

# VOCE—A Grid Environment for Central Europe

Jiří Chudoba, Lukáš Fiala, Jan Kmuníček, Jiří Kosina, Daniel Kouřil,  
Miloš Lokajíček, Luděk Matyska, Miroslav Ruda, Jan Švec

CESNET z. s. p. o., Zikova 4, 160 00 Praha 6, Czech Republic  
*email:* Jiri.Chudoba@cern.ch, fiala@fzu.cz, kmunicek@ics.muni.cz,  
kosina@fzu.cz, kouril@ics.muni.cz, lokajick@fzu.cz, ludek@ics.muni.cz,  
ruda@ics.muni.cz, svec@fzu.cz

## Abstract

Grids are distributed computing and storage systems with immense potential. However, this potential is usually not fully used, as the Grid systems are very complex, not easy to deploy and even harder to efficiently use. A lot of training is needed both for systems administrators and end users to deploy the full Grid potential and to provide a really efficient environment that helps end users to solve their problems. Also, the overhead to set up and maintain a Grid is not negligible and smaller user communities usually do not have nor the man power nor the knowledge to do it properly.

This paper provides an overview of VOCE—a Virtual Organization (VO) for Central Europe (CE). VOCE is an application neutral virtual organization for Central Europe. It offers its users—CE researchers—a complete set of Grid services. Grid novices can use the VOCE environment to get their first experience with a production level Grid environment, to undertake their initial training and to continue within the same environment for their actual work. Small scientific communities can use VOCE as their primary Grid environment, being freed from the burden of managing their own VO.

## 1 Introduction

Grids are large scale distributed systems that connect computing, storage and eventually other resources, belonging to different institutions. They use high speed network connections to transfer data and jobs between individual nodes or to coordinate a job on physically separated nodes. Through this remote access and especially the sharing of resources Grids can offer unprecedented computing power that adapts to actual needs. While they can support a single application spanning very large number of individual nodes—offering thus the truly high performance computing environment—their primary strength is in the support of high throughput computing. Tasks made from many independent jobs—e.g. the parametric studies or processing of individual parts of large experimental data sets—use Grids as an ideal computing substrate.

However, to efficiently manage and synchronize large numbers of potentially heterogeneous computing nodes and storage depots, rather sophisticated middleware must be deployed and configured. This task requires a lot of specific

administrative knowledge, whose lack may lead to less efficient or even unsatisfactory Grid environments. Also, end users must be well accustomed (i.e. trained) to this complex environment, otherwise they spend more time on tackling with Grid peculiarities than on solving the scientific problems of their own domain.

## 2 Project EGEE

The EGEE (Enabling Grids in E-SciencE) project[1] is one of the largest activities targeted to establish a production Grid infrastructure. Experts from almost 100 institutions work together on developing and operating a Grid environment that provides a production service for everyday use. The project is subdivided into several activities dealing with various aspects of Grid advancement, ranging from middleware development through deployment and operation of Grid services to end-user support and training.

Users of the EGEE Grid are organized into *Virtual Organizations (VO)*. Each VO targets one application area and is a logical meeting place of users on one side and resource owners on the other side. Resource owners provide their resources to individual VOs, reflecting thus user groups they are willing to support. VO users represent a community with common (or at least similar) scientific goals. The concept of VOs provides an abstraction of the community utilizing the Grid environment and makes the management more efficient and allows to coordinate better the effort necessary to achieve the goal of the group (e.g., by making sure that all the resources provide the necessary application software in up to date version).

Both resources of the VO and its users can be very distributed, spanning multiple organizations (and crossing multiple administrative domains) and connecting users from even larger set of organizations and countries. In order to ensure its operation, each VO utilizes several Grid services that allows e.g. to perform user management or provide access to the resources. These services are often configured and administrated by the VO itself, but a VO can also arrange with another VO to share the services among multiple VOs.

Having gone through several years of intensive development, the current Grids are getting closer to the end users and they are starting to use the Grid infrastructure in their everyday research work. However, despite the immense effort invested into the Grids over past years, current Grid environments are quite complex. Also, Grid administration is difficult due to the number of components, their different and often complex configuration, etc. Therefore, current Grids are almost entirely utilized by large research communities representing many researches and resources. Only such communities can afford to establish and operate all the services necessary for efficient use of the Grid environment.

Experiences gained by the larger groups show that Grids can be used efficiently for complex problems solving and also for very complex scientific collaboration. This potential could be also advantageous for other communities. However, smaller research groups (using the same application or working within

the same scientific area) may have enough resources, but do not have sufficient free man power to make up and continuously manage a basic VO infrastructure. And if their resources are not sufficient, they may not be strong enough to negotiate with large resource owner over the conditions to access these third party resources.

In the rest of this paper we describe an environment built in Central Europe that makes it possible for such smaller groups or even individual users to join an established Grid infrastructure and take advantage of the use of current Grid technologies.

### 3 VO for the Central Europe—VOCE

In terms of Grid management and operation, EGEE organizational structure reflects the geographical principle, with *federations* as its basic building blocks. The Central Europe federation (CE) is composed of institutions from Austria, Czech Republic, Hungary, Poland, Slovakia, and Slovenia. In order to ease access to the Grid for user from the region, the CE federation established and operates the *Virtual Organization for the CE federation—VOCE*.

The purpose of VOCE is to provide a complete production level Grid environment available for each and every user from the CE federation. The infrastructure established for VOCE contains all necessary Grid services and was built to be as independent as possible on components provided by third-party services outside the CE federation. The core services of VOCE are operated by CESNET, the Czech NREN provider, computing and storage resources are provided by all members of the CE federation. Currently, VOCE offers access to 500 CPUs and nearly 6 TB of storage. Despite these resources are usually shared with other projects and also change quite dynamically, the users of VOCE gain access to a very powerful environment.

VOCE primarily concerns with two user categories: Grid newcomers and users porting their applications for the Grid. The newcomers usually have no experience with Grid computing and are interested in getting started with the basic Grid technologies without having to configure and operate their own infrastructure. The latter users' group contains users who are familiar with the Grid environment and also have applications to test on the Grid but do not have resources to maintain their own infrastructure (or need an experimental environment to test whether such investment will eventually pay off). These users come usually from classic areas of distributed computing and want to find out how their applications behave in the Grid world. VOCE makes it possible for users from both these categories to join quickly a production Grid environment and to start its use. Since operation of the all middleware services is outsourced to the VOCE administrators, the users can fully concentrate on solving their problems without being bothered with issues concerning Grid administration.

Besides an environment for training, testing and evaluation, VOCE also provides an platform for everyday production. Users who have already got acquainted with the Grid environment and have their applications prepared for

running on the Grid can use the VOCE facilities for their day-to-day work. Users of VOCE smoothly shift from learning the Grid through experimenting with its tools to real production utilization without having to change the environment. If requirements of some user groups inside VOCE exceed the capability of the VOCE infrastructure (e.g. in terms of required number of resources or their availability) they can establish a new VO infrastructure that is optimized to fulfill their needs. VOCE administrators are prepared to provide guidance and basic support in order to help establish such a brand new VO.

### 3.1 VOCE Infrastructure

In order to be independent on other sites and administrators VOCE provides own installation of the Grid middleware services. In the first stage the infrastructure was based on the middleware distributed by the LCG project[2], later on we installed and tested the gLite software[3] and the services are being gradually replaced with gLite versions. So far we upgraded the Workload Management System and Logging and Bookkeeping services, which form the basic services for job management and monitoring. Other services will be upgraded shortly. We are also evaluating other services available from gLite which are not included in the LCG distributions (e.g. R-GMA, sophisticated data management services). When these services prove stable and useful, they will be offered for production usage.

VOCE provides a specialized machine containing all the software necessary for the users to access the Grid environment (*User Interface*). The machine can be accessed using the SSH protocol, the users can authenticate themselves using either their Grid proxy certificates or the username/password pair chosen during the registration. The UI machine provides an account for each VOCE user.

One of the VOCE goals is to provide a Grid environment that can be easily joined and used to test the Grid possibilities. However, in contrast to other “easily accessible” Grids, VOCE requires standard identification of all its users (including newcomers). This way the resource owners who provided their resources for VOCE can be sure that any misbehaving user can be easily tracked down and his or her access rights canceled.

Even when used only for training and demonstration, when most other similar training Grid environments (e.g. GILDA[4]) usually require very weak or no authentication (including anonymous access), based usually on a simple e-mail exchange over unsecure communication channels. These training environments often issue a certificate based on this message only. While allowing very simple registration process, such an approach inevitably weakens the environment and also leads to only limited amount of resources available for such a VO. With such a registration procedure in use, it is possible for a potential malicious users to get registered very easily, without revealing their true identity. Also, the users are not trained in proper security habits and conventions—this decreases the value of such training environments. In comparison with such environments VOCE does not support such semi-anonymous users and requires all applicants to possess a certificate issued by a Certification Authority approved by the EUGridPMA[5]

or IGTF[6] bodies. The minimal requirements put on the accreditation ensure that users must prove their identity during the certification process and also that the certificates can be reasonably trusted.

Unlike most current VOs in EGEE, VOCE is not tied with any particular application. On the contrary, one of the key purposes of VOCE is to attract users from various application areas and act as an incubator for introduction of new applications to the Grid. Experts in VOCE are prepared to assist with porting applications from other computing environments.

The VOCE infrastructure can be also used as a basis for *t*-Infrastructure in the CE federation that is undistinguishable from the production environment. This is used to demonstrate the possibilities of the Grid environment to the newcomers without sacrificing any individual aspect of fully production Grid environment.

VOCE users are controlled by a sophisticated user management system Perun[7] developed originally for a Czech nation-wide Grid environment. This system allows to mount very complex mechanisms to handle VOCE users. It is used as the primary storage of the user records, and using various modules it is able to push user data to supported services. Currently, records to the LDAP and VOMS servers are propagated, so the resources are able to generate Gridmap-files and mount access control mechanisms based on the VOMS attributes. The Perun system also controls generation of user accounts on the User Interface machine. Users' passwords are stored in a Kerberos service, which makes it possible to retain a single password for accessing multiple password-based service (currently UI and the Genius portal). The schema created for user management also allows administration of users by different administrators which are assigned only a particular set of users to maintain. Internally, users are organized into groups according to their certificate subject names and each group is maintained by administrators that are closest to the group (e.g. the SZTAKI users are primarily taken care of by the SZTAKI administrators).

### 3.2 Joining VOCE

VOCE is open to all users from the CE federation who agree with the VOCE usage rules. These rules define very basic principles of the use of the infrastructure, their current version is compatible with recommendation from the EGEE documents.

Basic prerequisite for acquiring a membership in VOCE is possession of a public-key certificate signed by a trusted certification authority (CA). VOCE relies on recommendations issued by the EUGridPMA and IGTF bodies and accepts certificates issued by all accredited CAs. Currently, all member states of the CE federation are covered with a national CA so each user in the CE federation can easily receive a certificate from an acknowledged CA. Therefore, VOCE does not operate any catch-all CA.

In order to ask for a membership in VOCE, a new user is required to fill

in an on-line form available from the VOCE home page<sup>1</sup>. Access to this form is only allowed for clients who have a valid certificate loaded in their browsers. Registration procedure consists of following steps:

1. First, the user enters his or her contact information (name, e-mail, affiliation, and telephone number). After filling in this information, the registration application sends a verification e-mail to the e-mail address specified in the request. The purpose of this e-mail is to verify the e-mail address provided.
2. After receiving the e-mail, the user must access the page referred to from the e-mail. Its URL is formed using a random string, unique for each request.
3. The user fills in requested username and password to his or her account on the VOCE User Interface machine and submits this information to finish the registration application.

After the user has finished the registration procedure, the user management system sends a notification e-mail to the VOCE administrators about the new request. The message contains the identification information of the user as well as a name of institution that is supposed to manage the user<sup>2</sup>.

The application enters a three-day period during which the responsible administrators can cancel the user's request. If no cancellation is performed and no objections from other administrators are raised, the user's account gets activated automatically after this period passes.

During the activation process a user record is generated in the VOCE LDAP and VOMS service so it can be propagated to the gridmap-files and also an account is created on the VOCE UI machine. The user is notified by an email about the activation of the account. The procedure is the opposite of what other production VOs usually perform, it requires the administrators to act if they want to reject users instead of grant them access. We believe this procedure provides a good trade-off between the speed of the registration process and level of risk. Of course, the administrators as well as the resource owners have mechanisms how to ban users later on whenever they feel such a need (e.g. when the user starts to cause operational problems).

In addition to command-line oriented access from the User Interface machine, the users can also use specialized web portals that hide a lot of technical details and allow user to familiarize quickly with the infrastructure and utilize it. Two portals are available for VOCE—Genius[8] and P-GRADE[9]. The Genius portal was developed to support training and demonstration activities and is often used as a tool in training courses that aim at Grids. The P-GRADE portal offers mechanism for solving complex problems. It supports workflow management and provides support for legacy applications, it is also possible to use the P-GRADE portal to access multiple Grids.

VOCE users can also use a unique framework Charon[10]. Charon creates a layer upon the standard Grid middleware environment and makes it easier to

---

<sup>1</sup><http://egee.cesnet.cz/en/voce/>

<sup>2</sup>Currently all users are managed centrally by CESNET

access the complex Grid infrastructure. Unlike the concept of portals, Charon provides a command-line oriented interface and is supposed to be used by users that require to have a full control over their running jobs. Charon is very modular and can be adapted for many different environments. In its current implementation, Charon supports environments using either pure PBS system or LCG-based Grids. Regardless the particular environment used on lower layer, Charon retains a single interface for the users, so they can easily use both these environments, which are very different.

## 4 Conclusion

The VOCE environment described in this paper provides a complete Grid infrastructure which is available for all users from Central Europe. It is based on the latest Grid technologies and makes them available to every admissible user in the region. It can be utilized by Grid novices to easy join and test the Grid services as well as by smaller research groups that do not have resources or skills to establish and maintain their own VO infrastructure. VOCE provides an application neutral environment which is open to researchers from all scientific domains. In addition to a full set of common Grid services known from other VOs, the VOCE users can also leverage less usual tools such as the Charon system or P-Grade portal for more efficient use of the Grid environment.

## References

1. The Enabling Grids for E-science (EGEE) Project.  
<http://www.eu-egee.org/>
2. The LHC Computing Project (LCG).  
<http://lcg.web.cern.ch/LCG/>
3. gLite—Lighthouse Middleware for Grid Computing.  
<http://www.glite.org/>
4. Grid INFN Laboratory for Dissemination Activities (GILDA).  
<https://gilda.ct.infn.it/>
5. The European Policy Management Authority for Grid Authentication.  
<http://www.eugridpma.org/>
6. International Grid Trust Federation.  
<http://www.gridpma.org/>
7. A. Křenek, Z. Sebastianová. Perun – Fault-Tolerant Management of Grid Resources. In *Krakow Grid Workshop 2004*. Krakow : Academic Computer Centre CYFRONET AGH, 2005 ISBN 83-915141-4-5. p. 133-140. 2005
8. Grid Enabled web eNvironment for site Independent User job Submission (GENIUS).  
<https://genius.ct.infn.it/>
9. Parallel Grid Run-time and Application Development Environment (P-GRADE).  
<http://www.lpds.szaki.hu/pgrade/>.
10. J. Kmuníček, P. Kulhánek, P. Petřek. CHARON System – Framework for Applications and Jobs Management in Grid Environment. *Krakow Grid Workshop 2005*. Krakow.