IL NUOVO CIMENTO Vol. 32 C, N. 2 Marzo-Aprile 2009
DOI 10.1393/ncc/i2009-10444-2

Colloquia: CSFI 2008

Grid: From EGEE to EGI and from INFN-GRID to IGI

A. Ghiselli and M. Mazzucato
INFN CNAF - Bologna, Italy

(ricevuto il 4 Agosto 2009; pubblicato online il 29 Ottobre 2009)

Summary. — In the last fifteen years the approach of the "computational Grid" has changed the way to use computing resources. Grid computing has raised interest worldwide in academia, industry, and government with fast development cycles. Great efforts, huge funding and resources have been made available through national, regional and international initiatives aiming at providing Grid infrastructures, Grid core technologies, Grid middleware and Grid applications. The Grid software layers reflect the architecture of the services developed so far by the most important European and international projects. In this paper Grid e-Infrastructure story is given, detailing European, Italian and international projects such as EGEE, INFN-Grid and NAREGI. In addition the sustainability issue in the long-term perspective is described providing plans by European and Italian communities with EGI and IGI.

PACS 89.20.Ff – Computer science and technology.

1. – Introduction

Since the end of the last century we have observed a strong computing and networking technology evolution leading to more powerful computers and to high speed networks as low-cost commodity components. In parallel there has been a growth of large-scale computing needs with high number of users widely distributed. This scenario has changed the way to think about and use computing resources. One of the most interesting and challenging approaches has been the "*computational Grid*" concept.

The concept of "*Grid computing*" was coined almost fifteen years ago by scientists in Argonne National Laboratory and the University of Chicago to denote a distributed computing infrastructure for advanced science and engineering [1]. The name "Grid" comes from the metaphor of "Electrical Grids" and the idea to get access to a resource (*e.g.*, electricity) by using a plug. However, the definition of the term "Grid" evolved on the basis of experience from Grid implementation projects. Back in 1998, C. Kesselman and I. Foster [1] provided a first Grid definition: *"A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities"*. Then in a subsequent paper, I. Foster *et al.* [2] refined the previous definition to also address social and policies issues, adding that Grid computing is concerned with *"coordinated resource sharing and problem solving in*

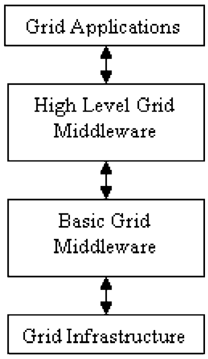

Fig. 1. – Grid layers.

dynamic, multi-institutional virtual organizations". In 2002, I. Foster [3] explained the essence of the Grid in a simple checklist. A Grid is a system that 1) coordinates resources that are not subject to centralized control . . . 2) . . . using standard, open, general purpose protocols and interfaces . . . 3) . . . to deliver nontrivial qualities of service.

DataGrid [4] was the first European project (January 2001) aiming to build a computational grid prototype on the basis of requirements [5] coming from users in the application fields of the High Energy Physics, Earth Observations, and Bio-Informatics. This project demonstrated the validity of the model and other European, national and international projects followed in order to evolve the Grid model as distributed computing infrastructure able to share geographically distributed resources (such as data, storage, computers, software, tools, applications, instruments, and networks) securely, at anytime, and anywhere. Currently, a variety of applications ranging from aerospace, astronomy, bioinformatics, chemistry, environment, finance, physics and health care are running on the Grid. The Grid software layers are shown in fig. 1. These layers reflect the architecture of the services developed so far by the most important European and international projects (like DataGrid, EGEE III [6], OSG [7], NAREGI [8]), and that are already providing a production service. The Grid architecture (shown in fig. 2) is characterized by the following components: Fabric with computational, storage, network, code repositories and catalog resources; Connectivity with secure access (like Globus GSI [9] combined with Virtual Organization Management System (VOMS) [10]); Resource with

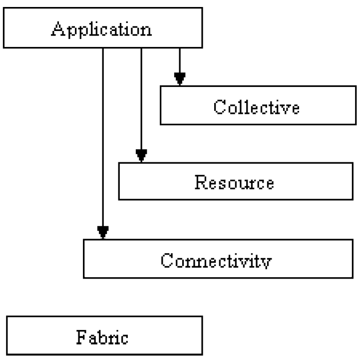

Fig. 2. – Grid architecture.

virtual interfaces and protocols (like GRAM [11], CREAM [12], and GridFTP [13]); Collective with monitoring (like GridICE [14]), diagnostics, data replication, accounting (like DGAS [15]), authorization workload management systems (like GPbox [16] and WMS [17]).

This paper is organized as follows. Section **2** describes the story of Grid e-Infrastructures emphasising European and Italian projects. Section **3** details the importance of sustainability of Grid projects today. Section **4** provides middleware information. Finally sect. **5** summarizes conclusions.

2. – Grid e-Infrastructures story

During the last eight years, several European and national projects (hundred of Millions euros) have developed and implemented a set of top level robust open source Grid middleware services (*e.g.*, the gLite release [18]) providing a solid foundation for researchers activities. Moreover these projects have deployed in the world the largest Grid e-Infrastructure (*i.e.* Internet plus Grid) identified by Enabling Grid for E-sciencE (EGEE) [19] then by EGEE II [20] and finally by EGEE III. EGEE is not the only existing Grid infrastructure, and many others are distributed around the world. Several of them are European Community (EC) co-funded projects, like EU-China [21], EU-India [22], EUMEDGrid [23], and EELA [24]; in Europe there are also ARC [25] and DEISA [26]; OSG, and TeraGrid [27] in USA; and NAREGI in Japan. All these Grids have set up an activity working group called "Grid Interoperability Now", currently renamed in "Production Grid Initiative (PGI)", aiming to study interoperability problems between them and find solutions on order to allow users to submit jobs to them in a transparent way. This collaborating Grid infrastructure is potentially linking eighty countries, making up a worldwide Grid infrastructures.

Members States and European Community (EC) are now guaranteeing the long-term sustainability of the baseline Grid middleware and e-Infrastructures in the European Grid Initiative (EGI)-Design Study (DS) [28] (*i.e.* a FP7 Project) and the National Grid Initiatives (NGIs) as described later.

2·1. *What EGEE delivers.* – EGEE implements a service-oriented architecture that virtualises resources. It is providing real services, not prototypes, daily offered by federations of well managed resource centers to thousands of users of several research communities. Most of the services adheres to the International Standards (like Web Services) or to standards defined by OGF [29] and OASIS [30], and are based on high level of security (like x.509 PKI). EGEE distributes the gLite production services under business friendly open source licence. gLite services enable users to be part of a Virtual Organisation (VO) for authentication and authorization on the Grid, to submit jobs with different flavors (*e.g.*, batch-like, dag workflow, collection, parametric, MPI, and interactive) to a set of committed resources, selected by a broker as a result of a best match mechanism between job description and resources characteristics, made available by the Grid information system. Standard interfaces and protocols have been defined and implemented to access computing resources and storage systems. Data replication mechanisms allows management and access to different copies of a file for optimizing data access and providing fault tolerance. There are few numbers describing the gLite production services today: 150 KJobs per day successfully completed in 250 centers; up to >10 GByte/s of aggregated sustained access to scientific data and transparent access to hundred TB of Storage.

TABLE I. – *List of applications.*

Application	Description
Simulation LHC Monte Carlo simulations	Fusion; WISDOM Jobs needing significant processing power; Large number of independent jobs; limited input data; significant output data.
Bulk processing HEP	Processing of satellite data distributed input data; Large amount of input and output data; Job management; Metadata services; complex data structures.
Parallel jobs	Climate models; Computational chemistry; Large number of independent but communicating jobs; Need for simultaneous access to large number of CPUs; MPI libraries.
Short-response delays prototyping new applications	Grid monitoring grid; Interactivity limited input and output data; processing needs but fast response and quality of service.
Workflow medical imaging	Food analysis; Complex analysis algorithms; Complex dependences between jobs
Commercial applications non-open source software	Geocluster (sismic platform); FlexX (molecular docking); Matlab, Mathematics.

2·1.1. Infrastructure operation. The infrastructure consists of hundred of sites distributed across many countries (55) with large quantity of CPUs and storages. There is a continuous monitoring of grid services and automated site configuration (or management) tools and procedures are available. An essential part of grid operations is that of operation security coordination, driven by the needs of ensuring that security problems are well coordinated across the distributed infrastructure. Multiple VOs from diverse research disciplines are supported through a managed process from the first contact to the production usage. This is done by training programs, experiencing in grid-enabling applications and on-line help-desk. The organisation of networking events (like User Forum, and Conferences) contributes to information dissemination for users. Communities using EGEE are: Archeology, Astronomy, Astrophysics, Civil Protection, Comp. Chemistry, Earth Sciences, Finance, Fusion, Geophysics, High Energy Physics, Life Sciences, Multimedia, Material Sciences and others. Table I contains a list of applications supported by EGEE [31].

2·1.2. EGEE III vision and mission. Given the high degree of maturity, the main vision of EGEE III is to make a strong move towards a sustainable world-wide production quality Grid infrastructure by appropriate technical and organisational decisions. This must be capable of providing services to an increasing rapid number of application areas, and make Grid technology easily accessible and usable for these communities. The second year program of EGEE III foresees a transition plan to a more sustainable organisation, in syncronisation with the parallel EGI-DS project, as described in the following.

3. – Why sustainability today?

After the success of EGEE, *dependence*, *protection of investments* and *perspective* are the major reasons justifying the "*sustainabilty*" approach for a Grid infrastructure organisation: *dependency* because some application domains depend on grids already today due to the good production service provided by EGEE; *protection of investments* because investments in grids, both from funding organizations and from users, need to be protected; and *perspective* because current grid users are enthusiastic about grid, therefore grid users will ask for a longer term perspective. Moreover at the EGEE'06 Conference (September 25, 2006) Viviane Reding, Commissioner, and EC said: "*... for Grids we would like to see the move towards long-term sustainable initiatives less dependent upon EU-funded project cycles*".

3˙1. *EGI-DS*. – The EGI-DS project, started in September 2007 and partially funded by the European Commission over 27 months, has the aim of bringing about the creation of a new European organizational model, capable of fulfilling this vision of a sustainable European computing grid infrastructure for e-science. The foundation for e-Infrastructure sustainability relies on establishing NGIs as legal organizations, supported by governments, and providing a unique representation at the European and international levels for all the communities related to national grid infrastructures from resources providers to scientific users. EGI-DS has collected and consolidated the requirements of a wide range of research disciplines within a large number of NGIs. It has designed the required functionality and modelled the organization that could consolidate, operate, manage and continue to develop a sustainable European e-Infrastructure model as described in the EGI Blueprint [32]. Examples of NGIs are: Austrian Grid Initiative, BEgrid (Belgium) DutchGrid (Netherland), Institute des Grilles (France), D-Grid (Germany), HellasGrid (Greece), Grid Ireland, IGI (Italy) NDGF (Nordic countries), Portuguese Grid, Swiss Grid, UK e-Science, National Grid Service, OMII, GridPP for UK. By the end of 2009, EGI and NGI should become stable organizations as it was for the NRNs and DANTE about the European Research Network Infrastructure. Each NGI should be a recognized national body with a single point-of-contact, able to mobilise national funding and resource, to operate the national e-Infrastructure, to support user communities (application independent, and open to new user communities and resource providers), to contribute and adhere to international standards and policies. Responsibilities between NGIs and EGI are split to be federated and complementary.

3˙2. *National Grids*

3˙2.1. INFN Grid. In Italy the first project about Grid was INFN Grid [34]. It is the INFN special project, started in 2000 and still on going with several goals: to promote computational grid technologies research and development middleware through European and international projects (see table II) and through internal research and development activities; to implement the INFN Grid infrastructure and to set up the national Grid infrastructure for the national research community (see table II); and finally to participate to the implementation of the global Grid infrastructure for the LHC community (WLCG: Tier1 and n*Tier2 [35]) through the setting up of the e-Infrastructure for the European research area EU FP6-7 (EGEE for the European production infrastructure, and EGI for Grid sustainability based on National Grids).

3˙2.2. Italian participation to EGEE. The Italian participation to EGEE has been organized by a joint research unit (JRU) coordinated by INFN. The federation includes five partners: INFN, ENEA, SPACI-UNILE, SPACI-UNINA, and SPACI-UNICAL. It also includes five more research centers: Istituto Tecnologie Biomediche (CNR/BARI with LIBI Project), PERUGIA University, Istituto Linguistica Computazionale, CNR-PISA,

TABLE II. – *List of European, International and National Grid projects.*

Category	Project	Main results
European and International	DataGrid	RB
	DataTag	VOMS, GLUESCHEMA, GridICE
	EGEE, EGEEII, EGEE III	WMS, VOMS, DAG, GPbox, VOMS Admin
	OMII-Europe [33]	GLUESCHEMA II, VOMS SAML
	CoreGrid	Studies and prototypes of new technologies and services for Grid
National	Grid-IT, INFN-Grid	operation model for the national Grid Infrastructure, Grid-Data Source Engine, StoRM, G-Pbox, GridICE II

Scuola Normale Superiore in PISA, ESA-ESRIN. Today there are 45 sites, 4500 CPUs and almost 100 TB.

3˙3. *IGI: the Italian NGI*. – IGI (Italian Grid Infrastructure) is the evolution of the federation participating to EGEE. and it is open to new partners. It is offically supported by the Italian Ministry of University and Research. It is recognised by the EC as the Italian Grid representative. IGI is foreseen to become an official entity in 2009.

The IGI partners are: Istituto Nazionale di Fisica Nucleare (INFN), Ente per le Nuove tecnologie, l'Energia e l'Ambiente (ENEA), Consiglio Nazionale delle Ricerche (CNR), Istituto Nazionale di Astrofisica (INAF), Istituto Nazionale di Geofisica e Vulcanologia (INGV), Università degli Studi di Napoli Federico II, Università degli Studi della Calabria, Sincrotrone Trieste S.C.p.A. (ELETTRA), Consorzio COMETA, Consorzio COSMOLAB, Consorzio SPACI, Consortium GARR, University of Perugia, University of Piemonte Orientale.

4. – Re-design of current Middleware organisation

The main middleware distributions currently used in European production grids are ARC [25], gLite and UNICORE [36]. They support now a large number of user communities with complementary requirements and dimensions, ranging from teams of a few individuals to very large international collaborations with thousands of researchers and tens to hundreds thousand jobs daily. These middleware solutions have become a reference in many countries and also outside Europe (like Asia-Pacific, Africa, India, China, and South America). Although the existing middleware has demonstrated the ability to support a production infrastructure, there is also clear evidence that it will need to evolve in response to the evolution of technology and to the continuous flow of user and operation requirements in areas such as functionality, robustness, usability, deployment, adherence to standards, interoperability with other infrastructures, with the additional constraint to maintain interface stability.

As a collaboration of NGIs to run grid infrastructures and to provide accompanying services, EGI itself is not directly responsible for the middleware development, and the maintenance, support, interoperability, and development of the middleware will be an

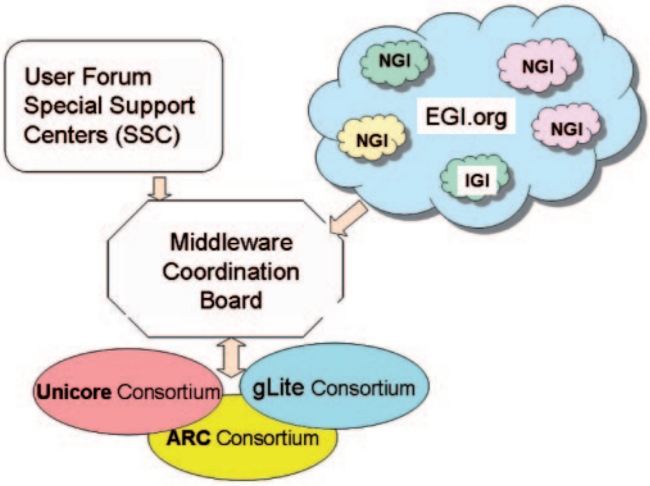

Fig. 3. – Middleware coordination board.

activity of the MW consortia gLite, ARC and UNICORE. They have two main tasks from EGI:

1) to guarantee the maintenance, support and interoperability of the middleware currently deployed on the e-Infrastructure —this is a task similar to any software maintenance, with its rules and costs provided as a general service offered to all European grid users.

2) To provide the further development required by the research communities, VOs and operational teams.

To guarantee the continuity of the grid infrastructures in Europe, EGI sees further evolution in the form of an EGI Unified Middleware Distribution (UMD), similar to the Virtual Data Toolkit (VDT) [37] in USA, which does not constitute the development of a new middleware stack but implements a unified distribution of certified components of the current stacks. As a consequence of the new organisation, the institutions involved with the gLite developments (*i.e.* Elsag-Datamat (IT), CESNET (CZ), INFN (IT), SWITCH (CH), FOM (NL), UH.HIP, CERN, STFC, UNIMAN) must set up a Consortium. This Consortium will be open and it can be seen also as an opportunity for the EU companies and research institutions needing to preserve and evolve the EGEE middleware in the new scenario of EGI and NGIs.

Since past experience has shown that only the strong interaction between operation, application and middleware activities is able to deliver the best results, the strategic planning of the activities related to the EGI middleware function will be made by the Middleware Coordination Board, an EGI.org body including representatives from the operations (including NGIs), applications and middleware development Consortia (see fig. 3). Figure 4 shows the schedule towards the new EGI and NGI organisations.

5. – Conclusions

Grids are all about sharing, representing means of working with groups around the world. EGEE operates the world's largest multi-disciplinary grid infrastructure for sci-

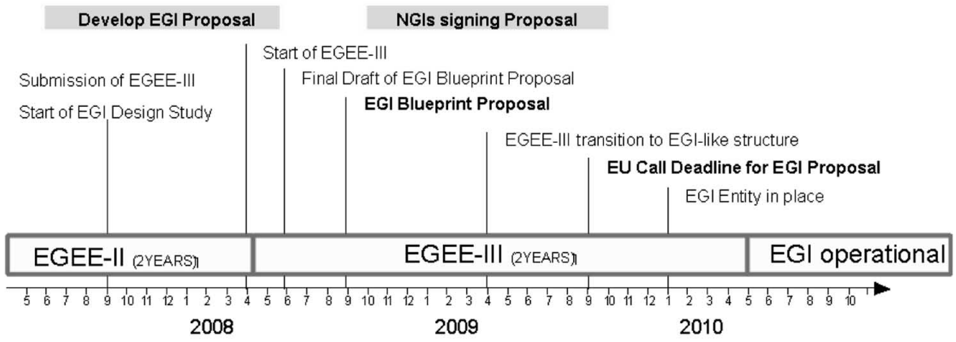

Fig. 4. – Schedule towards EGI.

entific research in constant and significant production usage. EGI needs to prepare the long-term perspective. Today we have a window of opportunity to move grids from research prototypes to permanent production systems (as networks did a few years ago). EGEE, collaborating projects, national grid initiatives and user communities are working to define a model for a sustainable grid infrastructure that is independent of short project cycles. IGI has to become a legal entity quickly in order to define the financial model and the list of services able to offer, and to play a major role in the establishment of EGI. The coming "*Consorzio gLite*" will offer an important opportunity to commercial enterprise to profit by public investments done until today in order to develop a business model based on Consorzio gLite support.

* * *

This work was produced, supported and perpetrated under the auspices of the Nuclear Physics National Institute (Istituto Nazionale di Fisica Nucleare).

REFERENCES

[1] Kesselman C. and Foster I., *The Grid: Blueprint for a New Computing Infrastructure*, (Morgan Kaufmann Publishers) 1998.
[2] Foster I., Kesselman C. and Tuecke S., *The Anatomy of the Grid*, at `http://www.globus.org/alliance/publications/papers/anatomy.pdf`.
[3] Foster I., *What is the Grid? A Three Point Checklist*, at `http://www-fp.mcs.anl.gov/~foster/Articles/WhatIsTheGrid.pdf`.
[4] *The EU DataGrid Project*, at `http://www.eu-datagrid.org/`.
[5] Kunszt P., *European DataGrid project: status and plans*, in *Proceedings of Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Vol. **502** (Elsevier Science), 2003, pp. 376-381.
[6] *Enabling Grids for E-sciencE (EGEE)*, at `http://www.eu-egee.org/`.
[7] *Open Science Grid*, at `http://www.opensciencegrid.org/`.
[8] *About NAREGI*, at `http://www.naregi.org/index_e.html`.
[9] Foster I., Kesselmann C., Tsudik G. and Tuecke S., *A security Architecture for Computational Grids*, in *Proceedings of 5th ACM Conference on Computer and Communications Security Conference*, 1998, pp. 83-92.

[10] Alfieri R., Cecchini R., Ciaschini V., dell'Agnello L., Frohner A., Gianoli A., Lorentey K. and Spataro F., *VOMS, an Authorization System for Virtual Organizations*, in *Across Grid 2003, LNCS 2004*, edited by Fernandez F. *et al.* (Springer-Verlag, Berlin, Heidelberg) 2004, pp. 33-34.
[11] Czajkowski K., Foster I., Karonis N., Kesselman C., Martin S., Smith W. and Tuecke S., *A Resource Management Architecture for Metacomputing Systems*, at ftp://ftp.globus.org/pub/globus/papers/gram97.pdf.
[12] Aiftimiei C., Andreetto P., Bertocco S., Dalla Fina S., Dorigo A., Frizziero E., Gianelle A., Marzolla M., Mazzucato M., Sgaravatto M., Traldi S., Zangrando L., Mendez Lorenzo P. and Miccio V., *Using CREAM and CEMON for job submission and management in the gLite middleware*, to appear in *Proceedings of CHEP'09, 17th International Conference on Computing in High Energy and Nuclear Physics, 21-27 March 2009 Prague, Czech Republic*.
[13] Allcock W., *GridFTP: Protocol extensions to FTP for the Grid*, in *Global Grid Forum*, 2003, http://www.ggf.org/documents/GWD-R/GFD-R.020.pdf.
[14] Andreozzi S., Aiftimiei C., Cuscela G., Dal Pra S., Donvito G., Dudhalkar V., Fantinel S., Fattibene E., Maggi G., Misurelli G. and Pierro A., *Next Steps in the Evolution of GridICE: a Monitoring Tool for Grid Systems*, in *J. Phys.: Conf. Series: CHEP'07*, Vol. **119** (2008).
[15] Piro R. M., Guarise A. and Werbrouch A., *Simulation of Price-Sensitive Resource Brokering and the Hybrid Pricing Model with DGAS-Sim*, in *Proceedings of the 13th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, 2004, pp. 325-330.
[16] Ciaschini V., Ferraro A., Forti A., Ghiselli A., Venturi V., Gianoli A., Luppi E., Stagni F. and Tomassetti L., *Distributed policy framework across multiple grid domains*, in *Proceedings of Nuclear Science Symposium*, Vol. **1** (IEEE), Oct. 26 2007-Nov. 3 2007, pp. 892-897.
[17] Cecchi M., Capannini F., Dorigo A., Ghiselli A., Giacomini F., Maraschini A., Marzolla M., Monforte S., Pacini F., Petronzio L. and Prelz F., *The gLite Workload Management System*, in *Proceedings of 4th International Conference on Grid and Pervasive Computing Geneva, 4-8 May 2009*.
[18] *gLite - Lightweight Middleware for Grid Computing*, at http://glite.web.cern.ch/glite/.
[19] *Enabling Grids for E-sciencE (EGEE)*, at http://egee1.eu-egee.org/.
[20] *EGEE II*, at http://egee2.eu-egee.org/.
[21] *The EU China Grid Initiative*, at http://www.euchinagrid.org/.
[22] *EU-IndiaGrid*, at http://www.euindiagrid.eu/.
[23] *The EUMEGrid Initiative*, at http://www.eumedgrid.org/.
[24] *E-science Grid facility for Europe and Latina America (EELA) 2*, at http://www.eu-eela.eu/.
[25] *Know ARC Project*, at http://www.knowarc.eu/.
[26] *DEISA*, at http://www.epcc.ed.ac.uk/research/collaborations/deisa/.
[27] *TeraGrid*, at http://www.teragrid.org/.
[28] *EGI*, at http://web.eu-egi.org.
[29] *Open Grid Forum*, at http://www.ogf.org/.
[30] *Advancing open standards for the global information society (OASIS)*, http://www.oasis-open.org/home/index.php.
[31] EGEE INFNSO-RI 222667 http://www.eu-egee.org/.
[32] *The EGI Blueprint*, at http://web.eu-egi.eu/press/releases/egi-blueprint-endorsed-by-the-egi-policy-board/.
[33] *OMII Middleware Infrastructure Institute (OMII) Europe*, at http://omii-europe.com/.
[34] *Welcome to INFN Grid*, at http://grid.infn.it.
[35] *Worldwide LHC Computing Grid (WLCG)*, at http://lcg.web.cern.ch/LCG/.
[36] *UNICORE (Uniform Interface to Computing Resources)*, at http://www.unicore.eu/.
[37] *About VDT*, at http://vdt.cs.wisc.edu/components/vdt.html.