Grid Initiatives: Lessons Learned and Recommendations *)

Version 2.0

With a Summary from the 2nd International Workshop on Campus and Community Grids,
held in Manchester, UK on 7th May 2007, as part of OGF20

Wolfgang Gentzsch
RENCI, Duke & D-Grid, January 21, 2007

1 Introduction

The aim of the following study of six major grid projects is to better understand, design, build, manage and operate national and community grids, based on the experience of early adopters and on case studies and lessons learned from selected grid projects. For this purpose, we have selected and analyzed a few major grid projects and grid infrastructures around the world, namely the UK e-Science Programme, the US TeraGrid, Naregi in Japan, the ChinaGrid, the European EGEE, and the German D-Grid initiative. In the following, we briefly describe these projects, summarize Lessons Learned, and provide Recommendations for those who intend to build national or community grids in the near future.

Our research so far is based on information from project Web sites and reports, and from interviews with major representatives of these grid initiatives. Major focus of our research and of the interviews was on applications and strategic direction, government and industry funding, national and international cooperation, and on strengths and weaknesses of the grid projects as described by the interviewees. The interviewees answered questions about their role in the project, successful and less successful approaches, sustainability of the resulting grid infrastructure, integration of individual results into the bigger picture, available funding, international collaboration, commercial services, and the future of e-Science. As a result, we have compiled a list of "Lessons Learned" and "Recommendations and Conclusions" which recently have helped us and others to successfully plan, implement, operate and fund grid projects.

In the Appendix, we include the proceedings from the 2nd International Workshop on Campus and Community Grids, held in Manchester, UK on 7th May 2007, as part of OGF20. We believe that this information about several additional and important grid initiatives might be useful for the reader. The appendix also includes links to the presentations that were accepted by the program committee.

2 What's Grid Computing ?

Any infrastructure such as road or rail, water or electricity, aims at expanding our environment and improving quality of life. Therefore, we are investing money and resources to build and maintain such infrastructures which are an indispensable part of our lives.

In the last few decades, a new infrastructure evolved with the Internet, and we recognized that this new technology facilitates communication and collaboration and helps increasing and preserving our knowledge. In the 90s, the World Wide Web has dramatically simplified the access to this 'Information Highway' such that everybody can make use of it, today, for their business or personal benefit.

---

*) This report has been sponsored by the Renaissance Computing Institute at the University of North Carolina in Chapel Hill

During the past 10 years, the World Wide Web has become pervasive in many facets of our professional and personal lives. And recently, it has evolved into what experts call "Web 2.0" [1]. Today, another Internet revolution is emerging: Grid Computing. While the Web offers easy access to mostly static information via Hypertext, the Grid adds another fundamental layer to the Internet, by enabling direct access to and use of underlying resources, such as computers, storage, scientific instruments and experiments, sensors, applications, data, and middleware services. Based on widely accepted grid and web services standards, resources communicate with each other and deliver results as services back to the user. These resources are part of a service-oriented architecture, called OGSA, the Open Grid Services Architecture, [2]. For the past several years, early adopters in research and industry have been building and operating prototypes of grids for global communities, virtual organizations, and within enterprises.

However, so far, direct access to the 'Information Power Plants' (e.g. the computers) of the Internet is mostly left to computer experts, scientists and engineers who are currently developing the next-generation e-infrastructure, the so-called Grid, which will enable any user easy access to any resource.

Goal is to get IT from the wall socket, as a service, in a similar way we are getting electricity from the electrical power grid, today; that's where the term Grid is coming from: Ian Foster and Karl Kesselman used it in their famous book [3]. However, there are differences to the electrical power which drives our appliances. We don't get light, heat, cleaned dishes and linen, and cold beer from the wall socket (how nice would that be); we are just getting the necessary electrical power to light, heat, clean, and to refrigerate. In contrast, in the future, complete IT services will be delivered to our homes, or to the cell phone or PDA while we are away from home. Your application doesn't have to sit in your PC anymore, as a piece of licensed software, but it will come as a grid-enabled service, from an external service provider. In contrast to the electrical power paradigm, the IT Grid provides the whole service to the wall socket. The computer, the software tools, the applications and tedious maintenance will disappear from our surroundings and will be provided as a service from experts specialized in building and operating data centers at low cost, and from those specialized on applications and services.

This sounds revolutionary. However, all the necessary building blocks are available today: fast networks ('fat pipes') interconnecting compute and storage nodes, intelligent middleware matching user requirements with available and suitable resources, numerous 'grid-enabled' application software, and client software (browsers and portals) which guarantee secure, transparent and remote access to resources, applications and data. Thus, complementing the Web, the Grid offers the platform for data collection, processing and cataloguing, in a distributed environment. While the Web, until today, is kind of a modern, dynamic, omniscient, multimedia book which is permanently updated by the community, the Grid will enhance the Web with capabilities such as the global workplace of the future where we will collaboratively collect and process the data, transform them into information and knowledge, and present it as services on the Web.

3 How does a Grid function?

Simply speaking, grid middleware interconnects all the distributed resources in a network. Light-weight software sensors, often called daemons or agents, reside within each resource, monitoring its status, providing resources with work, and reporting their status back to the "supervisor" software and then to the user. This supervising "resource broker" hosts a catalogue (similar to the 'yellow pages') with information about characteristics, availability and status of the resources and their workload, and it distributes (schedules) work to suitable resources according to predetermined policies and service level agreements (SLAs).

Other components of the grid services architecture securely manage applications and data; authenticate and authorize users; provide access to distributed resources, applications, and data; and deliver the results of a computation or a database query back to the user. All interfaces between these heterogeneous components (services) are standardized, e.g. via the Web Services Resource Framework (WSRF, [4]) and thus enable full interoperability among them.

4 Benefits of Grid Computing for Research and Industry

Grid infrastructures provide a wide spectrum of benefits: transparent access to and better utilization of resources; almost infinite compute and storage capacity; flexibility, adaptability and automation through dynamic and concerted interoperation of networked resources [5]; cost reduction through utility model; higher quality of products designed and developed via grid tools; shorter time-to-market; and more. This grid revolution is already well underway in scientific and engineering organizations with high demand of computing and data processing, mostly as prototype grids, and a few already in full production, see e.g. [6] – [14]. But also (and especially) for those research departments and businesses which cannot afford powerful supercomputing resources grid computing is of great benefit.

An outstanding example for grid computing benefits in research is the data collection and processing infrastructure for the high-energy physics experiment at the European research center CERN in Geneva, the so-called Large Hadron Collider, LHC, [15], which will be operational from early 2007. Over 5000 physicists world-wide will analyze the collision events of the largest particle accelerator in the world, resulting in petabytes of data per year, which will be filtered, sorted and stored into digital repositories and accessed by scientists. This LHC Grid is built in four tiers, with tier zero being the compute servers at CERN, tier one the national research centers, tier two the servers in the local research centers and universities, and finally the desktops of the researchers.

Besides the obvious benefits for the researchers, grid technology has great benefits also for the industry, [16]. In an era of increasing dynamics, shrinking distances, and global competition, those organizations are in an advantageous position which have access either to natural or to highly specialized resources, on demand, in an efficient and effective way. Countries like Germany for example don't have enough natural resources; thus, competition has to be strengthened via specialization, e.g. an excellent education for everybody, use of highly modern tools and machines, optimized development and production processes, and highly efficient communication and sales processes. Here, grid technology can provide great benefit. It enables engineers to access any IT resource (computer, software, applications, data, etc) in an easy and efficient way, to simulate any process and any product (and even the whole product life cycle) in virtual reality before it is build, resulting in higher quality, increased functionality, and cost and risk reduction. Grid technology helps to adjust an enterprise's IT structures to real business requirements (and not vice versa). For example, global companies will be able to decompose their highly complex processes into modular components of a workflow which can be distributed around the globe such that on-demand access to suitable workforce and resources is assured, productivity increased, and cost reduced. Application of grid technology in these processes, guarantees seamless integration of and communication among all distributed components and provides transparent and secure access to sensitive company information and other proprietary assets, world-wide.

5 Grid Business and Services

From a bird's eye view, the business model for grid services will be similar to those for electrical power, water, or telephony: our payments will be based on widely agreed billing units which include cost for computers, storage, software, applications, work, electrical power, and square footage for the equipment.

In the future, we will see many different grid-based services and and an army of new service providers. For example, providers offering compute cycles or storage, such as Amazon's Elastic Computing Cloud (EC2, [17]) or its Simple Storage Service (S3, [18]). But also 'Application Service Providers' offering a specific application service to engineering firms, accessible via browser. At the end of each month, we will receive an invoice from our favorite service provider, broken down into the services which we received from the different providers, very similar to our Telekom invoice today.

The Web will be the platform for many of our future grid-based businesses. We will surf to the service provider's Web portal, login and set up a personal account. Service providers will offer any services, securely, on demand, with highest quality, at reasonable cost, according to the Service Level Agreement (SLA) negotiated with the customer.

Grid technology will also revolutionize society. Let's look at education, for example. On one hand, our knowledge is increasing exponentially – e.g. in bioinformatics it's doubling every 12 months -, on the other hand, schools can't keep pace with this exponential development, especially in the natural sciences, [19]. Grid technologies will become the fundament for novel teaching and learning tools such as virtual laboratories which enable children (and teachers) to interactively experience the secrets of nature, engineering and society. This will dramatically increase their motivation, creativity and knowledge.

All this will take a few more years to happen. Firstly, we have to 'grid-enable' our data, our applications, our knowledge repositories. We need security technologies which guarantee that one's identity can't be stolen and that confidential data can be stored in highly secure containers if needed. This requires close collaboration among computer scientists, researchers, engineers, and businesses. Now that we share specific resources because it's more efficient and fosters communication and collaboration, we have to make sure that we only pay for what WE use and that our computational results remain confidential. Good news is that thousands of experts in research and industry are working very hard on solving these problems, in hundreds of grid projects such as the ones presented herein, and on Grid standards in the Open Grid Forum, OGF [22].

6 Major Grid Initiatives

In the following we briefly summarize six of the major grid projects around the world: ChinaGrid, German D-Grid, European EGEE, Japanese NAREGI, UK e-Science Programme, and the US TeraGrid. Information for this chapter has mostly been collected from the Web. The following table presents the different phases of the projects, their funding, estimated number of experts involved, and type of users (research or industry):

Initiative	Time	Funding	People	Users
UK e-Science-I:	2001 - 2004	$180M	900	Res.
UK e-Science-II:	2004 - 2006	$220M	1100	Res. Ind.
TeraGrid-I:	2001 - 2004	$90M	500	Res.
TeraGrid-II:	2005 - 2010	$150M	850	Res.
ChinaGrid-I:	2003 - 2006	$3M	400	Res.
ChinaGrid-II:	2007 – 2010	$15M	1000	Res.

NAREGI-I:	2003 - 2005	$25M	150	Res.
NAREGI-II	2006 - 2010	$40M	250	Res. Ind.
EGEE-I:	2004 - 2006	$40M	800	Res.
EGEE-II:	2006 - 2008	$45M	1000	Res. Ind.
D-Grid-I:	2005 - 2008	$32M	220	Res.
D-Grid-II:	2007 - 2009	$35M	440	Res. Ind.

6.1 China: The ChinaGrid, [21]

In 2002, the China Ministry of Education (MoE) launched the largest grid computing project in China, called the ChinaGrid project, aiming at providing the nationwide grid computing platform and services for research and education among 100 key universities in China. The vision for the ChinaGrid project is to deploy the largest, most advanced and most practical grid computing project in China or even around the world.

Context: There are currently at least three large grid initiatives in China:

- China National Grid – CNGrid (Ministry of Science and Technology)
- China Education and Research Grid – ChinaGrid (Ministry of Education)
- China Science Grid Project (National Science Foundation)

The most widely known and perhaps advanced grid initiative is ChinaGrid with its focus on education and research. It’s funded by the Ministry of Education. The first phase was 2003-2005, with 12 key universities involved (20 universities at the end of 2004). At that time, the systems in the grid had a performance of more than 16Tflops, with 180TB storage.

The underlying infrastructure for the ChinaGrid project is the CERNET (China Education and Research Network), which began operation from 1994, covering more than 800 universities, colleges and institutes in China. Currently, it is the second largest nationwide network in China. The bandwidth of the CERNET backbone is (currently) 2.5Gbps, connecting 7 cities, called local network centers. The bandwidth of the CERNET local backbone is 155Mbps.

The focus of the first stage of ChinaGrid is on the compute grid platform and on applications (e-science). These applications are from a variety of scientific disciplines, from life science to computational physics. The second stage of ChinaGrid project is from 2007 to 20010, covering 30 to 40 key universities in China. The focus will extend from computational grid applications to information service grid (e-information), including applications for a distance learning grid, digital Olympic grid, etc. The third stage will be from 2011 to 2015, extending the coverage of the ChinaGrid project to all the 100 key universities. The focus of the third stage grid application will be even more diverse, including instrument sharing (e-instrument).

The underlying common grid computing middleware platform for the ChinaGrid project is called ChinaGrid Supporting Platform (CGSP), to support all of the above mentioned three stages: e-science, e-information, and e-instrument. CGSP integrates all kinds of resources in education and research environments, making the heterogeneous and dynamic nature of the resources transparent to the users, and providing high performance, high reliable, secure, convenient and transparent grid services to the scientific computing and engineering research communities. CGSP provides both a ChinaGrid service portal, and a set of development environments for deploying various grid applications.

The current version, CGSP 2.0, is based on Globus Toolkit 4.0, and is WSRF [4] and OGSA [3] compatible. The previous version, CGSP 1.0, has been released in October 2004. There are the following five 5 building blocks in CGSP 1.0:

1. Grid Portal: The grid portal is the entrance for the end user to use grid services. By using grid portal, users can submit their jobs, monitor the running of jobs, manage and transfer data, and inquire the grid resource information. Grid portal also provides other facilities such as user management and accounting of grid resource usage.

2. Grid Development Toolkits: to provide the resources as grid services; deployment and management toolkit for grid; and programming model to deploy complex grid applications in the grid environment.

3. Information Service: responsible for the management of various resources within the grid environment, it provides a global resource view and grid information services, and updates grid resource information in real time. The main purpose is to provide real time information of various grid resources for end users and other modules in grid environment.

4. Grid Management: provides basic support for various jobs in grid environment. It consists of four parts:
- Service container: grid service installation, deployment, running, and monitoring environment on each node in the grid environment. It also provides necessary support to monitor the status of each grid node in real time.
- Data manager: responsible for the management of various storage resources and data files. It provides a global file view, so that users can access various data files transparently.
- Job manager: based on information services and data management, it provides support for job management, scheduling, and monitoring for end users' computational tasks, so that data and resources can be accessed transparently within the grid, and enables cooperatively working among distributed resources.
- Domain manager: ChinaGrid is organized in domains. A domain refers to an independent grid system to provide services to others. A domain can be a specialized grid, or a regional grid. The domain manager is responsible for user management, logging, accounting within a domain and interacting with other domains. It enables the domain administrator to easily manage the users, services, and resources within a domain, and establish policies among domains.

5. Grid security: provides user authentication, resources and services authorization, encrypted transmission, and the mapping of users to resources authorization.

6.2 Europe: EGEE and EGEE-II, Enabling Grids for E-sciencE, [10]

EGEE-II is the second phase of a four year programme of work which started with the EGEE project. The aim of EGEE is to build on recent advances in Grid technology and develop a service Grid infrastructure which is available to scientists 24 hours a day, providing researchers in academia and industry with access to major computing resources, independent of their geographic location. The EGEE project also focuses on attracting a wide range of new users to the Grid. The project concentrates primarily on three core areas:

- The first area is to build a consistent, robust and secure Grid network that will attract and incorporate additional computing resources on demand.
- The second area is to continuously improve and maintain the middleware in order to deliver reliable services to users.
- The third area is to attract new users from industry as well as science and ensure they receive the high standard of training and support they need.

The EGEE Grid is built on the EU Research Network GÉANT and exploits Grid expertise generated by many EU, national and international Grid projects to date. In its first phase, EGEE comprised over 70 contractors and over 30 non-contracting participants, and was divided into 12 partner federations, covering a wide range of both scientific and industrial applications.

The work carried out in the first phase of the project was organized into 11 activities, and with funding of over 30 million Euro from the European Commission (EC), the project was one of the largest of its kind. The first phase provided the basis for assessing subsequent objectives and funding needs, and gave way to a second phase which started on 1 April 2006. This project saw its consortium grow to over 90 contractors and a further 48 non-contracting participants in 32 countries, and its funding levels to over 36 million Euro from the EC. It maintains its organizational structure into geographical federations and comprises 10 activities.

The initial focus of the project was on two application areas, namely High Energy Physics (HEP) and Biomedicine. The rationale behind this was that these fields were already grid-aware and would serve well as pilot areas for the development of the various EGEE Grid services.

The number of applications has grown over the project's lifetime to include, among others, Earth Sciences, Astro-particle Physics, Computational Chemistry, Drug Discovery, Hydrology and Cosmology. In addition, there are several applications from the industrial sector running on the EGEE Grid, such as applications from geophysics and the plastics industry. At present there are more than 20 applications from 9 domains on the EGEE Grid infrastructure:

- Astrophysics - MAGIC
- Computational Chemistry
- Earth Sciences - Earth Observation, Solid Earth Physics, Hydrology, Climate
- Financial Simulation - E-GRID
- Fusion
- Geophysics - EGEODE
- High Energy Physics - 4 LHC experiments (ALICE, ATLAS, CMS, LHCb), BaBar, CDF, DØ, ZEUS
- Life Sciences - Bioinformatics (Drug Discovery, GPS@, Xmipp_MLrefine, etc.), Medical imaging (GATE, CDSS, gPTM3D, SiMRI 3D, etc.)
- Multimedia

The EGEE project now provides a stable and reliable Grid infrastructure with its own middleware stack, gLite. EGEE began work using the LCG-2 middleware, provided by the LCG project (which is itself based on the middleware from EU DataGrid, EGEE's predecessor). In parallel it produced the gLite middleware, using reengineered components from a number of sources to produce lightweight middleware that provides a full range of basic Grid services, part of which is based on Globus version 2.4. As of September 2006, gLite is at version 3.0, and comprises some 220 packages arranged in 34 logical deployment modules.

Who uses the EGEE Grid Infrastructure? Generally, the EGEE Grid infrastructure is perfect for any complex scientific researchwhich cannot be done on a single PC or that perhaps needs calculations that would take impractical lengths of time to compute even in a dedicated cluster of PC's. The EGEE Grid consists of over 20,000 CPU available to users 24 hours a day, 7 days a week, in addition to about 10 Petabytes (10 million Gigabytes) of storage, and maintains on average 20,000 concurrent jobs.

More than a thousand scientists from all over the world submit over 2 million jobs per year. These jobs vary from very short batches to large single computations, and their results are relating to any one of the scientific areas already mentioned.

6.3 Germany: The D-Grid Initiative, [20]

In 2003, German scientists and scientific organizations started the D-Grid initiative [20], jointly publishing a strategic paper in July 2003. This paper examined the status and consequences of grid technology on scientific research in Germany and recommended a long-term strategic grid research and development initiative. This resulted in the German e-Science Initiative founded by the German Ministry for Research and Education (BMBF) in March 2004, together with a call for proposals in the areas of Grid Computing, e-Learning, and Knowledge Management. In November 2004, the BMBF presented the vision of a new quality of digital scientific infrastructure which will enables our globally connected scientists to collaborate on an international basis; exchange information, documents and publications about their research work in real time; and guarantee efficiency and stability even with huge amounts of data from measurements, laboratories and computational results.

The e-Science Initiative and the first phase of D-Grid started on September 1, 2005. BMBF is funding over 100 German research organizations with 100 Million Euro over the next 5 years. For the first 3-year phase of D-Grid, financial support is approximately 25 Million Euro. The goal is to design, build and operate a network of distributed, integrated and virtualized high-performance resources and related services to enable the processing of large amounts of scientific data and information. The Ministry for Research and Education is funding the assembling, set-up and operation in several overlapping stages:

1. D-Grid 1, 2005-2008: IT services for scientists, designed and developed by the ‘early adopters’ of the computer science community. This global services infrastructure will be tested and used by so-called Community Grids in the areas of high-energy physics, astrophysics, medicine and life sciences, earth sciences (e.g. climate), engineering sciences, and scientific libraries.

2. D-Grid 2, 2007-2009: IT services for scientists, industry, and business, including new applications in chemistry, biology, drug design, economy, visualization of data, and so on. Grid services providers will offer basic services to these users.

D-Grid 3 (around 2008- 2010) will extend the grid infrastructure with an SLA and a knowledge management layer, and adding several virtual competence centres, encourage global service-oriented architectures in the industry, and use this grid infrastructure for the benefit of our whole society, as discussed in chapter 6.7.

D-Grid consists of the DGI Infrastructure project and (currently) the following seven Community Grid projects:

- AstroGrid-D (Astronomy)
- C3-Grid (Earth Sciences)
- HEP Grid (High-Energy Physics)
- InGrid (Engineering)
- MediGrid (Medical Research)
- TextGrid (Scientific Libraries, Humanities)
- WISENT (Knowledge Network Energy Meteorology)

Short-term goal of D-Grid is to build a core grid infrastructure for the German scientific community, until the end of 2006. Then, first test and benchmark computations will be performed by the Community Grids, to provide technology feedback to DGI. Then, climate researchers of the C3-Grid, for example, will be able to predict climate changes faster and more accurately than before, to inform governments about potential environmental measures. Similarly, astrophysicists will be able to access and use radio-telescopes and supercomputers remotely via the grid, which they wouldn't be able to access otherwise, resulting in novel quality of research and the resulting data.

The D-Grid Infrastructure Project

Scientists in the D-Grid Infrastructure project DGI, [23], are developing and implementing a set of basic grid middleware services which will be offered to the other Community Grids. For example, services include access to large amounts of data distributed in the grid, the management of virtual organizations, monitoring and accounting. So far, a core-grid infrastructure has been built for the community grids for testing, experimentation, and production. High-level services will be developed which guarantee security, reliable data access and transfer, and fair-use policies for computing resources. This core-grid infrastructure will then be further developed into a reliable, generic, long-term production platform which can be enhanced in a scalable and seamless way, such as the addition of new resources and services, distributed applications and data, and automated "on demand" provisioning of a support infrastructure.

An important aspect in every grid is security, especially with the industry expected to join soon, such as automotive and aerospace. Therefore, an important DGI work package is "Authentication and Authorization" [24]. It's obviously important to know that a user is really the one she pretends to be, and that she is authorized to access and use the requested resources and information. While enterprise grids are mostly operating behind firewalls, global community grids use security technology like VOMS, Virtual Organization Membership Service, [25]. However, building and managing so-called Certificate Authorities is still a very cumbersome activity to date.

The following D-Grid DGI infrastructure services are available for the current and D-Grid 2 community projects, at the end of 2006:

- The core D-Grid infrastructure offers central grid services. New resources can be easily integrated in the help-desk and monitoring system, allowing central control of resources to guarantee sustainable grid operation.
- DGI offers several grid middleware packages (gLite, Globus und Unicore) and data management systems (SRB, dCache und OGSA-DAI). A support infrastructure helps new communities and "Virtual Organizations" (VOs) with the installation and integration of new grid resources via a central Information Portal („Point of Information“). In addition, software tools for managing VOs are offered, based on the VOMS and Shibboleth [25] systems.
- Monitoring und Accounting prototypes for distributed grid resources exist, as well as an early concept for billing in D-Grid.
- DGI offers consulting for new Grid Communities in all technical aspects of network and security, e.g. firewalls in grid environments, alternative network protocols, and CERT (Computer Emergency Response Team).
- DGI partners operate „Registration Authorities“ to support simple application of internationally accepted Grid Certificates from DFN (German Research Network organization) and GridKA (Grid Project Karlsruhe). DGI partners support new members to build their own „Registration Authorities“.
- Core D-Grid is offering resources for testing, via middleware systems (gLite, Globus and UNICORE). The Portal Framework Gridsphere serves as the user interface. Within the D-Grid environment the dCache system takes care of the administration of large amount of scientific data.

Community Grid Projects

- The High-Energy Physics (HEP, [26]) Grid community is developing applications and components for evaluating terabytes of data from large high-energy physics experiments, including the Large Hadron Collider at CERN.
- ASTRO Grid [27] combines research institutions in astronomy and astrophysics into a single, nationwide virtual organization for distributed collaboration and integration of distributed astronomical data archives, instruments and experiments.
- MEDI Grid [28] represents the medical and bio-informatics community in Germany. It focuses on application scenarios in medical image processing, bioinformatics, and clinical research, and their interaction.
- C3-Grid [29] for the Collaborative Climate Community has the goal to develop a highly proficient grid-based research platform for the German earth-system research community to efficiently access and analyze distributed, high-volume scientific data from earth-system modeling and observation.
- For InGrid [30], the Industry applications grid project, a grid framework will be developed to enable modeling, optimization, and simulation of engineering applications from areas such as foundry technology, metal forming, groundwater flows, turbine simulation, and fluid-structure interaction.
- The TextGrid project [31] is developing tools and standard interfaces for publication software, modules for scientific text processing and editing, and administration and access to distributed data and tools on the grid.
- Finally, WISENT [32] is developing tools and methods in the area of energy meteorology to accurately forecast energy usage to be matched with energy provisioning on demand.

In the first half of 2005, an important initial milestone was achieved for D-Grid -- all participating scientific institutions collaborated in nine different working groups, eventually agreeing on technologies and procedures in the areas of resource management and monitoring, user portals and interfaces, storage management and archiving, metadata and data management, license management, accounting, virtual organizations, security, privacy, basic grid services, and business models, and reliability and sustainability. This was indeed a promising beginning, considering the heterogeneity of the different projects among multiple organizations, each with its own agenda. In the past, organizations strived to successfully collaborate within their own communities. But in D-Grid, for the first time ever, all these different communities are working together on a single, inter-community grid middleware platform to share computing resources, middleware tools, applications, and expertise. This will result in an IT infrastructure which is interoperable with other international grids, scalable and extensible for more community grids in the future, and available for all of our scientists for national and international collaboration.

The Next Phase: D-Grid 2, 2007 - 2009

In April 2006, the German Ministry of Education and Research published the second D-Grid call, to fund IT services which will be offered on top of the D-Grid infrastructure to research and industry. These "Service Grids" are characterized as follows::

- They offer specific services to communities, institutes, teams, in a distributed way.
- They utilize and complement the D-Grid middleware and take international standards into account.
- Established or new service providers will offer these services in a professional and utilitarian way.
- They will focus on sustainable and reliable business models.

D-Grid 2 will complement the current infrastructure and application projects and strengthen reliability and sustainability of the technology developments and services. In addition, new communities in research and industry as well as service providers shall be integrated into D-Grid, enabling national and international collaboration, finally leading to a new IT-services market for research and industry.

The following areas will be funded in D-Grid 2:

1. Generic ("horizontal") Service Grids: targeting at widely usable grid services for heterogeneous user groups or service providers (IT industry, compute and data centers), including corresponding business models.

2. Community ("vertical") Service Grids: funding new community grids which will use the existing D-Grid middleware and services. The gaps between existing core services, generic service grids, and the specific requirements of individual communities will be closed by service providers which develop and offer these services. This, new communities don’t have to develop specific services themselves, instead they just specify there requirements to their service providers.

3. Filling D-Grid Gaps: necessary enhancements of the current D-Grid infrastructure will be funded, which will lead to improved or new grid services (e.g. billing, pricing, security).

In December 2006, 15 new projects have been selected, focusing on new applications and services. They will start around mid of 2007

6.4 Japan: The NAREGI Project, [7]

The main objective of NAREGI is to research and develop grid middleware according to global standards to a level that can support practical operation, to implement a large-scale computing environment (the Science Grid) for widely-distributed, advanced research and education. NAREGI is carrying out R&D from two directions: through the grid middleware R&D at the National Institute of Informatics (NII), and through an applied experimental study using nano-applications, at the Institute for Molecular Science (IMS). These two organizations advance the project in cooperation with industry, universities and public research facilities.

The National Research Grid Initiative NAREGI was created in 2003 by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). From 2006, under the “Science Grid NAREGI” Program of the “Development and Application of Advanced High-performance Supercomputer project “ being promoted by MEXT, research and development is continuing to build on current results, while expanding in scope to include application environments for next-generation, peta-scale supercomputers.

A large number of research bodies from academia and industry are participating in this program, with research and development of grid middleware centered at the National Institute of Informatics (NII), and empirical research into grid applications being promoted by the Institute for Molecular Science (IMS). Also, in order to promote use of grid technology in industry, the Industrial Committee for Super Computing Promotion gathers research topics from industry and promotes collaborative work between academic and industrial research bodies. The results of this research will support construction of the Cyber Science Infrastructure (CSI), which is the academic research base being promoted by NII, as well as construction of the peta-scale computing environment for scientific research. Through this, NAREGI will accelerate research and development in scientific fields, improve international cooperation, and strengthen competitiveness in an economically effective way.

The middleware being developed by NAREGI will present heterogeneous computation resources, including supercomputers and high-end servers connected by network, to users as a single, large, virtual computing resource. In order to build a global grid, the middleware is being developed according to the Science Grid environment standards specifications from the Open Grid Forum (OGF, [22]).

The National Institute for Informatics (NII) is promoting the construction of the Cyber Science Infrastructure (CSI), which is the base for next-generation academic research. A core technology of CSI is the science grid environment, and it will be made up of academic data networks like SuperSINET.

The infrastructure provides a user-friendly environment to the user, who can then focus on his/her computational science research without concern for the scale of computing resources or environment required. High-throughput processing and meta-computing can be applied to large-scale analysis using the grid, allowing the supercomputers to be used to their maximum capabilities.

This environment allows multi-scale/multi-physics coupled simulations, which is becoming very important in computational sciences, in a heterogeneous environment. Resource allocation is suited to each application, so that coupled analysis can be done easily. Virtual Organizations (VOs), separate from the real organizations to which researchers and research bodies belong, can be formed dynamically on the Grid according to the needs of the research community. The project hopes to promote research and development by forming this kind of optimized research community where researchers can share and cooperate with data and resources.

In 2003, NAREGI developed a component technology based on UNICORE, and in 2004, released an alpha-version prototype of middleware based on UNICORE to test integrated middleware functions. In 2005, research and development was advanced on beta-version grid middleware, based on newly-established OGSA specifications, [2], to align with global activity. This beta version was released as open-source software in May 2006, and included enhanced functions supporting virtual organizations. In 2007, NAREGI Version 1.0, based on this beta version, will be released. From 2008, the scope of research and development will be expanded to include application environments for next-generation supercomputers, and the results of this will be released as NAREGI Version 2.0 in 2010.

6.5 UK: The e-Science Program, [13]

The UK e-Science program was proposed in November 2000 and launched the following year. The total funding for the first phase was $240M with a sum of $30M allocated to a Core e-Science Program. This was an activity across all the UK's Research Councils to develop and broker generic technology solutions and generic middleware to enable e-Science and to form the basis for new commercial e-business software. This $30M funding was enhanced by an allocation of a further £$40M from the Department of Trade and Industry which was required to be matched by equivalent funding from industry. The Core e-Science Program, which is managed by the UK Engineering and Physical Science Research Council (EPSRC) on behalf of all the Research Councils, is therefore the generic part of e-Science activities within the UK and thus ensured a viable infrastructure and coordination of the national effort.

The first phase of the Core e-Science Program (2001 – 2004) was structured around six key elements:

1) A National e-Science Center linked to a network of Regional e-Science Grid Centers
2) Generic Grid Middleware and Demonstrator Projects
3) Grid Computer Science based Research Projects

4) Support for e-Science Application Pilot Projects
5) Participation in International Grid Projects and Activities
6) Establishment of a Grid Network Support Team

The National Center, with funding of $6 million plus infrastructure funds, acts as the national focus point for grid computing, data resources and facilities. Like the Regional e-Science Centers it initiates projects with industrial partners investing a matching level of funding. It established an e-Science Institute with an international, multidisciplinary research seminar program and is managing a 'network of excellence' in grid technologies across the UK. More information can be found under local activities below.

The e-Science Core Program invited proposals for the large scale development of applications which would substantially test and exercise generic e-Science middleware; these were projects with significant industrial involvement. Also, eleven smaller demonstrator projects were funded to demonstrate the potential of the grid in a range of disciplines.

The EPSRC funded three, 6 year, computer science (CS) oriented, Interdisciplinary Research Collaborations (IRCs). These are major projects that fund key CS research groups from a number of universities to undertake long-term research which has relevance to issues in software development and middleware. Thus the Core Program funded the IRCs to develop an e-Science research agenda. The IRC projects include:

- Advanced Grid Interfaces for Environmental e-Science in the laboratory and in the field
- CoAKTinG: Collaborative Advanced Knowledge Technologies in the Grid
- Grid enabled knowledge services: collaborative problem solving environments in medical informatics
- Grid-Based Medical Devices For Everyday Health
- MIAS - Grid. A Medical Image and Signal Research Grid

To ensure that researchers developing e-Science applications are properly supported, especially in the initial stages, the Grid Support Center was established. The UK Grid Support Center (see local activities) supports all aspects of the deployment, operation and maintenance of grid middleware and distributed resource management for the UK grid test-beds. The Grid Network Team (GNT) works with application developers to help identify the network requirements and help map these on to existing technology. It also considers the long-term networking research issues required by the grid.

Due to the importance of actively communicating and collaborating with international community, a number of activities were initiated including the GridNet project (further details under international activities).

The second phase of the Core e-Science Program (2004 -2006) is based around six key activities:

- A National e-Science Center linked to a network of Regional e-Science Centers,
- Support activities for the UK e-Science Community,
- An Open Middleware Infrastructure Institute (OMII),
- A Digital Curation Center (DCC),
- New Exemplars for e-Science,
- Participation in International Grid Projects and Activities.

Of particular significance in the second phase are the OMII and DCC. The Open Middleware Infrastructure Institute (OMII) is an institute based at the University of Southampton, located in the School of Electronics and Computer Science. The vision for the OMII is to become the source for reliable, interoperable and open-source grid middleware, ensuring the continued success of grid-enabled e-Science in the UK. OMII intends to:

- Create a one-stop portal and software repository for open-source grid middleware, including comprehensive information about its function, reliability and usability;
- Provide quality-assured software engineering, testing, packaging and maintenance of software in the OMII repository, ensuring it is reliable and easy to both install and use;
- Lead the evolution of grid middleware at international level, through a managed program of research and wide-reaching collaboration with industry.

The Digital Curation Center (DCC) supports UK institutions with the problems involved in storing, managing and preserving vast amount of digital data to ensure its enhancement and continuing long-term use. The purpose of this DCC is to provide a national focus for research into curation issues and to promote expertise and good practice, both nationally and internationally, for the management of all research outputs in digital format. The DCC is based at the University of Edinburgh.

In addition to the UK e-Science program, there have been UK initiatives in the social sciences and the arts and humanities: The National Centre for e-Social Science (http://www.ncess.org/) has begun on an ambitious programme of developing e-Social Science tools and evaluating their social implications. Further, there is now an Arts and Humanities e-Science Support Centre (http://www.ahessc.ac.uk/) which is creating a community around the uses of e-science in, for example, history and linguistics.

As a result of this initiative the UK e-Science program has enjoyed a number of strengths including:

- An Advanced National Grid Infrastructure, which was built specifically for use with grid computing. The National Grid Service (NGS) is one of the facilities available to UK researchers which provides access to over 2000 processors, and over 36 TB of “data-grid” capacity.
- Availability of Funding: new research and industrially related funding from the UK government and different funding bodies. Over $500M has been invested in the e-Science program over the last five years. This has been followed by smaller-scale funding more recently for e-social science and e-research in arts and humanities
- Industrial involvement: Over a 100 companies are involved in UK e-Science projects including IBM, Intel, Oracle, and Sun, and a vast number of other national and international industries in different domains ranging from finance to pharmacy.
- The UK has extended its e-science capability to include not only the sciences and engineering, but also social sciences and arts and humanities, which will provide benefits across the academic community
- New research advances: Large scale multidisciplinary teams of scientist have worked together and made advances in a wide range of disciplines.

6.6 US: TeraGrid

The US TeraGrid, [6], is an open scientific discovery infrastructure combining leadership class resources at nine partner sites to create an integrated, persistent computational resource. Using high-performance network connections, the TeraGrid integrates high-performance computers, data resources and tools, and high-end experimental facilities around the country.

TeraGrid is coordinated through the Grid Infrastructure Group (GIG) at the University of Chicago, working in partnership with the Resource Provider sites: Indiana University, Oak Ridge National Laboratory, National Center for Supercomputing Applications, Pittsburgh Supercomputing Center, Purdue University, San Diego Supercomputer Center, Texas Advanced Computing Center, University of Chicago/Argonne National Laboratory, and the National Center for Atmospheric Research.

Terascale Initiatives 2000-2004: In response to the 1999 report by the President's Information Technology Advisory Committee, NSF embarked on a series of "Terascale" initiatives to acquire: (1) computers capable of trillions of operations per second (teraflops); (2) disk-based storage systems with capacities measured in trillions of bytes (terabytes); and (3) networks with bandwidths of billions of bits (gigabits) per second. In 2000, the $36 million Terascale Computing System award to PSC supported the deployment of a computer (named LeMieux) capable of 6 trillion operations per second. When LeMieux went online in 2001, it was the most powerful U.S. system committed to general academic research.

In 2001, NSF awarded $45 million to NCSA, SDSC, Argonne National Laboratory, and the Center for Advanced Computing Research (CACR) at California Institute of Technology, to establish a Distributed Terascale Facility (DTF). Aptly named the TeraGrid, this multi-year effort aimed to build and deploy the world's largest, fastest, most comprehensive, distributed infrastructure for general scientific research. The initial TeraGrid specifications included computers capable of performing 11.6 teraflops, disk-storage systems with capacities of more than 450 terabytes of data, visualization systems, data collections, integrated via grid middleware and linked through a 40-gigabits-per-second optical network.

In 2002, NSF made a $35 million Extensible Terascale Facility (ETF) award to expand the initial TeraGrid to include PSC and integrate PSC's LeMieux system. Resources in the ETF provide the national research community with more than 20 teraflops of computing power distributed among the five sites and nearly one petabyte (one quadrillion bytes) of disk storage capacity.

In 2003, NSF made three Terascale Extensions awards totaling $10 million in 2003, to further expand the TeraGrid's capabilities. The new awards funded high-speed networking connections to link the TeraGrid with resources at Indiana and Purdue Universities, Oak Ridge National Laboratory, and the Texas Advanced Computing Center at The University of Texas, Austin. Through these awards, the TeraGrid put neutron-scattering instruments, large data collections and other unique resources, as well as additional computing and visualization resources, within reach of the nation's research and education community.

In 2004, as a culmination of the DTF and ETF programs, the TeraGrid entered full production mode, providing coordinated, comprehensive services for general U.S. academic research.

The TeraGrid 2005-2010: In August 2005, NSF's newly created Office of Cyberinfrastructure extended support for the TeraGrid with a $150 million set of awards for operation, user support and enhancement of the TeraGrid facility over the next five years. Using high-performance network connections, the TeraGrid now integrates high-performance computers, data resources and tools, and high-end experimental facilities around the country. As of early 2006, these integrated resources include more than 102 teraflops of computing capability and more than 15 petabytes (quadrillions of bytes) of online and archival data storage with rapid access and retrieval over high-performance networks. Through the TeraGrid, researchers can access over 100 discipline-specific databases. With this combination of resources, the TeraGrid is the world's largest, most comprehensive distributed cyberinfrastructure for open scientific research.

6.7 The Future: What Comes After These Grid Projects ?

How will the Internet evolve under the influence of these new grid technologies? It will certainly take another few years until we see the next-generation Internet which allows access to compute resources and services as easily as the access to billions of Web sites today. For this to happen we have to continue to improve the new e-infrastructure in projects such as the ones mentioned herein, to fully benefit from the availability of vast amount of resources and services in a transparent way. In my interviews, I have collected a few thoughts on a potential roadmap for research, industry and society to achieve this goal:

Research:

- Development of user-friendly and automated grid infrastructure building blocks with standard interfaces to easily build local and special grids (e.g. campus grids in universities) and global grids for international research projects, to collaboratively use resources distributed in the Internet, such as computers, storage, applications, and data.
- Adaptation of application software for grid infrastructure and services, in areas like physics, chemistry, biology, weather, climate, environment, bioinformatics, medicine, aero- and fluid mechanics, oil and gas, economy, finance, and so on.
- Participation and contribution to standardization organizations such as OGF [22], OASIS [33], and W3C [34], and to European organizations such as ESFRI [35] and e-IRG [36].
- Development of training material and organization of training courses to learn how to build, operate and use grid infrastructures.
- Encourage independent grid resource and application service providers, developing new operational and accounting models, utility computing, and service level agreements.
- Integration of local, national, and community grids into European and international grid infrastructures.
- Overcome mental, legal and regulatory barriers, via case studies, demonstrators, and pilot projects.

Industry and Business:

- Development of new enterprise IT infrastructures based on OGSA (Open Grid Services Architecture) and SOA (Service Oriented Architecture), with SLOs (Service Level Objectives) and SLAs (Service Level Agreements) to mapping business processes to resource and application usage in an enterprise.
- Global enterprise grids to network all resources of globally distributed subsidiaries and branches, and for seamless integration of new companies after merger or acquisition.
- Close collaboration with research to efficiently transfer reliable global grid technology to the industry.
- Partner grids for close collaboration with business partners and suppliers, to optimize distributed product development, complex workflows for multi-disciplinary processes and applications, productivity and quality improvement through global “Six Sigma” processes.
- Sensor Grids und Wireless Grids, to enable communication and interaction of electronic devices e.g. for safety reasons in airplanes, cars, bridges, skyscrapers, etc.
- Development of local and global training grids to support active and interactive, flexible and dynamic education of enterprise personnel .

Society:

- Development of grids for the masses, in areas such as healthcare (illness, fitness, sensor-based monitoring of bodily functions), leisure (multi-player games, digital entertainment, sports), education (life-long learning, school grids, digital interactive laboratories), and work (Internet-based courses, online training, global teamwork, collaboratories).
- Starting with pilot projects in these areas, partnering with end-users (consumers), application service providers, and resource providers.

- Utilization of grid resources and services for education in schools, universities, and in enterprises. Integration of grid resources and interactive applications and simulations into existing curricula to dramatically improve motivation and creativity of the learners (and teachers).
- Development of personal digital assistants including technology and service infrastructure for the mass market.
- Integration of these new applications for the masses into user-friendly browser-based web portals.

In the near future, on an enhanced Internet, all kinds of service providers will offer their services for computing, data, applications, and many more. On an enhanced World Wide Web, via secure Web Portals, we will access grid components like Lego building blocks, which enable us to dynamically build grids ‘on the fly’, according to our specific needs. We will rent or lease the resources required and pay for what we use or on a subscription basis. We still might have our own resources, to fulfill a certain basic need, or for highly proprietary applications and data, which can be extended in a seamless way with resources from service providers, available on the grid. But, as already said, this will still take a few years.

As with any new infrastructure, development and deployment of the next Internet generation will require vision and endurance. We have to work continuously on strategic, long-term projects on a national or international scale, which demand collaboration of research and industry on complex inter-disciplinary projects, and which will enable and improve the tools of our scientists, business people, and educators and strengthen our position in the international competition.

7 Lessons Learned

In the following, we summarize the most important results and lessons learned from the grid projects analyzed and from the interviews, starting with the positive ones:

- Most of the successful projects in the early days had a strong focus on just one topic (middleware or application).
- It is very important that application scientists closely collaborate with computer scientists.
- Successful projects were often more pragmatic, with one strong focus and only a few selected aspects and requirements.
- Successful projects were mostly application and user driven, with a focus on the development on standard and commodity components, open source, and results easy to understand and to use.
- Professional service centers proved successful. E.g. in the UK, National Grid Service (NGS), Grid Operation Support Center (GOSC) and Open Middleware Institute (OMII) are extremely important factors to guaranty sustainability of the project results.
- Application-oriented and grid-enabled workflows and the separation of middleware and application layer helped the projects to deliver more sustainable results, and usability and integration became relevant.
- The newly introduced paradigm shift in resource funding in the UK, from the 46% lump sum model to the Full Economic Cost model, might better support the transition to utility computing.

However, there were also problems and challenges, especially with the early initiatives and projects:

- There was a lot of hype, especially in 2001 and 2002, and therefore high expectation in the projects and their results.

- There was a high risk often with projects which focused on both applications and infrastructure. Later projects either focused on the infrastructure with the applications as a driver, or focused on the application using existing core grid building blocks.
- Almost all projects in the early days developed their own infrastructure because the middleware in the early days (e.g. Globus, Condor, SRB, with new releases every 6 – 12 months) turned out to be immature. Middleware developed in these projects was often proprietary.
- In the early days, an integration of individual projects into a larger community or environment was not yet possible.
- One of the main reasons why some early projects failed was the then sudden change in 2003 from the classical, more proprietary grid technology approach to standard web services.
- Missing software engineering methods and especially low usability resulted in low acceptance of project results.
- The user point-of-view is paramount – a “build it and they will come approach” will not work. It is important to work with the user communities to ensure the resulting system is of a general nature and not limited in scope to a small number of applications.
- A lot of the grid middleware currently promoted is really intended for research and demonstrations but needs significant effort to be made suitable for large-scale production usage.
- Standards are evolving slowly and it is likely that initiatives to improve inter-operability between existing grids will produce meaningful results of benefit to the user communities on a shorter time scale. The experience gained with this inter-operability work will help identify the highest-priority points for standardization as well as a meaningful way to test if candidate standards can be implemented and deployed.
- It is challenging (but important) to establish an environment of constructive competition such that good ideas and performance are recognized and rewarded. There are still many areas where the “captive user” approach is viewed as a competitive advantage.

8 Recommendations (for future grid projects)

In this paragraph, we summarize major results and conclusions from the ‘lessons learned’, and present recommendations especially for those who intend to start or fund a grid initiative. Some of the recommendations seem trivial, but they strictly result from our analysis and findings and from the evaluation of the interviews:

- In any grid project, during development as well a during operation, the core grid infrastructure should be modified/improved only in large time cycles because applications and user environment might depend on this infrastructure. For example, if a project starts today, one recommendation could be to start building the grid infrastructure with Globus Toolkit 4 (currently 4.0.3) because we don’t expect too many changes anymore in the core of this technology, after the tremendous changes over the last 3 years.

- Of deep concern are differences and incompatibilities in the open source and commercial versions of grid technology. This has been experienced in the past for example with the distributed resource management system PBS and its commercial version PBS-Pro. The same is true for the SRB Storage Resource Broker from the San Diego Supercomputer Center and its commercial version from Nirvana.

- Continuity and sustainability especially for the infrastructure part of grid projects are extremely important. Therefore, additional funding should be available also after the end of the project, to guarantee service and support and continuous improvement and adjustment to new developments.

- Close collaboration in the grid development phase between the grid infrastructure developers and the application developers is mandatory for the applications to utilize the core grid services of the infrastructure and to avoid application silos.

- New application grids (community grids) should utilize the (existing) components of a generic grid infrastructure to avoid re-inventing wheels and building of silos.

- The infrastructure building block should be user-friendly to enable easy adoption for new (application) communities. In addition, the infrastructure group should offer an installation, operation, support and training services.

- Centers of Excellence should specialize on specific services, e.g. middleware development and maintenances, integration of new communities, grid operation, training, utility services, etc.

- We recommend to implement utility computing only in small steps, starting from enhancing existing service models moderately, and testing utility models first as pilots. Very often, today's existing government funding models are counter-productive when establishing new and efficient forms of utility services.

- After a generic grid infrastructure has been built, projects should focus first on one or only a few applications or specific services, to avoid complexity and re-inventing wheels.

- Usage of software components from open-source and standards initiatives is highly recommended to enable interoperability especially in the infrastructure and application-oriented middleware layer.

- For interoperability reasons, focus on software engineering methods especially for the implementation of protocols and the development of standard interfaces is important.

- In case of more complex projects, e.g. consisting of an integration and several application or community projects, a strong management board should steer coordination and collaboration among the projects and the working groups. The management board (Steering Committee) should consist of leaders of the different sub-projects.

- Participation of industry should be industry-driven. A push from the outside, even with government funding, doesn't seem to be promising. Success will come only from natural needs e.g. through already existing collaborations with research and industry, as a first step.

- For new grid projects, we recommend a close collaboration among grid-experienced computer scientists who build the (generic) grid infrastructure and the driving users who define their set of requirements for the grid infrastructure services.

- Application communities shouldn't start developing a core infrastructure from scratch on their own, but should - together with grid-experienced computer scientists - decide on using and integrating existing grid building blocks to avoid building proprietary application silo architectures and to focus more on the real applications.

- To encourage new communities to build/join/use grids, the grid infrastructure and especially the grid access portal have to be extremely user-friendly. The grid infrastructure group has to provide installation and operation support and training.

In their early stage, grid projects need enough funding to get over the early-adopter phase into a mature state with a rock-solid grid infrastructure such that other communities can join easily. We estimate this funding phase currently to be in the order of 3 years, with more funding in the beginning for the grid infrastructure, and later more funding for the application communities. Included in such a grid infrastructure funding are Centers of Excellence for building, managing and operating grid centers, for middleware tools, application support, and for training. Thus, parallel developments with re-inventing wheels can be avoided and funding efficiently spent.

- Building grids to provide utility computing support and services is still risky because of missing experience in this field and its mental, legal and regulatory barriers. We recommend to develop transitional strategies to migrate from existing resource usage models towards utilitarian models, in small steps and as pilot projects.

- Today's funding models in research and education are often project based and thus not ready for a utilitarian approach where resource usage is based on a pay-as-you-go approach. Old funding models first have to be adjusted accordingly before a utility model can be introduced successfully.

- Grid projects should have clear focus and objectives to avoid too diverse and diverging activities and to reduce complexity. Applications and users should be the key drivers, not the infrastructure.

- Success, especially in early-stage technology projects, is strongly proportional to the personality and leadership capabilities of the leaders.

- For several good reasons, industry in general is still in a wait-state with building and applying global grids, demonstrated by the moderate success so far in existing industrial global grid initiatives around the world. We recommend to closely work with the industry to develop appropriate funding and collaboration models which take into account the different technological, mental and legal requirements when adjusting the existing research community oriented approaches, ideally starting with already existing and successful research-industry collaborations.

- Try to study and/or use an existing grid if possible and connect your own resources once you are convinced of its value.

- If there are good reasons to create your own grid (on a university campus or in an enterprise) rather than join an existing one, better start with cluster based cycle savaging and when the users and their management are convinced of the value of sharing resources then extend the system to multiple-sites.

- Focus on understanding your user community and their needs. Invest in a strong communication/participation channel towards the leaders of that group to engage.

- Learn/keep up with what your peers have done/are doing. There is much useful experience to learn from partners.

- Instrument your services so that you collect good data about who is using which services and how. Analyze this data and learn from watching what's really going on, in addition to what users report as happening.

- Plan for an incremental approach and lots of time talking out issues and plans. Social effects dominate in non-trivial grids.

9 Acknowledgements

As always with this kind of substantial work, many colleagues from our community have contributed to this report. Therefore, I deeply appreciate the contributions from

- UK-e-Science: Tony Hey, Steven Newhouse, Carole Goble, Malcolm Atkinson, John Darlington, Trevor Cooper Chadwick, Monica Schraefel, Luc Moreau, Paul Watson, Aaron Turner, Ralph Schroeder
- EGEE: Bob Jones, Dieter Kranzlmueller, Anna Cook
- Naregi: Kazushige Saga, Satoshi Matsuoka, Kenichi Miura
- TeraGrid: Charlie Catlett, Dane Skow
- ChinaGrid: Hai Jin
- D-Grid: Uwe Schwiegelshohn, Wolfgang Guerich, Klaus Ullmann, Klaus Peter Mickel, Matthias Steinmetz, Matthias Kasemann, Wolfgang Hiller, Otto Rienhoff, Michael Resch, Elmar Mittler, Wilhelm Hasselbring, and their colleagues.

Thank you, guys !

10 References

[1] Tim O'Reilly, What Is Web 2.0, Design Patterns and Business Models for the Next Generation of Software, www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html
[2] OGSA Open Grid Services Architecture, www.globus.org/ogsa/
[3] I. Foster and C. Kesselman, "The GRID: Blueprint for a new Computing Infrastructure," Morgan Kauffman Publishers, 1st Edition 1999, 2nd Edition 2003.
[4] WS-RF Web Services Resource Framework, www.globus.org/wsrf/
[5] P.Dini, W.Gentzsch, M.Potts, A.Clemm, M.Yousif, A.Polze, "Internet, Grid, Self-Adaptability and Beyond: Are We Ready?", Proc. 2nd Intl. Workshop on Self-Adaptive & Autonomic Computing Systems, Zaragoza, Spain, Aug. 30 – Sept 03, 2004, www.dcl.hpi.uni-potsdam.de/papers/papers/134_SAACS_Panel_II_v3.0.pdf
[6] TeraGrid, www.teragrid.org
[7] NAREGI Japanese national grid project, www.naregi.org/index_e.html
[8] APAC Australian Partnership for Advanced Computing, www.apac.edu.au
[9] Website of CEC funded European grid projects, www.cordis.lu/ist/results
[10] EGEE, Enabling Grids for e-Science, http:// eu-egee.org/
[11] DEISA, Distributed European Infrastructure for Supercomputing Applications, www.deisa.org/index.php
[12] Large Hadron Collider Computing Grid Project, LCG, http://lcg.web.cern.ch/LCG/
[13] UK e-Science Programme, www.rcuk.ac.uk/escience/
[14] W.Gentzsch, Grid Computing in Research and Business, International Supercomputing Conference, Heidelberg 2005, www.isc2005.org/download/cp/gentzsch.pdf
[15] LHC Large Hadron Collider, http://lhc-new-homepage.web.cern.ch
[16] W. Gentzsch, Enterprise Resource Management: Applications in Research and Industry, In: Ian Foster and Carl Kesselman, Grid II: Blueprint for a new computing infrastructure, Morgan Kaufmann Publisher 2003.
[17] Amazon Elastic Computing Cloud, www.amazon.com/gp/browse.html?node=201590011
[18] Amazon Simple Storage Service, www.amazon.com/gp/browse.html?node=16427261
[19] IAETE School grids panel, www.iaete.org/soapbox/summary.cfm
[20] German D-Grid Initiative, https://www.d-grid.de/index.php?id=1&L=1
[21] Hai Jin, "ChinaGrid: Making Grid Computing a Reality", Digital Libraries: International Collaboration and Cross-Fertilization - Lecture Notes in Computer Science, Vol.3334, Springer-Verlag, December 2004, pp.13-24
[22] Open Grid Forum standardization organization, www.ogf.org

[23] D-Grid DGI Project, https://www.d-grid.de/index.php?id=61&L=1
[24] Authentifizierung im D-Grid (in German), www.d-grid.de/fileadmin/dgrid_document/Dokumente/vorschlagspapier-authz_v2.pdf
[25] Thesenpapier zum VO Management in D-Grid (in German), www.d-grid.de/fileadmin/dgrid_document/Dokumente/VOMS-Thesenpapier.pdf
[26] HEP-Grid, www.d-grid.de/index.php?id=44
[27] Astro-Grid, www.d-grid.de/index.php?id=45 and http://www.gac-grid.org/
[28] Medi-Grid, www.d-grid.de/index.php?id=42 and http://www.medigrid.de/
[29] C3-Grid, www.d-grid.de/index.php?id=46 and http://www.c3-grid.de/
[30] InGrid, www.d-grid.de/index.php?id=43 and http://www.ingrid-info.de/index.php
[31] Text-Grid, www.d-grid.de/index.php?id=167 and http://www.textgrid.de/
[32] WISENT Energy Meteorology, http://www.offis.de/projekte/projekt_e.php?id=181&bereich=bi
[33] OASIS, Organization for the Advancement of Structured Information Standards, http://www.oasis-open.org/
[34] W3C, The World Wide Web Consortium, http://www.w3.org/
[35] ESFRI, European Strategy Forum on Research Infrastructures, http://cordis.europa.eu/esfri/
[36] e-IRG, e-Infrastructure Reflection Group, http://www.e-irg.org/
[37] 2nd Int. Workshop on Campus and Community Grids (Draft) Eds: Laura McGinnis, David Wallom, Wolfgang Gentzsch, http://forge.gridforum.org/sf/go/doc14617?nav=1

Appendix:

2nd International Workshop on Campus and Community Grids (Draft)

Eds: Laura McGinnis, David Wallom, Wolfgang Gentzsch

http://forge.gridforum.org/sf/go/doc14617?nav=1

Status of This Memo

This memo provides information to the Grid community. It does not define any standards or technical recommendations. Distribution is unlimited.

Copyright Notice

Abstract

This is the proceedings from the 2nd International Workshop on Campus and Community Grids, held in Manchester, UK on 7th May 2007, as part of OGF20. This document includes the presentations that were accepted by the program committee.

1. Foreword

This workshop continues the series of Campus Grid workshops that was started at GGF-16 (Boston) where the workshop was hosted in collaboration with Harvard University. As well as providing a forum for local campus based grids this workshop also had the added benefit of being able to call on the experiences of the community and national/international grid attendees to ensure that issues arising and applicable lessons learned are not missed. This provided a balanced view of connectivity and expansion issues between local, regional, national and international grid endeavours.

With the success of the GIN activity at getting interoperability between national infrastructures it is important that we now expand and move this activity to show how already existent Campus Grid systems can be included in this activity and hence use this to connect them with the national grids. There is also an increasing realisation that a federation of smaller campus and community activities will eventually construct these themselves and here the lessons already learned by GIN can be most readily utilised.

The workshop used a mix of invited speakers from institutions that have already joined differing scales of infrastructures together and breakout sessions to identify key areas where further work is needed both within areas such as user lifecycle (including implementations of usage records and improvements to e-source usage specifications) and distributed systems management. There is also the large underlying area of data, storage, management and accounting that will be addressed.

2. Organizers

This workshop was jointly organized by the OGF Production Grid Services Research Group (PGS-RG) and German D-Grid project.

Prof. Dr. Wolfgang Gentzsch is Coordinator of the German D-Grid Initiative, Area Director for Major Grid Projects on the Steering Committee of the Open Grid Forum, visiting scientist at the Renaissance Computing Institute at UNC Chapel Hill, and an adjunct professor of computer science at Duke and NCState universities. Before, he was managing director for Grid Computing and Networking Services at MCNC and Sun Microsystems senior director for grid computing.

Dr. David Wallom is the Technical Manager of the Oxford e-Research Centre, Chair of the UK e-Science Engineering Task Force and co-chair of the Production Grid Services Research group within the OGF. Before arriving at Oxford he was Operations director for the Centre for e-Research Bristol at the University of Bristol.

Laura McGinnis is a Project Manager with the Systems and Operations Group at the Pittsburgh Supercomputing Center. She represents PSC to the National Science Foundation’s TeraGrid project as an active member of the Accounting and Education, Outreach and Training Working Groups. She also serves OGF as chair of the Usage Record Working Group and co-chair of the Production Grid Services Research Group.

3. Speakers

David Wallom, University of Oxford Campus Grid, OxGrid

The volume of computationally and data intensive research in a leading university can only increase. This though cannot be said of funding, so it is essential that every penny of useful work be extracted from existing systems. The University of Oxford has invested in creating a campus wide grid. This will be used to connect not only all large-scale computational resources within the university but also those shared use systems within teaching and student labs. This will also provide a uniform access method for ‘external’ resources such as the National Grid Service and the Oxford Supercomputing Centre.

Presentation located at:: http://www.ogf.org/OGF20/materials/715/OxGrid.pdf

Karim Djemame, White Rose Grid

The White Rose Grid (WRG) e-Science Centre brings together those researchers from the Yorkshire region that are engaged in e-Science activities and through these in the development of Grid technology. The initiative focuses on building, expanding and exploiting the emerging IT infrastructure, the Grid, which employs many components to create a collaborative environment for research computing in the region.

Presentation located at: http://www.ogf.org/OGF20/materials/715/ogf20-WhiteRoseGrid.pdf

Hugh Beedie and James Osborne, A Condor Grid @ Cardiff University

The introduction of Full Economic Costing in UK universities has meant that these organisations must capitalise on their entire computing infrastructure to support the computing needs of their researchers. It is also clear that cycle-stealing from existing IT infrastructure provides an extremely high return on investment (ROI). Information Services deployed the first version of Cardiff University's campus grid infrastructure back in April of 2004. Today the campus grid contains over 1400 Windows workstations providing a theoretical 800 GFLOPS of computing power to our researchers whilst at the same time increasing the ROI made by workstations originally purchased to support teaching and learning activities. With the current procurement of a cluster expected to provide a peak 20 TFLOPS of computing power, and the potential to expand the campus grid to provide an additional 10 TFLOPS of peak power, we expect to be able to support the research computing needs of our researchers both now and well into the future.

Presentation located at: http://www.ogf.org/OGF20/materials/715/OGF-Cardiff.pdf

Martin Dove, eMinerals MiniGrid

Many environmental problems, such as transport of pollutants, development of remediation strategies, weathering, and containment of high-level radioactive waste, require an understanding of fundamental mechanisms and processes at a molecular level. Computer simulations at a molecular level can give considerable progress in our understanding of these processes. The vision of the eMinerals project is to combine developments in atomistic simulation tools with emerging grid-computing technologies in order to stretch the potential for undertaking simulation studies under increasingly realistic conditions, and which can scan across a wide range of physical and chemical parameters. The project brings together simulation scientists, applications developers and computer scientists to develop UK escience/grid capabilities for molecular simulations of environmental issues.

Presentation located at: http://www.ogf.org/OGF20/materials/783/ogf20-eMinerals.pdf

Kashif Saleem, UKNEESGrid

Due to a growing requirement for state-of-the-art research facilities for conducting sophisticated and large-scale structural dynamic experiments, there is a paramount need to have a network of collaborative experiments and computational infrastructure. United Kingdom Network for Earthquake Engineering Simulation (UK-NEES) aims to collaborate to build such a network for the United Kingdom. It will act as another tool to benefit earthquake engineering technology in a quest to relieve the devastation caused by earthquakes. It will use grid technologies to enable a smooth interface with other similar network systems both in the UK and overseas, potentially allowing Oxford researchers to collaborate with other leading researchers across the world to further the understanding of seismic design through integrated experimentation, computation and simulation.

Presentation located at: http://www.ogf.org/OGF20/materials/783/OGF-UKNEESGrid.pdf

Andrew Richards, UK National Grid Service

The National Grid Service, funded by JISC, EPSRC and CCLRC, was created in October 2003 and the service entered full production in September 2004. The NGS is led and coordinated by the STFC in collaboration with the University of Manchester, the University of Oxford, the University of Edinburgh and the White Rose Grid at the University of Leeds. The UK's National Grid Service (NGS) provides a core e-Infrastructure that underpins UK research, providing standardized access to compute resources, data resources and large scale facilities, enabling collaborative computing across the UK. The NGS also provides a national "gateway" to international collaborations.

Presentation located at: http://www.ogf.org/OGF20/materials/783/ogf20-NGS.pdf

Wolfgang Gentzsch, D-Grid

Scientists in the D-Grid Infrastructure project are developing and implementing a set of basic grid middleware services which will be offered to the other Community Grids. Such services are, for example, access to distributed resources, applications, and large amounts of data in the grid, managing of virtual organizations, monitoring and accounting. In addition, a core-grid infrastructure is available to the community grids for testing and experimenting. High-level services will be developed which guarantee security, reliable data access and transfer, and fair usage of computing resources. This core-grid infrastructure will then be further developed into a reliable, generic, long-term production platform which can be enhanced in a scalable and seamless way, adding new resources, distributed applications and data, new services, and a support infrastructure, on the fly.

Presentation located at: http://www.ogf.org/OGF20/materials/783/OGF20-D-grid.pdf

Erwin Laure, EGEE

The EGEE project brings together experts from over 27 countries with the common aim of building on recent advances in Grid technology and developing a service Grid infrastructure which is available to scientists 24 hours-a-day.

The project aims to provide researchers in academia and industry with access to major computing resources, independent of their geographic location. The EGEE project will also focus on attracting a wide range of new users to the Grid. The project will primarily concentrate on three core areas:

- Build a consistent, robust and secure Grid network that will attract additional computing resources.
- Continuously improve and maintain the middleware in order to deliver a reliable service to users.
- Attract new users from industry as well as science and ensure they receive the high standard of training and support they need.

Presentation located at: http://www.ogf.org/OGF20/materials/783/ogf20-EGEE.pdf

John Brooke, NWGrid

The NW-GRID project, a collaboration between CCLRC Daresbury Laboratory and the Universities at Lancaster, Liverpool and Manchester, will establish a computational Grid comprising large-scale commodity computing systems coupled by a high-speed network. It will establish, for the region, a world-class activity in the deployment and exploitation of Grid middleware technologies (the software that glues together the various data and computing resources) and demonstrate the capabilities of the Grid in leading edge computational science and engineering applications.

Presentation located at: http://www.ogf.org/OGF20/materials/784/OGF20-NWGrid.pdf

Satoshi Matsuoka, NAREGI

The main objective of NAREGI is to research and develop grid middleware according to global standards to a level that can support practical operation, so that a large-scale computing environment (the Science Grid) can be implemented for widely-distributed, advanced research and education. NAREGI is carrying out R&D from two directions: through the grid middleware R&D at the National Institute of Informatics (NII), and through applied experimental study using nano-applications, at the Institute for Molecular Science (IMS). These two organizations advance the project in cooperation with industry, universities and public research facilities.

NAREGI, the National Research Grid Initiative, was created in 2003 by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). From 2006, under the “Science Grid NAREGI” Program of the “Development and Application of Advanced High-performance Supercomputer project “ being promoted by MEXT, research and development will continue to build on current results, while expanding in scope to include application environments for next-generation, peta-scale supercomputers.

Presentation located at: http://www.ogf.org/OGF20/materials/784/ogf20-NAREGI.pdf

Michael Gronager, Nordic Data Grid Facility

The Nordic Data Grid Facility, NDGF, is a collaboration between the Nordic countries (Denmark, Finland, Norway, Sweden). The motivation for NDGF is to ensure that researchers in the Nordic countries can create and participate in computational challenges of scope and size unreachable for the national research groups alone. NDGF is a production grid facility that leverages existing, national computational resources and grid infrastructures.

Currently, several Nordic resources are accessible with ARC and gLite grid-middleware, some sites with both. Today, the first operational user of the NDGF is the Nordic High Energy Physics community - the ALICE, ATLAS and CMS Virtual Organizations - through the operation of the Nordic Tier-1, which together with the Tier-0, CERN, and the other 10 Tier-1s collects, stores and processes the data produced by the Large Hadron Collider Experiment at CERN.

Presentation located at: http://www.ogf.org/OGF20/materials/784/OGF20-NDGF.pdf

John McGee, Open Science Grid

The Open Science Grid is a distributed computing infrastructure for large-scale scientific research, built and operated by a consortium of universities, national laboratories, scientific collaborations and software developers.

The OSG Consortium's unique community alliance brings petascale computing and storage resources into a uniform grid computing environment. Members of the OSG Consortium contribute effort and resources to the OSG infrastructure, and reap the benefits of a shared infrastructure that integrates computing and storage resources from more than 50 sites in the United States, Asia and South America.

Presentation located at: http://www.ogf.org/OGF20/materials/784/OGF20-OSG.pdf

Charlie Catlett, TeraGrid

TeraGrid is an open scientific discovery infrastructure combining leadership class resources at nine partner sites to create an integrated, persistent computational resource. Using high-performance network connections, the TeraGrid integrates high-performance computers, data resources and tools, and high-end experimental facilities around the country. These integrated resources include more than 102 teraflops of computing capability and more than 15 petabytes (quadrillions of bytes) of online and archival data storage with rapid access and retrieval over high-performance networks. Through the TeraGrid, researchers can access over 100 discipline-specific databases. With this combination of resources, the TeraGrid is the world's largest, most comprehensive distributed cyberinfrastructure for open scientific research.

Presentation located at: http://www.ogf.org/OGF20/materials/784/OGF20-TeraGrid.pdf

4. “Roundtable” Discussions

For each of the following topics a primer of the topic was displayed and the chair started discussion between each of the previous presenters.

Data in Grids: Authenticity and Integrity, Access Controls and Technology Evolution Management

Data on grids exacerbates existing issues with data provenance and conservation. In a traditional computing environment, if you have data currently in large volume, you know what it means. Data integrity, though important, is something that can be monitored. Within a project that has a longer lifetime, as well as when depositing research data into an institutional repository infrastructure, data integrity and provenance become serious issues, since the creators and primary users may not be interested in maintaining schemas etc. Since a lot of this data may be considered legacy with little or not metadata already, a lightweight way to authenticate and tag integrity context of data is needed.

Communities are collecting data and setting the standards for their organizations with respect to the integrity and identification of their data. Of the speakers within this workshop there are many different policies, for example:

- EGEE states when users arrive on the system that integrity is up to the user and community
- OSG plans inherently for data integrity, with periodic checking and if necessary reloading from a secondary store when there is corruption

It is clear though that there is a situation where a combination of engineering challenge versus identity and authentication mandates must be satisfied. Another issue that was then raised by the eMinerals project was whether this also included data provenance management? A suggested solution that could be considered is through commercial DRM.

From the production point of view there is also a threshold where retrieval of data from tape becomes insupportable, it is normally a product of equipment that is involved as well as pure data volume. An example was given by Reagan Moore of retrieval of 10TB within hours, though this required special hardware etc. The overall data loss threshold will vary from project to project but should be defined when the project starts and appropriate steps taken before the first data loss incident occurs. It should be remembered though that moving data around can trigger error correction that might not otherwise be needed, but it can actually help with the preservation of the data.

The other key problem is concerning metadata is that older collections are highly likely not to have it. Easy mechanisms must be devised for the addition of minimal metadata to legacy data sets being made available via grid technology.

Recommendations:

- We recommend the Data groups within OGF engage commercial DRM vendors in their discussions, either through attendance at one of their own meetings or off-site.
- There could be value in an information paper on data volumes as a reference which collects current practices in this area.
- Automated metadata generation should be another informational document, which would collate current practices in this area

User management, passing identities and accounting

As we move to federating local into regional and then national grids, the traceability of users, accounting for what they have done etc. becomes increasingly important. We should ensure that existing mechanisms are fully exploited before deciding that new mechanisms must be devised. It was noted that for most identity on the web Username/email address seems to be working, but has problems e.g. identity theft. It is an example, though, of a federated identity method based on local authentication.

The most pressing identity challenge is that for current grid methods, e.g. certificates, scalability challenges are still un-tested. It was also clear that the majority of groups that presented are at least looking at Shibboleth for identity management. Questions raised, though, include:

- Will Shibboleth deal with shortcomings identified in Kerberos?
- Will national federations also scale in their levels of trust to the same scale as other mechanisms?

Wit respect to accounting the question was raised about how to account for the heterogeneity of resources within a true grid environment. CPU weighing, for example, is based on the trustworthiness of the partners, who must report unbiased performance results for their resources to be fairly compared across the grid organization. It is also important to consider who wants/needs to look at accounting information and adjust solutions to their needs/requirements. There really can be no one size fits all solution.

Some sites are concerned about ”fair share”, especially if there are multiple consumers and providers, since there are real costs associated with usage (e.g. electricity, machine depreciation). Some sites need to be very precise with respect to usage so they can charge back to the resource user.

It was also noted that MPI across the WAN will change the charging model (CPU versus walltime). In general though schemes need to be agreed upon among the member organizations before service has started.

In any schema introduced by an organization there needs to be extensibility to include mechanisms for attaching measures of quality of service and charging for non-compute resources. An example of this could be taken currently from Amazon, who are charging for storage. In the end though how you charge is not too important for users as long as the model is clear. Users will normally decide where they want to compute and the charging method is part of their decision process. There are also technical requirements that must be taken into account including how charging schemas fit into the needs of metaschedulers and resource brokers.

Recommendation:

- OGF should continue to support activities that engage the support projects involved in user identity management, allocations, and usage tracking.

Support models, both systems and users

Within a production grid environment there are normally multiple levels and types of support, both formal and informal. In most instances though the balance between users and ‘resources’ needs for support is different. They can both be heavy, though, normally at around the same time, for example when a system change is made. All of the grids present try to deliver either a single point of contact for users, or to separate into user and site support (resource) calls.

Even though we do try to make grids easier to use, they are still very complicated and that users have to change how they do their work. These changes, ideally, should be as minimal as possible, which means the grids need to support users and site services to make the required changes on practice and process.

Since the majority of grids that were represented here make use of Open Software, the question was raised about who should be supporting it. It was resoundingly clear that this can only be done by the operating site, though this doesn’t scale in the long term. Generally grid software is complex; developers need to take this complexity into account as they further refine grid software.

Recommendations:

- Keep developers aware of the support costs of the software they are deploying, so that they can fix it; otherwise we'll keep paying in more complex user services.
- Identify current and best practices for user and site support. This will provide a reference for existing grid implementations, and for new groups that are considering deploying grid technology.

Licensing within cross organizational systems

About half of the speakers at this workshop have grid issues that include software licenses. The target populations include wholly academic users or increasingly collaborative projects between multiple organizations and industrial/academic collaborations. For wholly commercial organizations that use the grid there are problems where the internal methods of license control either don't scale or don't work. Non-optimal solutions discussed include

- Purchasing both commercial and non-commercial license, which could be cost prohibitive
- Mechanisms to isolate users from the inappropriate licenses adds extra management overhead to the system
- Utilizing traditional licensing models (e.g. per-CPU) that don't scale for grid computing

It is important that vendors are clear that we are not suggesting making their software free etc. But it is necessary for licenses to be both clear and sensible in their application so that both parties can benefit. We want to work with the vendors/suppliers to resolve these issues

Recommendation:

- OGF should support the creation of a Software Licensing Research Group. A BoF session is scheduled for later in this (OGF20) conference.

5. Author/Editor Information

Contact information for authors:

Wolfgang Gentzsch, D-Grid, Duke & RENCI An der Pirkacher Breite 11 93073 Neutraubling, Germany wgentzsch@d-grid.de	Laura F. McGinnis Pittsburgh Supercomputing Center 300 S.Craig Street, #313 Pittsburgh,PA 15213, USA lfm@psc.edu	David C .H. Wallom Oxford e-Research Centre University of Oxford 7 Keble Road Oxford, UK david.wallom@ierc.ox.ac.uk

6. Intellectual Property Statement

The OGF takes no position regarding the validity or scope of any intellectual property or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; neither does it represent that it has made any effort to identify any such rights. Copies of claims of rights made available for publication and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the OGF Secretariat.

The OGF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights which may cover technology that may be required to practice this recommendation. Please address the information to the OGF Executive Director.

7. Disclaimer

This document and the information contained herein is provided on an “As Is” basis and the OGF disclaims all warranties, express or implied, including but not limited to any warranty that the use of the information herein will not infringe any rights or any implied warranties of merchantability or fitness for a particular purpose.

8. Full Copyright Notice

Copyright (C) Open Grid Forum (2007). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the OGF or other organizations, except as needed for the purpose of developing Grid Recommendations in which case the procedures for copyrights defined in the OGF Document process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the OGF or its successors or assignees.

9. Presenter contact details

Hugh Beedie, Cardiff University, beedie@Cardiff.ac.uk
John Brooke, University of Manchester, john.brooke@manchester.ac.uk
Catlett Charlie, Argonne National Laboratory, catlett@mcs.anl.gov
Karim Djemame, University of Leeds, karim@comp.leeds.ac.uk
Martin Dove, University of Cambridge, martin@esc.cam.ac.uk
Wolfgang Gentzsch, RENCI, wgentzsch@mcnc.org
Michael Gronager, Nordic Data Grid Facility, gronager@ndgf.org
Erwin Laure, CERN, Erwin.Laure@cern.ch
Satoshi Matsuoka, Tokyo Institute of Technology, matsu@is.titech.ac.jp
James Osborne, Cardiff University, osborneja1@Cardiff.ac.uk
Andrew Richards, Science and Technology Facilities Council, A.J.Richards@RL.AC.UK
Kashif Saleem, University of Oxford, kashif.saleem@eng.ox.ac.uk
David Wallom, University of Oxford, david.wallom@oerc.ox.ac.uk