EU Funded Grid Development in Europe

Paul Graham[3], Matti Heikkurinen[1], Jarek Nabrzyski[2], Ariel Oleksiak[2], Mark Parsons[3], Heinz Stockinger[1], Kurt Stockinger[1], Maciej Stroiński[2], Jan Węglarz[2]

[1] CERN, European Organization for Nuclear Research, Switzerland
[2] PSNC, Poznan Supercomputing and Networking Center, Poland
[3] EPCC, Edinburgh Parallel Computing Centre, Scotland

Abstract. Several Grid projects have been established that deploy a "first generation Grid". In order to categorise existing projects in Europe, we have developed a taxonomy and applied it to 20 European Grid projects funded by the European Commission through the Framework 5 IST programme. We briefly describe the projects and thus provide an overview of current Grid activities in Europe. Next, we suggest future trends based on both the European Grid activities as well as progress of the world-wide Grid community. The work we present here is a source of information that aims to help to promote European Grid development.

1 Introduction

Since the term "Grid" was first introduced, the Grid community has expanded greatly in the last five years. Originally, only a few pioneering projects such as Globus, Condor, Legion and Unicore provided Grid solutions. Now, however, many countries have their own Grid projects that provide specific Grid middleware and infrastructure.

In this paper, in order to give a comprehensive overview of existing technologies and projects in Europe, we establish a general taxonomy for categorising Grid services, tools and projects. This taxonomy is then applied to existing projects in Europe. In particular, within the GRIDSTART [5] framework we have analysed 20 representative Grid projects funded by the European Commission in order to highlight current European trends in Grid computing. The guiding principle behind this taxonomy is to enable the identification of trends in European Grid development and to find out where the natural synergies between projects exist.

Since the need for this taxonomy was practical - and relatively urgent - a certain amount of guidance in the form of "pre-classification" was deemed necessary in the information gathering phase. This meant that rather than asking open questions about the activities of the projects and creating the classification based on the answers, the projects themselves were asked to identify which layers and areas (see later) they worked on according to a classification presented to them in a series of questionnaires. Thus, it is likely that this taxonomy will evolve as the contacts and collaboration between projects increases.

This taxonomy is based on the IST Grid Projects Inventory and Roadmap [4](a 215 page document). In this paper we extract the key aspects of the data presented in that document and refer to the original document for further details.

The paper should also prove of interest to the broader distributed computing community since the results presented provide a clear overview of how European Grid activities are evolving. The paper supersedes previous work reported in [7] (describing the initial work towards this survey) and [1] (reporting on a preliminary overview). The more up-to-date overview provided in this paper covers new trends and Grid services which are rapidly evolving from standardisation work as well as benefiting from insight into the latest developments in the various projects, that have occurred since the initial overviews were prepared.

2 Taxonomy

Development of Grid environments requires effort in a variety of disciplines, from preparing sufficient network infrastructure, through the design of reliable middleware, to providing applications and tailored to the end users.

The comparison of Grid projects is made according to three different categorisation schemes. The first is by different technological **layers** [2, 3] that separate the Grid user from the underlying hardware:

- **Applications and Portals.** Applications such as parameter simulations and grand-challenge problems often require considerable computational power, access to remote data sets, and may need to interact with scientific instruments. Grid portals offer web-enabled application services, i.e. users can submit and collect results for their jobs on remote resources through a web interface.
- **Application Environment and Tools.** These offer high-level services that allow programmers to develop applications and test their performance and reliability. Users can then make use of these applications in an efficient and convenient way.
- **Middleware (Generic and Application Specific Services).** This layer offers core services such as remote process management, co-allocation of resources, storage access, information (registry), security, data access and transfer, and Quality of Service (QoS) such as resource reservation and trading.
- **Fabric and Connectivity.** Connectivity defines core communication protocols required for Grid-specific network transactions. The fabric comprises the resources geographically distributed and accessible on the Internet.

The second categorisation scheme concerns technical **areas**, which include topics such as dissemination and testbeds and which address the wider issues the impact of Grid technology. All areas with their projects are listed in Figure 2 which categorises different the aspects of Grid projects.

The third main categorisation scheme in this article focuses on the scientific domain of **applications** as well as the computational approaches used (see

Section 3.3). Further related work on an earlier taxonomy of Grid resource management can be found in [6].

3 Major Trends in Grid Development

In the Grid inventory report we analysed the major Grid projects in Europe that are referred to as *Wave 1* ("older" projects that received funding prior to 2001) and *Wave 2* ("younger" projects). Links to all project web-sites can be found at [5].

- Wave 1 projects are formally part of the EU-funded GRIDSTART project and are as follows: AVO, CrossGrid, DAMIEN, DataGrid, DataTAG, EGSO, EUROGRID, GRIA, GridLab and GRIP.
- Wave 2 projects are informal partners of the GRIDSTART project and are as follows: BioGrid, COG, FlowGrid, GEMSS, GRACE, GRASP, MammoGrid, MOSES, OpenMolGRID, SeLeNe.

Apart from these EU-funded projects, there are several other national and multi-national Grid initiatives like INFN Grid (Italy), NorduGrid (Northern European countries) and the e-Science Programme (UK) that each encompasses a range of projects. Most of these projects have informal ties with one or more GRIDSTART projects, but the analysis of these ties is beyond the scope of this document.

The analysis presented in this document is based on a survey, categorised by Grid areas, that has been submitted to each of the projects. For further details on the analysis methodology we refer to [4].

3.1 Development in Grid Layers

Generally, one can observe the following *trend*: projects which started later are biased towards the development of higher-level tools and applications (this trend is continued by Wave 2 projects). This is justified since several projects (such as DataGrid and EuroGrid) are preparing a good basis for further work by developing low-level tools in the Fabric and Middleware layer. However, it is not a general rule. For instance, projects such as DataGrid, DataTAG and CrossGrid, which are co-operating with each other in order to prepare an environment for data-intensive applications, work on Fabric layer components although they started at different times. This complementary work is beneficial, since the application domains of the projects are different.

In the GRIDSTART cluster there are large projects with activities covering many Grid layers (DataGrid, GridLab, CrossGrid: these projects work on complimentary aspects in the specific layer) and smaller projects focused on particular layers (DataTAG, DAMIEN). All Wave 2 projects belong to this second group. Many of them focus on the highest layer and/or on a single application domain (e.g., COG, OpenMolGRID, SeLeNe). Wave 2 projects rarely work on the fabric layer.

The choice of the underlying Grid system obviously influences the architecture of projects too. The two principle Grid toolkits in this study, Globus and UNICORE, are used (see Figure 1). Globus is a more "horizontal" solution in the form of a toolkit offering much necessary functionality while UNICORE is more "vertical" and provides software up to the Graphical User Interface. The influence of these characteristics on project architecture can be noted, for example, in the case of EUROGRID and GRIP. These projects "Grid-enable" applications through preparing application-specific UNICORE plug-ins. They also add more dynamic functionality through extending the UNICORE system itself.

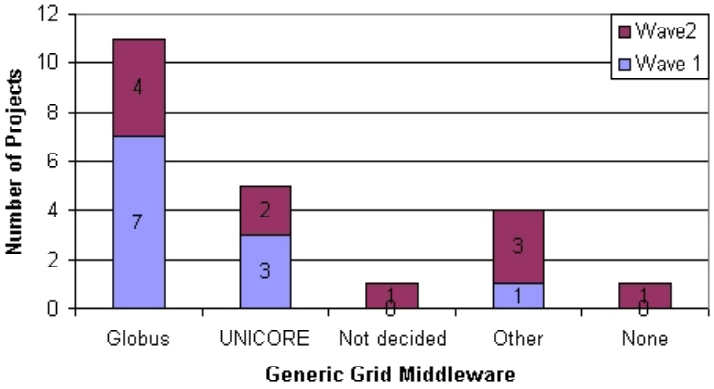

Fig. 1. Generic Grid middleware used by projects analysed in this paper divided into Wave1/Wave2 projects. "Not decided" concerns projects that were in an early stage of development and various solutions were tested. "None" concerns ontology Grid projects and therefore no need for submission of computation jobs has been identified.

Generally, one notice that differences between project architectures result from the *different types of Grids* that are being developed. Although the layers defined in Section 2 can still be distinguished, in Data and Information Grids, replication or data search services are placed above various data archives while, in the case of Computational Grids, the global scheduler and job submission systems are built on top of local resource management systems. The next major difference that occurs in the *architectures* of Grid projects results from the trend towards a service-oriented model. Some projects (GEMSS, GRASP, GRIA) represent the service-oriented approach. The difference here is that the stress is put on services (and their performance) rather than specific hardware resources, or a specific scientific application.

3.2 Development in Areas

Development in particular areas is given in Figure 2 which displays the number of projects putting significant effort into a given area. For example, 7 projects develop portals, 17 projects deal with applications, 11 out of the 20 projects focus on Resource Management (for more details see [4]).

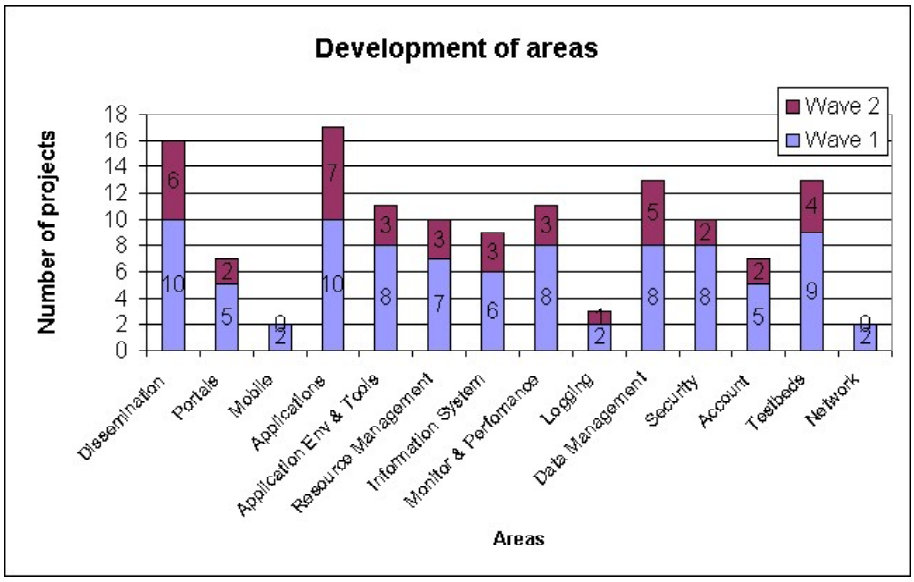

Fig. 2. Areas developed by analysed projects divided into Wave1/Wave2 projects

We also distinguish between Wave 1 and Wave 2 projects in order to indicate the directions and trends of both groups of projects. We can observe the following phenomena:

- Focus on some areas is significantly less for Wave 2 projects, e.g. Resource Management, Information Services, Application Development Environment and Tools, possibly due to the existence of solutions in previous projects.
- Although the scope of development in a certain area in Wave 2 projects may be similar to the one of Wave 1 projects, there is a different level of abstraction. For example, in the case of data management, Wave 2 projects may work on knowledge management rather than low level data access techniques.
- Although Wave 2 projects are more oriented towards high-level Grid functionality, there has been little concentrated, cross-project effort in the development of user friendly access methods such as portals or mobile access. Instead the emphasis is placed on techniques that add semantics to data (and in consequence facilitate access for end users).

Figure 2 does not communicate exactly to the extent of developments in given areas, since projects put different emphasis on specific areas. Furthermore, some technologies belonging to a specific area may be developed to a greater extent than in others.

In Table 1 we give a summary of different solutions provided by the projects. For many of the areas analysed, there are now existing tools that can be incorporated into other projects. Possible examples of existing solutions that may be (or may already have been) applied by other Grid initiatives in order to profit from synergies are:

- security: VOMS [4] (DataGrid/DataTAG)
- schema for information system : GLUE schema (DataTAG)
- data management: Spitfire (DataGrid)
- developer tools: PACX-MPI (DAMIEN) and MARMOT (CrossGrid)
- framework for portals: GridSphere (GridLab)

Additionally, there are ongoing developments that may provide the basis for further interesting initiatives in the near future. Examples of such solutions are: Resource Management - GRMS (GridLab); Security - GAS (GridLab), CoPS (AVO); Application Development Environments and Tools - UNICORE plugins (EUROGRID, GRIP), GAT [5] (GridLab); Accounting - Accounting and Billing services (EUROGRID, GRASP).

Despite all these solutions there are several problems yet to be overcome which include:

- Transfer of current solution components to a service based approach.
- Focus on learning from mistakes to build what will become the first reliable, resilient and robust "production" Grids.

3.3 Development in Applications

This section is devoted to applications, as they are the main stimulators of Grid infrastructure development. Their domains, requirements and user communities have a great influence on the structure of many of the current projects.

Figure 3 shows the numbers of applications from particular domains. We have distinguished the following general domains: Earth and Environmental Sciences, Biology and Medicine, Physics and Astronomy, Engineering and Multimedia. All remaining domains fall into the category *other domains*, which includes many commercial and business applications.

Although many Grid projects have their roots in physics or are driven by other scientific domains such as biology, medicine or earth sciences, there are also industrial applications including engineering and multimedia. The distribution of application domains has changed considerably between the Wave 1 and Wave 2 projects. Several projects apply Grid technology to the fields of biology and

[4] VOMS: Virtual Organization Membership Service

[5] GAT: Grid Application Toolkit

Dissemination	Industry and research forum Tutorials	CrossGrid, DataGrid DataGrid, GridLab, CrossGrid
Portals	Genius portal, Earth Observation Portal End user support of several existing portals Migrating desktop GridSphere framework, portlets and portlet development environment UNICORE portal	DataGrid, MammoGrid AVO CrossGrid GridLab EUROGRID, GRIP, OpenMolGRID
Applications	Bio-technology applications Cactus-based applications (general purpose framework, large user community in the astrophysics)	EUROGRID, GRIP, GEMSS GridLab
Application env. and tools	PACX-MPI Extended Code Coupling Interface MpCCI, MetaVampir, DIMEMAS UNICORE plugins MARMOT MPI verification tool GAT	DAMIEN, CrossGrid (used by) DAMIEN EuroGrid, OpenMolGRID CrossGrid GridLab
Resource Mgmt and Scheduling	Integration to batch systems enhanced Job description languages. Job submission system QoS Manager fro MPI applications Resource Mgmt System (UNICORE based) Use of Globus resources through UNICORE GridLab Resource Mgmt System (GRMS)	Several of the Wave 1 projects DataGrid DAMIEN EUROGRID GRIP GridLab
Information	GLUE UNICORE IDB information/knowledge management	DataTAG EUROGRID, GRIP BioGrid,COG,MOSES,SeLeNe
Monitoring and Performance	R-GMA, Fabric Monit. & Fault Tolerance Monitoring System MetaVampir, DIMEMAS G-PM	DataGrid GridLab DAMIEN CrossGrid
Logging	Logging and Bookkeeping service	DataGrid
Data Mgmt	Hierarchical Replica Management Spitfire search or categorisation	DataGrid DataGrid EGSO, GRACE
Security	VOMS Plug-in for GSI-proxy generation GridLab Authorisation Service (GAS)	DataGrid/DataTAG GRIP GridLab
Testbeds	(Intercontinental) testbed	DataGrid, DataTAG, GridLab, FlowGrid

Table 1. Solutions provided by analysed projects

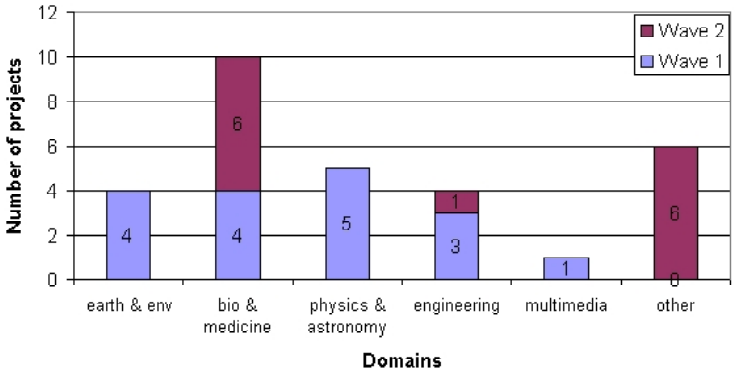

Fig. 3. Areas developed by analysed projects divided into Wave1/Wave2 projects

medicine. New applications have also appeared including the ERP sector, e-Learning and solutions such as the semantic web designed for multiple domains. A classification can be found in Figure 4.

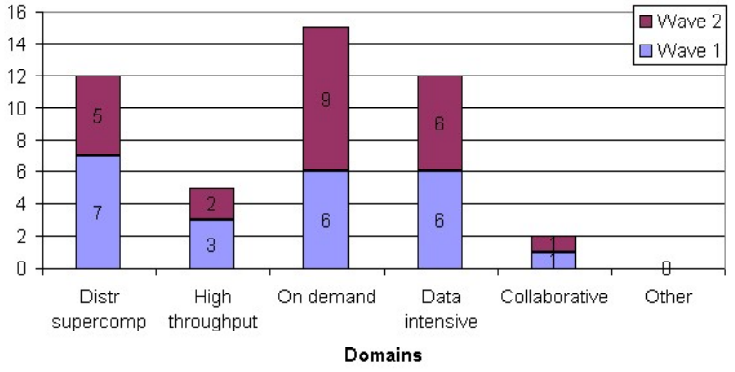

Fig. 4. Applications

The majority of applications for Wave 1 projects deal with large amounts of data (data intensive applications) or require huge computing power (distributed supercomputing applications). However, we should also notice the increasing need, especially in the case of Wave 2 projects, for on demand and collaborative applications, which have additional requirements for higher-level services including mechanisms for controlling quality of service, and sometimes even new

architectures (e.g. in the form of distributed services). Additionally, applications that need remote resources for a certain amount of time (on demand applications) often require efficient payment mechanisms. All these trends must be taken into consideration while developing Grid middleware and infrastructure.

Comparing the applications from Wave 2 with those from Wave 1, the following conclusions may be drawn:

- Although present in Wave 1 projects, there is a greater focus on industrial applications in Wave 2.
- Many of of the Wave 2 applications are in the medicine and bio-technology field.
- The trend that about half of the projects deal with data-intensive applications continues, but Wave 2 projects focus on semantics of data and knowledge extraction rather than on low-level data management.
- New applications are emerging for instance in the financial sector (GRASP), ERP (GRASP) and with regard to corporate ontologies targeted to various industries (COG).
- Most Wave 2 projects focus on a single specific area, however, there are also projects such as GRASP or COG targeted to wider communities of users.

There are also areas being developed by only a few projects that might need more consideration in the future:

- Accounting services serve as an example of such an area. Their development is one of the main goals of the GRIA project, which is developing business models for the Grid. ASP services that include accounting and billing are also being implemented in the scope of EUROGRID.
- Mobile access is another example of an activity specific to some of projects. This is one of the objectives of both GridLab and CrossGrid.
- Activities such as knowledge management and semantic Grids do not belong to the goals of "older" Wave 1 projects; however, there are projects concerning these areas in the scope of several Wave 2 projects such as COG, MOSES or BioGrid.

Real industrial applications are being used even in the early Grid projects, which is quite unusual for an emerging technology and demonstrates the validity of the IST funding model. Overall, there is a strong requirement for business involvement since it is increasing the speed of Grid development and is attracting broad communities of end users.

4 Conclusion and Future Trends

In this paper we presented a simple taxonomy of Grid projects based on an inventory of Grid projects in Europe funded in part by the European Union. The main trend is for more recent Wave 2 projects to focus on the high layers of technology, and on biomedical applications in particular. On the other hand, distributed supercomputing and its application has been deemphasised in Wave 2. Based on the current status, we foresee the following future trends:

– International and inter-project collaborations and interoperability will gain more importance. Strong working groups - and organisational support for their work on all the levels involved in the European Grid research - are required in order to profit from synergies and to deal with interoperability.
– There is a strong requirement for quality, reliability, security and above all interoperability for Grid systems. As a result, web services and in particular OGSA will most probably "dominate" the Grid "market" in the short to medium term: we see this tendency already in the newer projects of our survey.

Acknowledgements

This work was partially funded by the European Commission program IST-2001-34808 through the EU GRIDSTART Project. We thank:

F. Grey; M. Dolensky, P. Quinn; P. Nowakowski, B. Krammer; R. Badia, P. Lindner, M. Mueller; R. Barbera, F. Bonnassieux, J. van Eldik, S. Fisher, A. Frohner, D. Front, F. Gagliardi, A. Guarise, R. Harakaly, F. Harris, B. Jones, E. Laure, J. Linford, C. Loomis, M. B. Lopez, L. Momtahan, R. Mondardini, J. Montagnat, F. Pacini, M. Reale, T. Roeblitz, Z. Salvet, M. Sgaravatto, J. Templon; R. Cecchini, F. Donno, JP Martin-Flatin, O. Martin, C. Vistoli; R. Bentley, G. Piccinelli; HC Hoppe, D. Breuer, D. Fellows, KD Oertel, R. Ratering; M. Surridge; M. Adamski, M. Chmielewski, Z. Balaton, M. Cafaro, K. Kurowski, J. Novotny, T. Ostwald, T. Schuett, I. Taylor; P. Wieder; E. v.d. Horst; M. Christostalis, K. Votis; N. Baltas, N. F. Diaz; J. Fingberg, G. Lonsdale; M. Cecchi; S. Wesner, K. Giotopulos, T. Dimitrakos, B. Serhan; A. Ricchi; S. Sild, D. McCourt, J. Jing, W. Dubitzky, I. Bagyi and M. Karelson; A. Poulovassilis, P. Wood

References

1. Marian Bubak, Piotr Nowakowski and Robert Pajak. An Overview of EU-Funded Grid Projects. 1st European Across Grids Conference, Santiago de Compostella, Spain, February 2003.
2. Desplat J.C., Hardy J., Antonioletti M., Nabrzyski J., Stroinski M., Meyer N. Grid Service Requirements, ENACTS report, January 2002.
3. Ian Foster, Carl Kesselman, Steve Tuecke. The Anatomy of the Grid: Enabling Scalable Virtual Organizations, Intl. J. Supercomputer Applications, 15(3), 2001.
4. Fabrizio Gagliardi, Paul Graham, Matti Heikkurinen, Jarek Nabrzyski, Ariel Oleksiak, Mark Parsons, Heinz Stockinger, Kurt Stockinger, Maciej Stroinski, Jan Weglarz. IST Grid Projects Inventory and Roadmap, GRIDSTART-IR-D2.2.1.2-V1.3, 14 August 2003.
5. GRIDSTART project website: http://www.gridstart.org
6. Klaus Krauter, Rajkumar Buyya, and Muthucumaru Maheswaran, A Taxonomy and Survey of Grid Resource Management Systems for Distributed Computing, International Journal of Software: Practice and Experience (SPE), ISSN: 0038-0644, Volume 32, Issue 2, 2002, Wiley Press, USA, February 2002.
7. Jarek Nabrzyski, Ariel Oleksiak. Comparison of Grid Middleware in European Grid Projects, 1st European Across Grids Conference, Santiago de Compostella, Spain, February 2003.