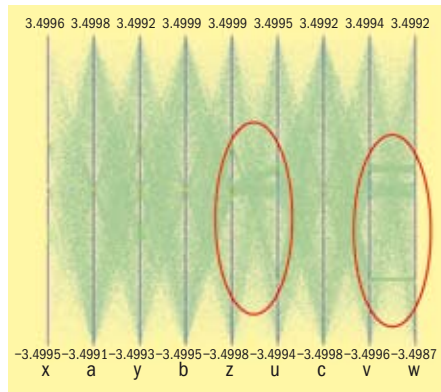


Fig. 5. Using dotted lines is a simple method that can reduce cluttering.

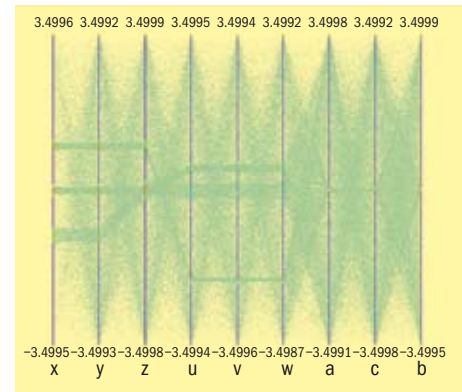


Fig. 6. Axis order is also important in highlighting data clusters.

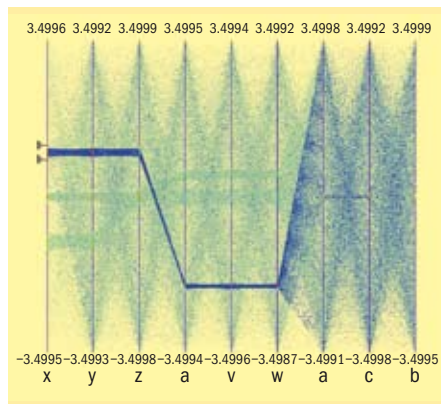


Fig. 7. Selections are set of ranges that can be defined interactively.

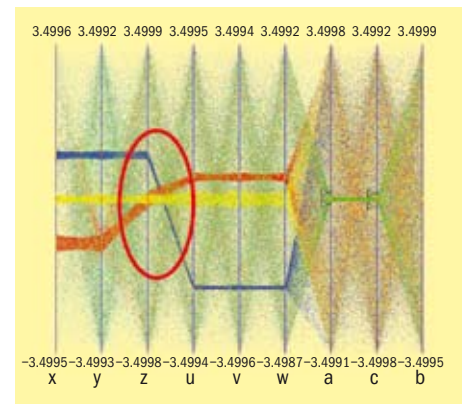


Fig. 8. Several selections can be defined each of them having its own colour.

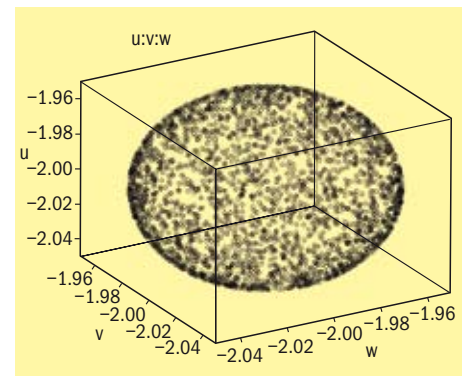
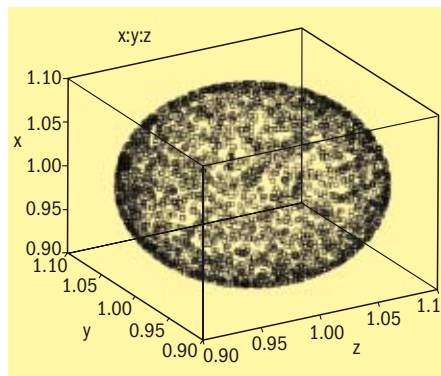


Fig. 9. Output of `nt->Draw("x:y:z")` (left) and `nt->Draw("u:v:w")` (right) after a selection.

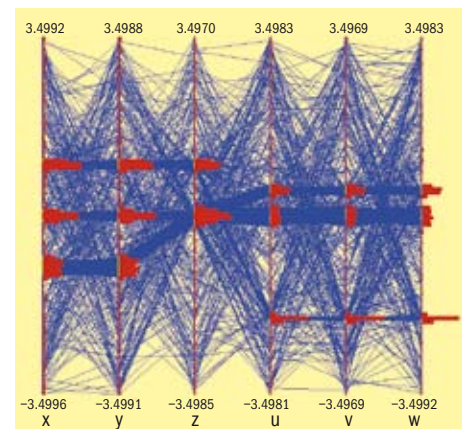
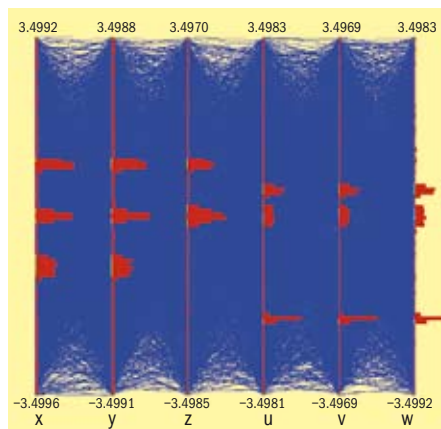


Fig. 10. Before (left) and after (right): applying a "weight cut" highlights data clusters.



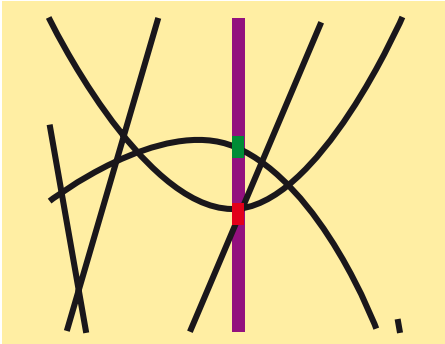
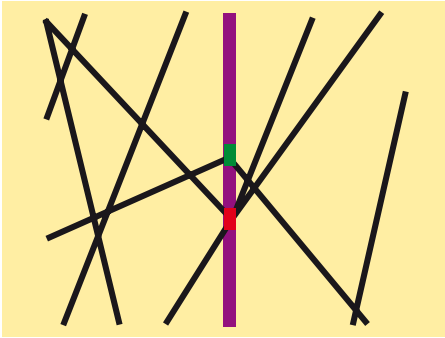


Fig. 11. By drawing the events as curves, this can differentiate them even further.

accurate events filtering. For example a single outlying event is clearly visible when the lines are solid, therefore it is easy to make cuts to eliminate a single event from a selection. Such selection (to filter a single event) on a scatter plot would be much

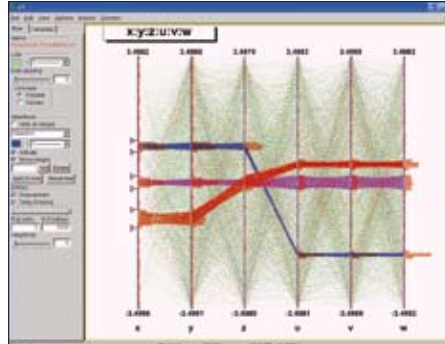
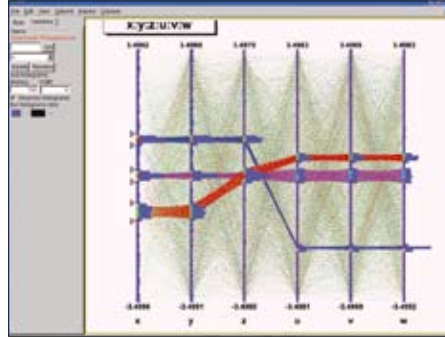


Fig. 12. A graphical user interface is vital in manipulating parallel co-ordinates plots.

more difficult.

Once a selection has been defined it is possible to use it to generate a *TEntryList* which is applied to the tree and used at drawing time. In our example the selection we defined allows us to select exactly

the two correlated “random spheres” as shown in figure 9.

Another technique has been implemented to show clusters when the picture is cluttered. A weight is assigned to each event, which is computed as:

Where:

- $b_i$  is the content of bin crossed by the

$$weight = \sum_{i=1}^n b_i$$

event on the  $i$ th axis

- and  $n$  is the number of axes.

The events with bigger weights are those belonging to clusters. It is possible to paint only the events having a weight above a given value and the clusters appear. In figure 10 the “weight cut” of 50 was applied to the plot on the right. If only a few events are displayed, drawing them as smooth curves instead of straight lines helps to differentiate them as shown in figure 11.

Interactivity and therefore the graphical user interface (GUI) is important in manipulating parallel coordinates plots. The ROOT framework provides easy ways to implement direct interactions on the graphical area and the graphical editor facility provides a dedicated GUI.

#### Useful links

CHEP’07 paper: [www.iop.org/EJ/article/1742-6596/119/4/042007/jpconf8\\_119\\_042007.pdf](http://www.iop.org/EJ/article/1742-6596/119/4/042007/jpconf8_119_042007.pdf).

ROOT: <http://root.cern.ch/>

**Olivier Couet, PH-SFT**

# J2EE Public Service implements security and usability upgrade

The J2EE Public Service is the CERN central server infrastructure for the deployment of JAVA Web applications. This year, some new key features have been implemented to improve the security and usability of the service.

#### Security: CERN Authentication

CERN Authentication is the IT standard single sign on (SSO) solution for CERN Web-based applications (see <https://cern.ch/authentication>). It facilitates the login process and increases security by relying on a single login/password pair, a certificate or Windows local machine credentials. CERN SSO has the benefit of allowing the user to authenticate once and gain access to the resources of multiple software systems (including Indico, Twiki, NICE services, among others).

As a first step, in July the JPSManager (J2EE Public Service Manager JAVA Web application) adopted CERN Authentication.

This was transparent to the users as they were able to continue authenticating using their CERN login.

Since October, managers of JAVA Web applications have been able to request CERN Authentication. The aim is to replace the currently deployed authentication mechanism (JNDIRealm) with CERN Authentication at the container level. Authorization will continue to be based on the NICE Group Manager.

#### Usability: permanent storage solution relying on AFS

When the J2EE Public Service was first launched, it had the limitation that no permanent files belonging to owners of applications were allowed to be stored on service machines (other than the .war files themselves), only temporary files are allowed. However, some applications need to store and modify files (e.g. configuration files, generated graphs or reports and files

uploaded by users). Application owners need read-and-write access to those files, and the files need to be backed up.

A solution using a distributed file system (DFS) library that uses WebDAV to connect to the Windows DFS was implemented last year. Unfortunately this solution requires modifications to the source code of the Web application.

From September, an alternative solution using AFS has been in production. The service provides on-demand, dedicated AFS space to the JAVA Web application. Full access to this AFS location is granted to the owner and to the application itself. This solution has the advantage that it is transparent to developers as the JAVA Web application doesn’t need any modifications to the source code.

If you would you like to know more, visit the service Web page at <http://cern.ch/j2ee-public-service/>.

**Lucia Moreno, IT-DES**

# CSC2008 revisits Norway

Since it started 37 years ago, CERN School of Computing (CSC) has taught 1947 students and taken place in 18 countries. It was 34 years since the school was last held in Norway and we were happy to return there, this time to Gjøvik, 40 km south-east of Lillehammer, which hosted the 1994 Olympic Games.

The school took place from 25 August to 5 September and welcomed 52 participants. The scientific programme consisted of 50 hours of lectures and exercises organized around three themes and was delivered by an international team of 12 lecturers.

CSC2008 was the first school that took place after the four-year period of EU support via the Marie-Curie project, which provided partial grants to all participants from 2004 to 2007, as well as fully funding many participants coming from less affluent regions of the world. Even so, this year we had 19 different nationalities, including participants from non-CERN member states including China, Korea, Lithuania, Pakistan, Romania, Russia and Turkey.

Every school is co-organized by CERN together with one or more national scientific institutions in the host country. This year the school was organized in collaboration with Gjøvik University College assisted by the University of Oslo. CERN is indebted to Prof. Jørn Wroldsen, rector of Gjøvik University College, for his continued support to the school, to Prof. Are Strandlie, chairman of the local organizing committee, at the inception of the "CSC in Gjøvik" project, and to Kari Lauritzen, the local administrative manager who looked after many aspects, from budget and planning issues to logistics and recreational activities.

As usual, since 2002, the scientific programme culminated with an optional final examination. This year, 50 out of the 52 students registered – an exceptional ratio – and 47 of them were successful. This success rate should not be misinterpreted: the examination is extremely difficult and the participants worked exceptionally hard, spending hours – and nights for some – revising before the examination.

The examination jury was pleased to award special distinctions to: **Robert Zimmermann**, from Bonn University, who obtained the highest mark; **Constantin Mihai Cuciu**, **Matti Kortelainen** and **Oliver Oberst** who obtained the second best mark; and **David Gonzalez Maline** and **Paolo Tedesco** for obtaining the third best mark.

2008 will remain an historic year for the scientific programme: in addition to receiving the prestigious CSC Diploma from CERN, everyone who passed the exam



*CSC2008 consisted of 50 hours of lectures and exercises organized around three themes and was delivered by an international team of 12 lecturers.*



*The sports programme proved popular among computing school attendees.*

also received a formal European Credit Transfer System (ECTS) Certificate from Gjøvik University College. This certificate, recognized worldwide, not only certifies the value and the quality of the diploma, but in addition the students can count this towards their university studies. The CSC Diploma was assigned 5 ECTS points.

The scientific programme is not the only activity organized at CSC: it is complemented by a strong optional social and sports programme. Every afternoon a range of indoor and outdoor sports was proposed, including team sports (basketball, volleyball, football, floorball – a popular sport in Nordic universities) as well as badminton, tennis, squash and swimming. Many lecturers and organizers also turned into sports teachers, supervising activities but also permitting

several beginners to try new sports, or improve their skills in others.

As usual, the traditional football match and the mixed-doubles tournament (badminton this year) were organized in week two. The sports programme culminated with a special sports day, where students chose from several activities supervised by teams of specialists. The activities were ranked in decreasing order of physical difficulty: lake canoeing (level 5), climbing (level 4), orienteering (level 3), forest hiking (level 2), and sailing on Lake Mjøsa (level 1).

Although these activities were optional, 98% of the students participated, the majority registering for the most demanding activity. Once again, the slogan of the school proved true: work hard, play hard.

**François Fluckiger, IT-DI (CSC school director)**

## Istanbul hosts EGEE conference

What has 545 people from 48 countries, 285 presentations, 97 sessions, 12 demonstrations, and 50 partners? Answer: the EGEE conference in Istanbul.

“There was an excitement here,” said Bob Jones, EGEE project manager. “It all came together.” According to Jones the highlights included a Best Application Presentation award given to the CYCLOPS team of Marco Verlatto (INFN), Stefano Dal Pra (INFN) and Valerio Angelini (CNR-IMAA) for their “G-RISICO: A wild fire-risk-assessment application running on an advanced grid infrastructure.” The team said that their approach could help civil protection authorities predict not only wild fire, but the risks of many types of natural disasters.

Another highlight was the signing of a Memorandum of Understanding between EGEE and EDGES (Enabling Desktop Grids for E-Science). As the first of its kind, “it will bring together the large, community-based approach of EGEE with the volunteer, BOINC-style desktop world,” said Peter Kacsuk, director of EDGES and president of the Hungarian Academy of Sciences. Such an approach could potentially aid



*Peter Kacsuk, director of EDGES, and Bob Jones, project manager of EGEE, after signing the first memorandum of understanding. Image courtesy of Asli Zengin.*

researchers working in fields as diverse as fusion, biomedicine, engineering and chemistry, he added.

An outstanding range of posters covering topics from high-energy physics and bioinformatics to middleware development was displayed, and the Best Poster

Competition was closely fought. This year, conference delegates voted for Javier Rojas Balderrama’s “NEUROLOG: neuroscience application workflows execution on the EGEE grid”.

The Demonstration Competition winner was “EGEE application porting support group” presented by Gergely Sipos. The judges said that they were particularly impressed by the speed with which this group had got up and running in just a matter of months since the start of EGEE-III, designing the application, porting it and now training other teams.

The runner-up for Best Application Demonstration was “Grid-enabled virtual screening service based on grid application platform” from Mason Hsiung, Lee Hurng-Chun, Ueng Wei-Long, Chen Hsin-Yen, Eric Yen and Simon Lin.

The next EGEE user forum, OGF25, will be held in Catania, Italy from 2 to 6 March 2009, and will be hosted by INFN.

**Dan Drollette, iSGTW, and Catherine Gater, EGEE**

• *This article was published online in iSGTW on 8 October.*

## EGEE to EGI: we must not be afraid of the future

“We must not be afraid of the future,” was the sound advice from Mário Campolargo of the European Commission at the Enabling Grids for E-science (EGEE) conference in Istanbul, in his welcome address “eScience grids: Where does Europe stand?”.

Within the European Grid Initiative (EGI), planning is under way for the implementation of a sustainable, pan-European grid infrastructure to support collaborative eScience. The building blocks of this infrastructure are national grid initiatives – autonomous, national grid bodies, able to mobilize their own national funding and computing resources and to support their own research communities.

The projected transition of grids from the current short-term, project-based approach to a long-term model will require significant reorganization for the grid community; EGEE will no longer be the central organizing body for science grids within Europe. Careful planning is needed to ensure a smooth transition from today’s EGEE-based infrastructure into the future EGI model.

“During the EGI sessions today I felt that while there are still issues that exist, there is a real willingness to move forward and, above all, a new sense of urgency,” said Bob Jones, project director of EGEE. “This comes from the recognition that the end of



*Time to turn off this road and move to the next. Image courtesy of www.gerards.be.*



*Mário Campolargo addresses the European grid community during the opening plenary of EGEE 08. Image courtesy of Onno Zweers.*

EGEE-III and other collaborating projects is really just round the corner. Even when the blueprint is in place, it’s just the beginning. There is still much to be done.”

“At the same time, the media coverage

generated by the Large Hadron Collider [LHC] start-up makes now an important time to act. Decision-makers and politicians have the LHC fresh in their minds – they can see why large infrastructure projects are important,” he added.

Agreement by the EGI Policy Board to begin accepting bids for the location of the EGI organization marks a significant step forward. The EGI.org is expected to start operation in 2010. “Between now and Christmas, we need to come up with our transition plan,” said Jones. “This plan will outline how EGEE will move from where it is now to where it needs to be to fit in with the EGI planning.”

EGI Design Study project director Ludek Matyska notes that the transition must ensure continuation of the present grids and should also encourage wider adoption.

“Thanks to EGEE, and others, production grids are a reality today,” said Dieter Kranzlmüller of Ludwig-Maximilians University. “EGI is not a simple continuation of EGEE but should offer a possible solution for all European grid infrastructure projects to achieve sustainability.”

**Danielle Venton, EGEE**

*Prepared with the help of the GridTalk project.*

• *This article was published online in iSGTW on 1 October.*

## Phoney messages target CERN

Cybercriminals are making increasing use of fake e-mails and websites to steal account and identity information from users. This technique is called phishing. The e-mails often look real and try to trick you into giving your password or personal data.

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. Often you can see whether a link is fake by hovering the mouse over the link (without clicking on it). A common trick is to have a link like <http://123.www.company.com/sales> instead of <http://www.company.com/sales>, which at first sight looks genuine but has been slightly changed from the real one by adding a few extra characters.

### Suspicious activity

If you encounter any suspicious activity on your computer(s) or your account(s), please contact [Computer.Security@cern.ch](mailto:Computer.Security@cern.ch) immediately. If you have questions, please check the Computer Security web pages at <http://cern.ch/security/>.

### Fake e-mails

CERN has recently been targeted by fake e-mails that purported to be from the CERN Helpdesk and asked people to reply giving their username and password. These cases are quickly blocked by the CERN mail service, but some people will receive the mails before the mail service has been alerted. Users therefore need to be vigilant about any requests for passwords or personal data.

Please warn the security team about suspicious mails by sending them as an attachment to [Computer.Security@cern.ch](mailto:Computer.Security@cern.ch) and then delete them. You can contact [helpdesk@cern.ch](mailto:helpdesk@cern.ch) if you have any questions on this topic.

### Instant messaging

Several computers at CERN have been broken into by attackers who have tricked users of instant messaging applications (e.g. MSN or Yahoo Messenger) into clicking on weblinks that seemed to be from known contacts. The links appeared to be photographs from "friends" and requested software to be installed. However, this was attacker software.

Cybercriminals use fake messages to try to trick you into clicking on weblinks, which will help them to install malicious software onto your computer. Anti-virus software cannot be relied on to detect all cases so your vigilance is needed. If you have any questions, contact [helpdesk@cern.ch](mailto:helpdesk@cern.ch).

### Password protection

CERN's computing rules (Operational Circular No. 5) require that passwords remain confidential and must never be given to anyone, not even to the Helpdesk or other support personnel. If you think that your password could have been exposed, then you must change it immediately. For further information consult <http://computingrules.web.cern.ch/ComputingRules/>.

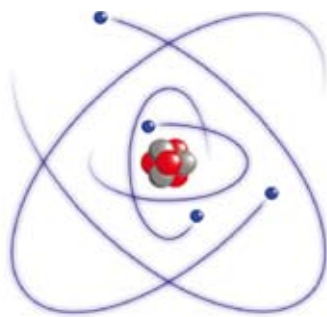
Thank you for your understanding and collaboration in this matter.

**CERN Computer Security Team**

## Next generation of CERN Linux

In an article in the September–October 2007 issue of *CNL* (<http://cerncourier.com/cws/article/cnl/31112>), we wrote that the production Linux version for LHC startup would be SLC4, despite the fact that it is rather old and not well-suited to more modern desktop hardware. The unfortunate delay associated with the LHC now increases the pressure to move off that platform, but also presents us with a window of opportunity to try to move to a newer version of Linux.

The next generation of CERN Linux (Scientific Linux CERN, SLC5) is currently being certified and ready for testing from the Linux install servers (<http://linuxsoft.cern.ch>). A next-generation LXPLUS (<http://cern.ch/plus>) cluster with SLC5 is available as [lx64slc5.cern.ch](http://lx64slc5.cern.ch). Some SLC5 LXBATCH capacity is available and is expected to grow quickly – new CPU



machines will be installed with SLC5.

SLC5 has not yet been validated for production by the experiments, and the timetable to get it certified in time for LHC is tight, but for now we are optimistic that we will be able to process LHC data on a newer platform than SLC4.

**The CERN Linux team**

## Calendar

### December 2008

1–2 **The World Summit of Cloud Computing conference IGT2008**

Israel  
<http://www.cloudcomputing.org.il/>

1–5 **9th International Middleware Conference**

Leuven, Belgium  
<http://middleware2008.cs.kuleuven.be/>

4–5 **Informatics Education Europe III**

Venice, Italy  
<http://www.dsi.unive.it/IEEIII/>

7–12 **4th IEEE International Conference on e-Science**

Indianapolis, Indiana USA  
<http://www.escience2008.iu.edu/>

17–20 **15th International Conference on High Performance Computing HiPC '08**

Bangalore, India  
[www.hipc.org/](http://www.hipc.org/)

### January 2009

14–17 **International Conference on Health Informatics 2009 HEALTHINF2009**

Porto, Portugal  
<http://www.healthinf.org>

### February 2009

2–4 **GEANT security toolset training**

Zürich, Switzerland  
<http://www.terena.org/activities/training/sectoolset/ws2/index.html>

12–15 **ICT-Professionalism, a Global Challenge 2009 Working Conference IPROF-09**

Arnhem, NL  
<http://www.IPROF09-arnhem.nl>

### March 2009

2–6 **Open Science Grid All Hands Meeting**

Livingston, Los Angeles, USA  
<http://indico.fnal.gov/conferenceDisplay.py?confId=2012>

2–6 **4th EGEE User Forum**

Catania, Italy  
<http://indico.cern.ch/conferenceCFA.py?confId=40435>

21–27 **Computing in High Energy and Nuclear Physics CHEP 09**

Prague, Czech Republic  
<http://www.particle.cz/conferences/chep2009/>

23–26 **5th International Conference on Web Information Systems and Technologies WEBIST**

Lisbon, Portugal  
<http://www.webist.org/>