

Roadmap for a European Healthgrid

Vincent BRETON^{a,1}, Ignacio BLANQUER^b, Vicente HERNANDEZ^b, Nicolas JACQ^c,
Yannick LEGRE^a, Mark OLIVE^d et Tony SOLOMONIDES^d

^a *Corpuscular Physics Laboratory of Clermont-Ferrand, CNRS-IN2P3, France*

^b *Universidad Politecnica de Valencia, Spain*

^c *HealthGrid, France*

^d *University of the West of England, United Kingdom*

Abstract. This paper proposes a 10-year roadmap to achieve the goal to offer to healthcare professionals an environment created through the sharing of resources, in which heterogeneous and dispersed health data as well as applications can be accessed by all users as a tailored information providing system according to their authorisation and without loss of information. The paper identifies milestones and presents short term objectives on the road to this healthgrid.

Keywords. Grids, medical research, healthcare, medical informatics, medical imaging

Introduction

The concept of grids for health was born in Europe in 2002 and has been carried forward through the HealthGrid initiative [1]. This European initiative has edited, in collaboration with CISCO, a white paper setting out for senior decision makers the concept, benefits and opportunities offered by applying newly emerging Grid technologies in a number of different applications in healthcare [2]. Starting from the conclusions of the White Paper, the EU funded SHARE project aims [3] at identifying the important milestones to achieve the wide deployment and adoption of healthgrids in Europe. The project will devise a strategy to address the issues identified in the action plan for a European e-Health [4] and set up a roadmap for technological developments needed for successful take up of healthgrids in the next 10 years.

In a previous paper [5], we presented an analysis of the adoption of grids for biomedical sciences and healthcare in Europe, identifying bottlenecks and proposing actions. These actions have been further assessed within the framework of the SHARE European. The present paper proposes a technical roadmap for the adoption of healthgrids for medical research in Europe.

1. The Goal

Our goal is to offer to healthcare professionals an environment created through the sharing of resources, in which heterogeneous and dispersed health data as well as applications can be accessed by all users as a tailored information providing system

¹ Correspondence : breton@clermont.in2p3.fr

according to their authorisation and without loss of information. The environment should allow multiple usages depending on the user community and offer to any groups involved in medical and life sciences research means to share information and access resources on demand. Persistent infrastructures are needed for the users who are looking for fully functional services for medical research and aim at scientific production. For the users who are designing distributed applications and projects, toolkits are also needed to enable use of the Grid in a secure, interoperable and flexible manner. These toolkits should allow creating virtual organizations, administering it, manipulating biological and medical data, and interfacing with the infrastructure resources and services whenever needed. The environment should be open and evolving: for those who wish to integrate resources, tools and content, standard specifications should be documented.

The infrastructure itself should give access through standard interfaces to public and private, free and non free computing and storage resources; data stored in public and private databases. The data accessible on the environment include molecular data (ex. genomics, proteomics), cellular data (ex. pathways), tissue data (ex. cancer types), personal data (ex. EHR), population (ex. epidemiology); services for medical research such as analysis tools (workflow, data mining); services for on line patient care (telemedicine, visualization), For the majority of the healthcare professionals, the technical aspects should be completely hidden behind a friendly user interface. As stated above, our goal is to offer services to the healthcare professionals but legal issues may very significantly delay the deployment of healthgrids for healthcare. As a consequence, we are not sure how relevant a purely technical roadmap like the one presented in this paper is for healthcare. As a consequence, we expect the technology to get adopted primarily for medical research. Our roadmap will therefore refer specifically to medical research.

2. The starting point

The state of the art of healthgrids has been described in several documents [2,5,6,7]. We summarize here the main conclusions reached by their authors. Large scale grid infrastructures are available for scientific production in Europe; these infrastructures offer unprecedented opportunities for intensive computing. Several toolkits are now available which start to offer grid services in a secure, interoperable and flexible manner; these toolkits have still to be tested at a large scale on biomedical applications. Europe has witnessed in the last few years the emergence of e-science environments such as ^{my}Grid [8] or VL-e [9]; these environments are science portals for distributed analysis which offer scientists the possibility to carry out their experiments in a familiar environment while using the most recent web service technology and developments. Successful deployment of CPU intensive biomedical applications has been convincingly achieved on grid infrastructures world wide and several projects in Europe and world wide are successfully deploying biomedical grid applications; most of the success stories involve groups which have been active in grid projects for more than 5 years, which shows the present difficulty to use these environments for scientific production. The successful deployments in Europe are almost strictly limited to CPU intensive applications while very few applications involving manipulation of distributed medical data have been demonstrated so far; those were done at a prototype level by recognized team of computer experts in relation to clinicians.

The documents also highlighted several issues which had to be addressed in order to enable the healthgrid vision.

First issues are grid technology: web services are identified as the most promising technology to enable the HealthGrid vision despite their present limitations. There are today many different grid middleware (Globus, gLite, SRB, Unicore, GRIA, etc.) but none of them fulfil the requirements for a healthgrid. The ones which have demonstrated their scalability (gLite, Unicore) have limited functionalities particularly in the area of data management. Some which offer powerful and demonstrated data management functionalities (SRB) do not provide job management services. Moreover, these middleware are not so far built on web services and therefore do not offer standard interfaces. More recent grid middleware based on web services have not yet demonstrated their robustness and scalability.

Second issues are grid deployment: the deployment of grid nodes in healthcare centres such as hospitals or medical research laboratories is still extremely limited. This is due mainly to the present limitations of the existing European grid middleware which do not offer the necessary functionalities for secure manipulation of medical data. Other present limitations include the absence of an easy to install middleware distribution and the lack of friendly user interfaces to the grid for non experts. Just like Windows allows users to use their PC without any knowledge of the operating system, there is a need for user friendly interfaces to the grid.

Third issues are standardization: the definition and adoption of international standards and interoperability mechanisms is required for storing medical information on the grid. This includes for instance recording and ensuring consent, anonymization and pseudonymization.

Fourth issues are communication: there is an evident lack of information on the grid technology in the biomedical community. Communication on grids has been mostly focussed on the particle physics and computer sciences academic communities. Raising the awareness of the biomedical community is a major challenge for the coming years.

3. The road to grid infrastructures for medical research

In the previous section, we have listed a number of issues which need to be addressed in order to build the environment we have set up as our goal. In the following, we focus this study on medical research where adoption of grids is less depending on the evolution of EC legal texts. On the road to healthgrids, we have identified milestones which correspond to important steps forward in the services offered to the medical research community. We need to stress at this point that we are strongly convinced that a bottom-up approach has to be adopted on the road to healthgrids. Starting from the services made available on the existing grid infrastructures, a persistent distributed environment for medical research has to be built. This environment will progressively be enriched with new functionalities as technology progresses. In the United States, the Biomedical Informatics Research Network [1] is a very successful example of this bottom-up approach. In parallel to the building of this persistent environment, more volatile projects using grid toolkits or e-science environments to manage distributed data and knowledge for specific medical applications are very important to develop high level data integration services and spread the grid culture in the medical community. These services will later be made available on the grid infrastructures through standardized interfaces.

3.1. Milestones

Grid deployment is still very limited in European healthcare centres. Several factors contribute to this situation. One is the present human cost for deploying grid elements. Another important factor is the present situation in hospitals where none of the resources is accessible from the outside world. Our first deployment milestone called **MD1** is the “**successful permanent deployment of computing grid nodes inside European medical research centres**”. The goal of this first deployment is to create an environment where groups active in medical research can find resources for large scale simulation and modelling. This requires the development of friendly user interfaces. As a consequence of the present organization, a grid node in a hospital would have to be located outside the hospital firewall. Only anonymized or pseudonymized data could be transferred from inside the hospital firewall to the grid node. Technical solutions exist but they have not been deployed yet. These issues must be addressed and successful deployment of grid nodes inside European healthcare centres must be achieved. The services made available on these nodes are dependent on the technology. As documented in [6], the presently available grid infrastructures are mostly offering services for large scale computing.

As documented in [6], some data management services are emerging which are relevant to manipulate medical images. These services have to reach their full maturity for the second milestone called **MD2**, “**Successful permanent deployment of data grid nodes inside European medical research centres**”. The challenge here is to store medical data on the grid and this will only be achieved once the grid middleware will allow doing it in a secure way.

Once medical data are securely stored on the grid, the next issue is to deploy the services to query these data, to build relationships between and to provide appropriate representation to the researchers and to healthcare professionals. Building relationships between these data requires agreeing on standards for representing and storing them, in order to develop knowledge management services to manipulate the data. Once the standards are defined and widely adopted, the next grid deployment milestone **MD3** can be achieved, namely a “**Successful permanent deployment of knowledge grid nodes inside European medical research centres**”. The difference here between the data grid and the knowledge grid is that data stored on the data grid are just exploited through simple queries while a knowledge grid offer services to manipulate concepts while ignoring the underlying data model and the grid storage architecture.

3.2. Reaching the milestones

For each of the milestones described above, we need the right technology. We are going to discuss now the technological issues as well as the standardization issues which have to be addressed on the road to healthgrids.

3.2.1. Grid technology issues

We need a grid operating system that allows handling distributed computing for milestone MD1, distributed data storage and query for milestone MD2, data integration and knowledge management for MD3. There are definitely progresses being made in this direction and several software stacks are now available which provide relevant services based on OGSA (GT4 [10], OGSA-DAI [11], GRIA [12], etc). Beside the fact

that these toolkits are not yet offering all the functionalities required by a grid environment for healthcare, a major issue is that most of them have not yet been tested on medical applications and/or at a significant scale. On the other hand, several academic consortia are developing grid middleware which are deployed on the large scale infrastructures in Europe. These consortia are aware of the requirements of the medical community but these middleware are not yet based on web services because their development effort started before the migration of grid standards to web services. These two trends need to converge to deliver a **“tested, robust and scalable grid middleware based on web services that allow job and data management and which complies with EC countries laws on manipulation of personal medical data”**. Such a middleware is not for tomorrow and we rather foresee a progressive evolution of the grid operating systems similar to the one observed for personal computers.

Once a grid operating system is developed, a crucial issue is to make it easy to install in a healthcare centre. The availability of a public distribution of the middleware allowing a quick installation and configuration of the grid elements is required. This distribution requires significant resources because a middleware is very complex software with multiple modules and dependencies. Moreover, the technology is still evolving quickly as its standards are still under definition and only few implementations of these standards are now available. As a consequence, the distribution will have to be regularly updated to adapt to this rapid evolution path and be kept backward compatible so that sites using old versions are still able to belong to the same virtual organization as sites configured with newer versions of the distribution. In summary, a major challenge for the development of healthgrids is the **“availability of a free, easy to install and configure, robust and documented distribution of the grid middleware, accompanied by a significant user support”**.

3.2.2. Standardization issues

As already mentioned above, standards are absolutely necessary to the deployment of services which integrate data in bioinformatics and medical informatics, data coming from different medical disciplines and data coming from different countries in Europe. The adoption of standards for the exchange of biological and medical information is still limited to a few specific fields. These standards are needed to build data models, to produce ontologies and to develop knowledge management services. Moreover, they need to be compatible with grid standards so as to allow their implementation on the healthgrids. The largest initiatives in the medical informatics field such as DICOM and HL7 are just starting to study the interface between their standards and web services technology. We suggest focussing the effort on two topics, medical imaging and Electronic Health Records, in a first stage. In view of the importance of these standardization efforts, we introduce standardization milestones **MS1** corresponding to the **“Production of a standard for the exchange of medical images on the grid based on DICOM”**, and **MS2** corresponding to the **“Production of a standard for the exchange of Electronic Healthcare Records on the grid”** compatible with HL7.

3.2.3. Communication issues

The lack of information on grids is frequently identified as one of the key reasons why they have raised so far very little interest in the field of medical research. On the other hand, the services offered by the grids have been too limited to really make them a serious alternative to the existing computing models. A convincing communication

relies on success stories demonstrating the impact of grids for medical research. So we identify the need for **a demonstration environment offering a very easy access to the grid for non experts and providing some convincing services for medical research**. The evolution of this demonstration testbed should follow the evolution of the healthgrid through the different deployment milestones. On this dissemination environment, dedicated efforts to promote the technology can be developed.

3.2.4. Security issues

In this document, we are not dissociating the security issues from the other technological issues because they have to be addressed from the lowest middleware layers. Deployment of a data grid for medical research discussed above will only be possible when the middleware will be able to provide all the necessary guarantees in terms of management of personal data. Here is what we perceived as the specific technical challenges related to the handling of medical data on the grid: manipulation of personal data on the grid must strictly obey regulations; these regulations change from country to country in Europe. Services for anonymization and pseudonymization of medical data must be provided. Medical data belong to the patient; a mechanism must be set up to allow any given individual to access to all his/her data on the grid. In the perspective of the usage of grids for healthcare, authentication of health professionals on the grid can not be handled by requesting all of them to get a grid certificate; a mechanism must be set up so that professional cards can be used to provide authentication on the grid.

3.2.5. Summary

We have identified 5 milestones on the road: MD1, called “Computing grid”, corresponding to the successful permanent deployment of computing grid nodes inside European medical research centres; MD2, called “Data grid”, corresponding to the successful permanent deployment of data grid nodes inside European medical research centres; MD3, called “Research K-Grid”, corresponding to the successful permanent deployment of knowledge grid nodes inside European medical research centres; MS1, called “Grid DICOM”, corresponding to the production of a standard for the exchange of medical images on the grid; MS2, called “Grid EHR”, corresponding to the production of a standard for the exchange of Electronic Healthcare Records on the grid.

The milestones MD1, MD2, MD3 correspond to the deployment of infrastructures while MS1 and MS2 are related to the availability of standards. The figure 1 illustrates how the different milestones follow each other on the road and how the progress on the road will depend on the availability of the grid operating system as well as its distribution. For each of the three deployment milestones, computing grid, data grid and knowledge grid, the figure illustrates how two environments are needed, one for scientific production and one for demonstration.

We estimate to about 10 years the time necessary to reach the goal of deploying an healthgrid after successfully achieving the different milestones discussed previously. The technology available today allows reaching the first milestone MD1 today. We estimate about 2 years are needed to achieve MD2 while at least 5 years will be needed for MS1, MS2 and MD3. Finally, our experience shows that about three years will be needed from the day a first knowledge grid is deployed to the day it is robust enough to become an environment for scientific production.

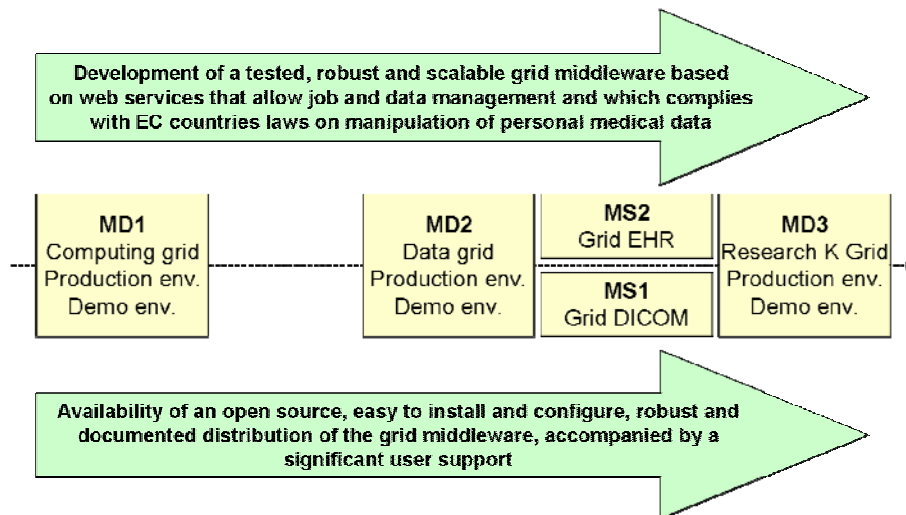


Figure 1. Illustration of the key challenges and the milestones on the road to healthgrids

3.3. Risk analysis

There are three keys to the successful progress on the technical road to healthgrids: the evolution of the technology, including development and distribution of the needed grid middleware; the deployment of stable infrastructures with guaranteed level of services; the deployment of medical applications on these infrastructures. As a consequence, we identify three key technical risks to achieve the vision: the absence of concentration of the critical mass of expertise to develop the grid middleware and its distribution; the absence of agreed standards to share medical images and Electronic Health Records on the grid; the non-adoption of the healthgrid infrastructures by the research community so that they are not used by medical applications. To these technical risks, we add a fourth major risk which is the absence of evolution of the legal texts to allow sharing of medical data: this could completely prevent the deployment of an healthgrid.

4. Short term roadmap (2-3 years)

Successful achievement of the 10-year target depends on immediate actions. We recommend developing R&D activities along three lines:

1. Develop healthgrid infrastructures
2. Deploy biomedical grid applications on the existing infrastructures;
3. Deploy biomedical grid applications using OGSA compliant grid toolkits and e-science environments.

Two out of the three lines (2 and 3) are now actively pursued around the world as biomedical applications are deployed on almost all grid infrastructures and many projects are now under development using web services toolkits or e-science environments. The three lines are needed because there is no dedicated grid environment for medical research and there are no OGSA compliant grid toolkits and e-science environments available on the existing infrastructures. However a convergence of these research axes should be achieved in about 2 to 3 years when the middleware deployed on the healthgrid and on the grid infrastructures will offer web

service interfaces to their grid services so that the grid toolkits and e-science environments will be available to all healthgrid users. In parallel to these R&D activities, we recommend to pursue actively the definition of standards for the sharing of medical images and electronic health records on the grid. Rather than developing new standards, grid experts should as much as possible get involved in the already existing medical informatics standardization bodies.

4.1. Development of healthgrid infrastructures

Having a dedicated infrastructure for medical research is a key to the adoption of grids for medical research. It also solves the problem of handling priorities on a multidisciplinary grid where the services made available have been chosen through a consensus process and specific requirements for the medical community like security issues have low priority for other research communities. This dedicated infrastructure should deploy the most up to date services relevant to biomedical research as soon as they have demonstrated their scalability and robustness. The infrastructure will initially mostly focus on computing services, but should aim at deploying very rapidly secure data management services and later knowledge management tools. In order to offer immediate service to the community, the healthgrid should be built using technologies which are interoperable with the existing infrastructures while keeping the perspective to offer as soon as possible web services interfaces to the grid services.

4.2. Deployment of biomedical grid applications on the existing infrastructures

One can wonder why it is interesting to keep deploying biomedical applications on the existing multidisciplinary infrastructures once an healthgrid is running. The first reason is that one should not reinvent the wheel and take advantage of the services already offered by infrastructures like EGEE [13] and DEISA [14]. The second reason is that the amount of resources and storage available on the healthgrid will hardly ever compare to what EGEE and DEISA provide. The healthgrid should provide the interface and the high level services needed for medical research but whenever there is a need for heavy CPU or storage resources, mechanisms should be available to use EGEE or DEISA to address these needs. As a consequence, through the collaboration between the developers of both healthgrid and infrastructure projects, tools should be designed to achieve this submission in the most transparent way. This means also that deployment of biomedical grid applications will increase on infrastructures like DEISA and EGEE once the healthgrid is deployed. These infrastructures should offer the largest possible palette of services relevant to medical research and exposed through web services interfaces. As a consequence, a continuous effort must be maintained to develop the synergy between the projects and to improve the environment for the deployment of biomedical grid applications on these infrastructures.

4.3. Deployment of biomedical grid applications using OGSA compliant grid toolkits and e-science environments

The services offered for medical research evolve with the grid operating system technology and we expect therefore to witness the evolution of the applications deployed with the progress of the middleware. On the other hand, most of the present European biomedical grid projects require already a high level of data integration

which is beyond the capacities of the existing grid infrastructures. Nowadays, we observe a gap between the needs of these projects and the services made available on the infrastructures. How to address this gap is an important issue. The roadmap exposed above presents the bottom-up approach where infrastructures will progressively offer new services with the evolution of the technology. Toolkits such as OGSA-DAI or GRIA are very relevant to address the specific needs of medical grid projects and their use should be strongly encouraged. As well, high level environments such as ^{my}Grid or VL-e allow now biologists and healthcare professionals to start manipulating concepts they are familiar with while accessing potentially distributed data. The next step is to achieve the deployment of these environments on grid infrastructures and evolve them towards improved usability by end users.

4.4. Development of standards

The R&D activities described above are going to require the manipulation of medical data on the grid. As a consequence, standards for the exchange of medical data will be more and more required. At all costs, grid experts should avoid developing new standards but rather should as much as possible get involved in the already existing medical informatics standardization bodies. There, they can disseminate the concept of grids and work on extending the existing standards to distributed environments. The Open Grid Forum is a potentially very interesting place where grid experts involved in the different medical informatics standardization bodies could meet and investigate the interface between these standards and the existing grid standards. However, this would require a deep reorganization of the present Life Sciences Research group at OGF which is not working properly. The Healthgrid initiative provides the right framework to coordinate the development of the different standards in collaboration with the OGF and the different medical informatics standardization bodies.

5. Conclusion

In this document, we are proposing a 10-year roadmap to achieve the goal to offer to healthcare professionals an environment created through the sharing of resources, in which heterogeneous and dispersed health data as well as applications can be accessed by all users as a tailored information providing system according to their authorisation and without loss of information. Starting from the state of the art on healthgrids described in several documents [2,5,6,7], we have described a way toward the environment. We have identified 5 milestones on the road:

- MD1, called “Computing grid”, corresponding to the successful permanent deployment of computing grid nodes inside European medical research centres,
- MD2, called “Data grid”, corresponding to the successful permanent deployment of data grid nodes inside European medical research centres,
- MD3, called “Research K-Grid”, corresponding to the successful permanent deployment of knowledge grid nodes inside European medical research centres,
- MS1, called “Grid DICOM”, corresponding to the production of a standard for the exchange of medical images on the grid,
- MS2, called “Grid EHR”, corresponding to the production of a standard for the exchange of Electronic Healthcare Records on the grid.

Achieving these different milestones require the availability of a grid operating system providing all the needed functionalities as well as an easy-to-install distribution of this middleware.

We have identified major technical risks which can prevent the vision to happen:

- the absence of concentration of the critical mass of expertise to develop the grid middleware and its distribution,
- the absence of agreed standards to share medical images and Electronic Health Records on the grid,
- the non-adoption of the healthgrid infrastructures by the research community so that they are not used by medical applications,

To these technical risks, we have added a fourth major risk which is the absence of evolution of the legal texts to allow sharing of medical data which can prevent the deployment of an healthgrid.

Finally, we have proposed in the next 2 to 3 years to develop R&D activities along three lines:

- Develop healthgrid infrastructures,
- Deploy biomedical grid applications on the existing infrastructures,
- Deploy biomedical grid applications using OGSA compliant grid toolkits and e-science environments.

In parallel to these R&D activities, we recommend to pursue actively the definition of standards for the sharing of medical images and electronic health records on the grid in the already existing medical informatics standardization bodies. We consider that the Healthgrid initiative provides the right framework to coordinate the development of the different standards in collaboration with the OGF and the different medical informatics standardization bodies.

6. Acknowledgements

The SHARE project is co-funded by the European Commission under contract n°FP6-2005-IST-027694.

7. References

- [1] Information on Healthgrid initiative available at <http://www.healthgrid.org>.
- [2] V. Breton, et al., editors on behalf of the Healthgrid White Paper collaboration, The Healthgrid White Paper, Proceedings of Healthgrid conference, IOS Press, Vol 112, 2005.
- [3] Information on Share project available at <http://www.eu-share.org>.
- [4] Action plan for a European e-Health Area, COM(2004) 356, European Commission, http://europa.eu.int/information_society/doc/qualif/health/COM_2004_0356_F_EN_ACTE.pdf.
- [5] V. Breton, et al., Proposing a roadmap for Healthgrids, Stud Health Technol Inform. 120:319-29., 2006.
- [6] Healthgrid technology baseline report, Share deliverable D3.2 available from <http://www.eu-share.org>.
- [7] V. Breton, et al., HealthGrid, a new approach to eHealth, proceedings of the eHealth 2006 conference, Malaga, 2006.
- [8] R.D. Stevens et al., MyGrid, personalised bioinformatics on the information grid, Bioinformatics 19(1) i302-i304, 2003.
- [9] H. Rauwerda, et al, The Promise of a virtual lab, Drug Discov. Today. 11(5-6):228-36, 2006.
- [10] I. Foster, Globus Toolkit Version 4: Software for Service-Oriented Systems, IFIP International Conference on Network and Parallel Computing, Springer-Verlag LNCS 3779, 2-13, 2005.
- [11] Information on OGSA-DAI middleware available at <http://www.ogsadai.org.uk>.
- [12] Information on GRIA middleware available at <http://www.gria.org>.
- [13] F. Gagliardi, et al., Building an infrastructure for scientific Grid computing: status and goals of the EGEE project, Philosophical Transactions: Mathematical, Physical and Engineering Sciences, 363, 1729-1742, 2005 and <http://public.eu-egee.org>.
- [14] Information on DEISA available at <http://www.deisa.org>.