Network of Excellence

GRID-based Systems for solving complex problems

European Research Network on foundations, software infrastructures and applications for large-scale distributed Grid and peer-to-peer technologies

Final Activity Report

Start date of project: September 1, 2004
Duration: 48 months

Organization name of lead contractor: ERCIM

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)

Keyword List: industry, sustainability, research application, dissemination

CoreGRID FP6-004265

I. Introduction

The CoreGRID Network is funded by the European Commission within the European Union's Sixth Framework Programme for research and technological development. A grant of 8.2 million has been assigned to the project for a duration of four years starting from September 2004.

CoreGRID comes under the framework of Europe's Information Society Technologies (IST) thematic priority. IST has defined Grid technologies as a crucial objective that will transform the European Union into the most competitive knowledge-based economy in the world.
By providing everyone with immense computing power and knowledge – currently unavailable to even the largest corporations and laboratories – Grids will improve the competitiveness of European industries and mark a new era of markets and services previously perceived as impossible to drive forward. The impact on our quality of life will be profound, allowing us to better monitor and model everything from global climate change to the way cars behave in collisions.

In order to put Europe in front and make sure today's research addresses tomorrow's market needs, CoreGRID is committed to structuring European research by integrating a critical mass of expertise and promoting scientific and technological excellence within and beyond the Grid research community. Through this commitment, CoreGRID is helping Europe to take Grids out of research labs and into industry. This initiative marks a critical step in ensuring that Europe realises the benefits of the information society.

II. Objectives & structure

The primary objective of the CoreGRID Network of Excellence is to build solid methodological and technological foundations for Grid and peer-to-peer technologies, and to stay at the forefront of scientific excellence. This objective has been achieved by structuring integrated research activities carried out by experts in parallel and distributed systems, middleware, programming models, algorithms, tools and environments.

This joint research contributed to realising the CoreGRID vision of a future Grid infrastructure: seamless integration of the existing Grid and other emerging architectures (such as peer-to-peer) using concepts and standards from the World Wide Web Consortium and other relevant standardisation bodies.

To comply with this long-term objective, the CoreGRID Network runs a joint programme of activities (JPA). The JPA integrates and co-ordinates the activities of the major European research teams in the field of Grid and peer-to-peer technologies. Composed of well-established researchers (161 permanent researchers and 164 PhD students) from 46 research centres and universities, the CoreGRID research teams bring high-level expertise in specific areas. They also influence their national Grid and peer-to-peer programmes, fostering better long-term integration.

Operated as the European Grid Research Laboratory, the JPA is structured around six research institutes (see Figure 1). Each institute represents a research area identified as being

of strategic importance to ensure the sustainable development and deployment of Grid infrastructure:

1. Institute on Knowledge and Data Management: Handling information, data, and knowledge that are required or produced by a wide range of diverse processing services.

2. Institute on Programming Model: Making the programming of Grid infrastructures as simple and transparent as possible.

3. Institute on Architectural Issues: Scalability, Dependability, Adaptability: studying adaptive and dependable Grid architectures and services to design next generation Grid middleware. Knowledge Layer, Information Layer, Computation/Data Layer, Middleware, Data, Knowledge Control.

4. Institute on Grid Information, Resource and Workflow Monitoring Services: Provide scalable information services to implement a consistent view of the Grid.

5. Institute on Resource Management and Scheduling: Addressing efficient scheduling and co-ordination of all relevant resources within a Grid environment.

6. Institute on Grid Systems, Tools and Environments: Integrating various middleware, tools and applications for problem solving.

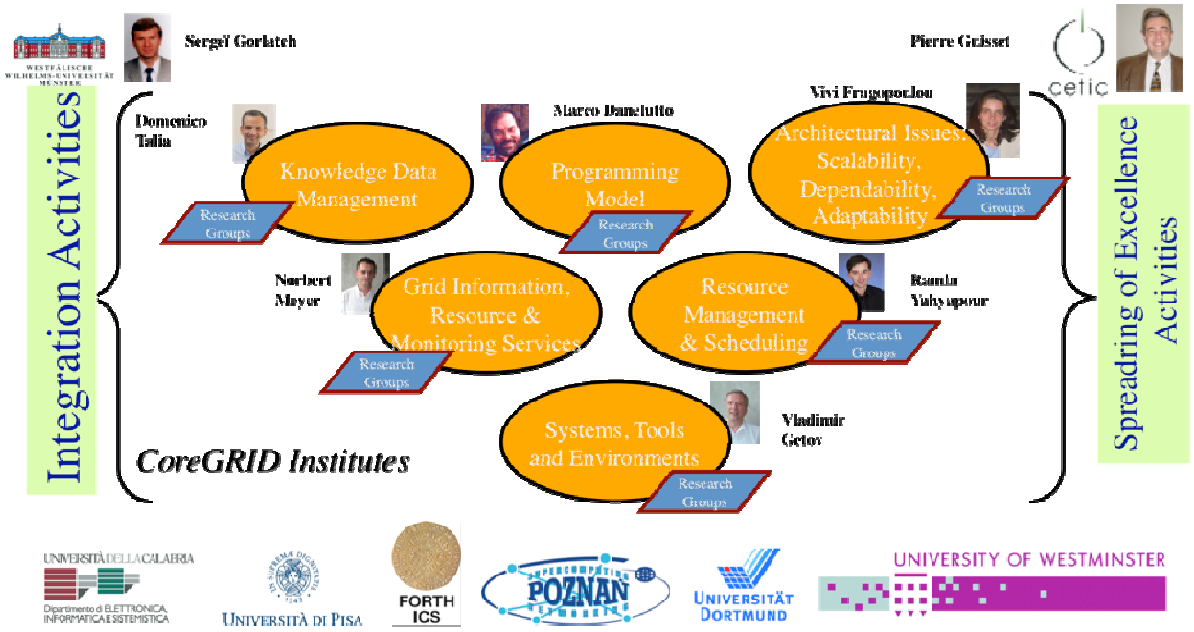


Figure 1: CoreGRID structure

These six institutes benefit from two other type of activities: those increasing integration and fostering collaboration between CoreGRID researchers and those spreading excellence outside the network.

This programme of integrated research activities carried out by the best teams in Europe meets the goal of a Network of Excellence, as defined by the European Commission. Driven

by the principles of integration, dissemination and sustainability, CoreGRID is clearly and successfully moving towards the accomplishment of its vision for the Next Generation Grid.

Figure 2 gives the organizational map of CoreGRID. Two coordinators head the network with the help of an executive committee (gathering all the institute and activity leaders, the chairman of the Members General Assembly and the Integration Monitoring Committee). The Members General Assembly is the voice of the network within which each CoreGRID partner is represented. Two other additional boards (Scientific and Industrial Advisory Boards) give advice to the network coordinators. A newparticipant task force aims at selecting new members (either full or associate) that want to join the network.

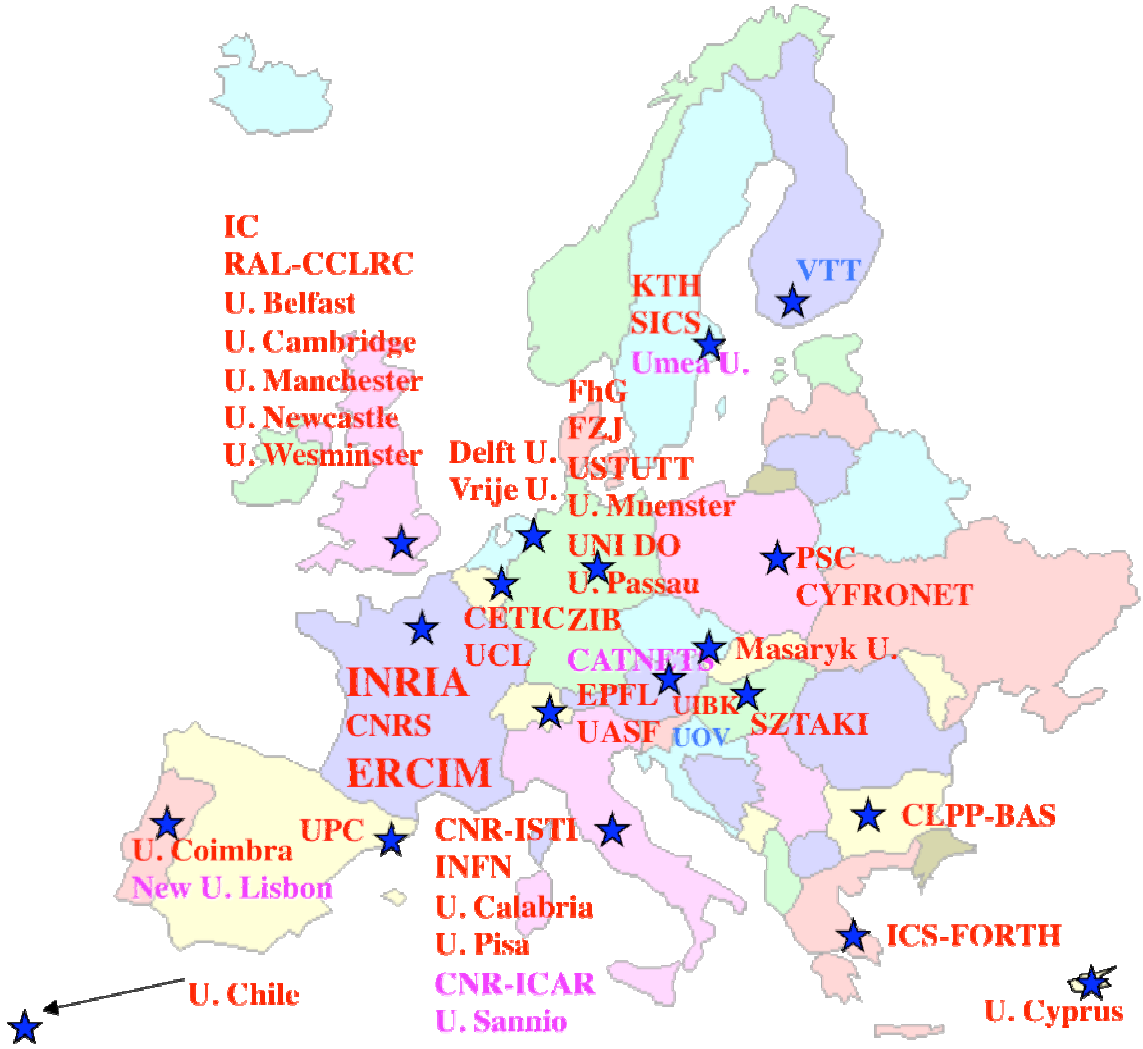
Figure 2: CoreGRID committees and boards

III. The Consortium

CoreGRID involves 39 full partners and 7 associate members from 19 countries (18 from Europe).

Full partner name	Short name	Country
European Consortium for Informatics and Mathematics	ERCIM	France
CETIC	CETIC	Belgium
Institute for Parallel Processing	IPP-BAS	Bulgaria
Information Science and Technologies Institute	CNR-ISTI	Italy
French National Research Council	CNRS	France
Delft University of Technology	TUD	Netherland
Swiss Federal Institutes of Technology	EPFL	Switzerland
Fraunhofer Gesellschaft	FhG	Germany
Forschungszentrum Jülich	FZJ	Germany
High Performance Computing Centre	USTUTT	Germany
Institute of Computer Science, Foundation for Research and Technology - Hellas	ICS-FORTH	Greece

National Institute for Research in Nuclear Physics	INFN	Italy
National Institute for Research in Computer Science and Control	INRIA	France
Royal Institute of Technology	KTH	Sweden
Masaryk University Brno	MU	Czech R.
Poznan Supercomputing and Networking Center	PSNC	Poland
Science and Technology Facilities Counsil	STFC	UK
Swedish Institute of Computer Science	SICS	Sweden
Computer and Automation Research Institute	SZTAKI	Hungary
The Queen's University of Belfast	QUB	UK
Univertity of Muenster	WWU Muenster	Germany
University of Calabria	UNICAL	Italy
University of Cardiff	UWC	UK
University of Chile	UCHILE	Chili
University of Coimbra	UCO	Portugal
University of Cyprus	UCY	Cyprus
University of Dortmund	UNI DO	Germany
Université catholique de Louvain à Louvain-la-Neuve	UCL	Belgium
University of Manchester	UoM	UK
The University of Newcastle Upon Tyne	UNCL	UK
University of Passau	UNI PASSAU	Germany
University of Pisa	UNIPI	Italy
Haute Ecole Spécialisée de Suisse Occidentale	HES-SO	Switzerland
University of Westminster	UoW	UK
Technical University of Catalonia	UPC	Spain
Vrije Universiteit	VUA	Netherland
Zuse Institute Berlin	ZIB	Germany
CYFRONET	CYFRONET	Poland
University of Innsbruck	UIBK	Austria



Associate member name	Country
Institute of High Performance Computing and Networking	Italy
New University of Lisbon	Portugal
CATNETS	Germany
University of Sannio	Italy
Umeå University	Sweden
University of Vienna	Austria
Technical Research Centre of Finland	Finland

IV. Main achievements of the Network

After four years of existence, CoreGRID has carved out a place for itself in the international Grid research arena. It has become one of the largest research centres in Grid computing, encompassing a vast range of research topics such as knowledge and data management, programming models, middleware, resource management and scheduling, workflow, service infrastructures and P2P systems, just to cite a few. It has reached its ideal objective: to become the European Grid beacon. The main achievements are presented below:

- Organisation of 60 meetings and 17 workshops, contributing to solving research challenges as described in the CoreGRID research roadmaps.
- Setting up of a document repository allowing CoreGRID researchers to instantly access documents produced by the network. More than 3000 documents are stored in the repository.
- Delivery of more than 450 joint technical papers accepted in peer-reviewed conferences, workshops and journals.
- Publication of 176 CoreGRID technical reports and 5 white papers, co-authored by at least two different CoreGRID partners showing the level of integration among the CoreGRID community.
- Publication of 11 CoreGRID books published by Springer.
- The effective kick-off of several spin-off research projects or participation to projects, funded either by the European Commission within the 6th and the 7th Framework Programmes, or through national and/or regional initiatives
    - GridComp: STREP - FP6-IST Call 5 (Starting date: 1 June 2006 – Joint partners: INRIA, ERCIM, UoW, CNR, UCHILE)
    - XtreemOS: IP - FP6-IST Call 5 (Starting date: 1 June 2006 – Joint partners INRIA, CNR, ZIB, VUA, STFC, UPC)
    - Selfman: SSA - FP6-IST Call 5 (Starting date: 1 June 2006 – Joint partners INRIA, UCL)
    - GridTrust: SSA- FP6-IST Call 6 (Starting date: 1 June 2006 – Joint partners CETIC, STFC, CNR, VUA)
    - EchoGRID: SSA - FP6-IST Call 6 (Starting date: 1 January 2007 – Joint partner ERCIM)
    - Phosphorus: IP - FP6-IST Call 6 (Starting date: 1 October 2006 – Joint partners PSNC, FhG, FZJ)
    - SmartLM: IP - FP7-ICT-2007-1 - Objective 1.2 (Starting date: 1 February 2008 – Joint members FhG, FZJ)
    - S-Cube: NoE - FP7-ICT-2007-1 - Objective 1.2 (Starting date: 1 March 2008 – Joint members CNR, INRIA, SZTAKI)
    - OGF-Europe: SSA - FP7-ICT-2007-1 - Objective 1.2 (Starting date: 1 March 2008)
- Establishing a database of publications by CoreGRID researchers in the area of Grid and peer-to-peer computing with around 1,050 references available today on the CoreGRID web site.
- A researcher's database with more than 270 entries to allow any CoreGRID researcher to quickly identify the best experts on a given research topic related to Grid and P2P computing, thereby facilitating joint research activities.
- Increased visibility of the Grid research community through the support of highly-reputed international conferences, such as EuroPar 2005, 2006 and 2007, HPDC 2006, IEEE conference on Grid Computing in 2006 and 2007 and events such as Grid@Work 2006 and 2007. CoreGRID also sponsored the Open Grid Forum as a silver member.
- Opening up academic research agendas in order to identify business-oriented research priorities, leading to the spin-off of new CoreGRID activities, for example in service-oriented architectures and systems.
- Developing new ideas to anticipate technological trends and to promote commercially relevant and promising research.

- Active involvement of industrial stakeholders to help identify take-up opportunities beyond publicly funded programmes, thus stimulating the involvement of industrial stakeholders in CoreGRID institutes. 4 meetings were organized with our Industrial advisory board.
- Setting up a Grid User Community in order to raise public awareness with 6 press releases, 3 press conferences and 53 press clippings. We also disseminated 3 annual reports (2005, 2006 and 2007) and 9 newsletters to a wide public audience. 2 brochures and posters were used to explain the objectives of CoreGRID during public events.
- Organisation of several scientific workshops jointly with highly reputed international conferences and the CoreGRID symposium jointly held with EuroPar 2007 and EuroPar 2008.
- Organisation of three annual CoreGRID Summer Schools.
- Implementing Mobility Programmes - a Fellowship Programme and a Researcher Exchange Programme - increasing integration between partners, and now involving industrial members of the IAB. A total of 45 researcher exchanges and 19 fellowships.
- Organisation of three annual Integration Workshops.

V. CoreGRID Institutes

1) Knowledge and Data Management

a. Objectives

Grids are changing their role, moving from a computation and data management platform to a pervasive information and knowledge management infrastructure. This trend requires new models, services and technologies so as to enable Grid computing systems to manage distributed data and knowledge, thus enabling complex applications according to the SOKU model.

The INSTITUTE ON KNOWLEDGE AND DATA MANAGEMENT (KDM) joins together thirteen institutions from eight European countries and involves more than 50 senior researchers and PhD students. The general goal of this Institute is to further integrate data management and knowledge discovery solutions with Grid technologies for providing data- and knowledge-intensive Grids. The Institute's approach is a vertical approach that encompasses all layers involved with knowledge management in Grids: storage management at the systems-level, information and knowledge management, and knowledge discovery. At the same time, single KDM solutions that bring benefits to the Grid community have been designed and developed.

A key KDM objective is to design and develop common solutions for data management and knowledge discovery and management on Grids. This promotes the wide diffusion and use of knowledge-based Grid services for the Semantic Grid and the Knowledge Grid. To this end, the partners of the KDM Institute focus on the problems of providing commodity-based connectivity among heterogeneous distributed storage devices, management automation of administration tasks traditionally handled manually, and storage virtualisation for serving well-defined requirements from multiple users.

The Institute provides a collaborative setting for European research teams working on: distributed storage management on Grids; knowledge techniques and tools for data-intensive applications; security and trust mechanisms for storage and data; and integration of data and computation Grids with information and knowledge Grids.
The goal of the second year has been to consolidate and expand the joint activity of research groups, thus promoting larger leading teams and supporting efforts towards standard models, services, middleware and solutions.

During the last four years, the Institute members (CETIC, FORTH, CNR-ICAR, INFN, CNR-ISTI, PSNC, STFC-RAL, SZTAKI, Universidade Nova de Lisboa, University of Calabria, University of Cyprus, University of Manchester, and University of Newcastle) worked on three main tasks:

1 DISTRIBUTED STORAGE MANAGEMENT
Providing infrastructures, techniques and policies for managing distributed storage resources in the Grid.
2 INFORMATION AND KNOWLEDGE MANAGEMENT
Developing metadata, semantic representation and protocols for Grid service discovery, information management and design of knowledge-oriented Grid services.
3 DATA MINING AND KNOWLEDGE DISCOVERY
Designing Grid services for distributed data mining and knowledge discovery on Grids and P2P systems.

In all those areas, the partners jointly produced scientific results implemented in software prototypes and published in scientific journals and conference proceedings. Additionally, security issues have been covered with a particular focus on security requirements and models for storage and data management.

b. Achievements

Several tangible results are today visible in terms of joint research activities: new European projects involving KDM partners, joint publications and research visit exchanges. The main research contributions of the Institute covers the following topics:

- Data storage access and management architecture
- Storage security
- Data Integration Models and Architectures
- Methods for deriving GRID trust and security policies for managing VOs
- Distributed Data Mining in GRIDs and P2P Systems
- Adaptivity in Distributed Query and Workflow
- Metadata and Semantics for Grid Information Services
- Discovery in Large Collaborative Networks

To effectively show the impact of the Institute in terms of integration, it can be noticed that the number of joint publications involving KDM partners was rather low at the beginning of the network with only two clusters of partners (UCY/UoM and UNICAL/CNR-ICAR). After four years, all the partners involved in the institute have contributed to at least one joint publication with another partner. Moreover, one book on “Knowledge and Data Management in GRIDs”, co-edited by the KDM task leaders, has been published in the CoreGRID Springer

series, and another three post-conference proceedings were published by Springer as a result of the three workshops on Grid Middleware co-organized with three other CoreGRID institutes. Finally the institute published 38 Technical reports.

It has also to be noticed that the KDM institute has established a close cooperation with Hitachi with a view to transfer some of its research results. This collaboration has been implemented thanks to the Industrial Fellowship Programme. A post-doc was hire to start the collaboration between CETIC, UNICAL and Hitachi at Sophia Antipolis.

Overall, the institute has contributed to solving the following research problems:

- Information Services for Large-Scale Grids A Case for a Grid Search Engine
- Data integration and query reformulation in service-based Grids: Architecture and Roadmap
- Tree Vector Indexes: Efficient Range Queries for Dynamic Content on Peer-to-Peer Networks
- Semantic Support for Meta-Scheduling in Grids
- A Study of Languages for the Specification of Grid Security Policies
- Security Requirements Analysis for FileStamp Distributed File System
- Systems and techniques for distributed and stream data mining
- Algorithms and frameworks for stream mining and knowledge discovery on Grids
- A DHT-based Peer-to-Peer Framework for Resource Discovery in Grids
- Modeling and Supporting Grid Scheduling
- Conductor: Support for Autonomous Configuration of Storage Systems
- Analyzing the Workload of the South-East Federation of the EGEE Grid Infrastructure
- Reputation-based trust management systems and their applicability to grids
- A Super-Peer Model for Multiple Job Submission on a Grid
- A Service-Oriented System to Support Data Integration on Data Grids
- Modular Adaptive Query Processing for Service-Based Grids
- Adapting to Changing Resource Performance in Grid Query Processing
- Peer-to-Peer Metadata Management for Knowledge Discovery Applications in Grids
- Multi-set DHT for interval queries on dynamic data
- An Analysis of GRID Storage Element Architectures: High-end Fiber-Channel vs. Emerging Cluster-based Networked Storage
- An Analysis of Security Services in Grid Storage Systems
- An ActOn-based Semantic Information Service for EGEE
- Use of P2P Overlays for Distributed Data Caching in Public Scientific Computing
- Self-optimizing Block Transfer in Web Service Grids
- A P2P Job Assignment Protocol for Volunteer Computing Systems
- Designing data analysis services in the Knowledge Grid
- Mechanisms for High Volume Data Transfer in Grids
- Defeating Colluding Nodes in Desktop Grid Computing Platforms
- Data Consistency and Peer Synchronization in Cooperative P2P Environments
- Distributed Data Mining in Desktop Grids
- Providing security to the Desktop Data Grid
- A data-centric security analysis of ICGrid
- A Scalable Architecture for Discovery and Planning in P2P Service Networks
- Usage Control in Data Grids

c. Future roadmap and vision

Research activities in the area of data and knowledge intensive services and systems in Grids, and P2P systems are being pursued in Europe, in USA and in Asia by many research teams and, at the same time, big companies such as IBM, Yahoo!, Amazon, Google, HP, ORACLE, and SUN Microsystems, are very active in the area with a special focus on service design and implementation. This demonstrates the key role of data management in such distributed infrastructures and the importance of developing knowledge-based applications that exploit GRID, P2P and Cloud features to achieve high performance and high availability. CoreGRID KDM researchers will work towards providing a variety of models, architectures, prototypes, and services offering different technological solutions to current problems faced by applications in Grid and dynamic distributed infrastructures.

Future research activities will focus on core technologies needed to implement the SOKU model, including ontologies, data mining and knowledge discovery, data management, also taking into account SLAs. Partners of the KDM Institute will proceed beyond the funded period and will continue their joint activities by addressing key research challenges in the KDM area such as:

- Support for autonomous storage systems and management of replication, consistency and data placement;
- Metadata definition and searching for scalable information systems;
- Distributed content-based retrieval services;
- Study of scalable services for accessing, querying and managing data centres;
- Design of services for open analytics environments for science and business;
- Data intensive workflows: service scheduling, data security and privacy, data partitioning, techniques to bring computations as close to the data as possible;
- Web Mashup for data management.

d. List of selected publications

- Wei Xing, Marios Dikaiakos and Rizos Sakellariou. A Core Grid Ontology for the Semantic Grid. Proceedings of the 6th IEEE International Symposium on Cluster Computing and the Grid (CCGrid 2006), Pages 178-184, IEEE Computer Society, Singapore, May 2006.
- Anastasios Gounaris, Rizos Sakellariou, Norman W. Paton and Alvaro A.A. Fernandes. A novel approach to resource scheduling for parallel query processing on computational grids. Distributed and Parallel Databases, Vol. 19(2-3):87-109, May 2006.
- Anastasios Gounaris, Carmela Comito, Rizos Sakellariou and Domenico Talia. A Service-Oriented System to Support Data Integration on Data Grids. Proceedings of Seventh IEEE International Symposium on Cluster Computing and the Grid (CCGrid 2007), Rio de Janeiro, Brazil, IEEE Computer Society, May 2007.
- Domenico Talia, Angelos Bilas and Marios Dikaiakos (editors). Knowledge and Data Management in Grids. Springer, December 2006.
- Pasquale Cozza, Carlo Mastroianni, Domenico Talia and Ian Taylor. A Super-Peer Protocol for Multiple Job Submission on a Grid. In Wolfgang Lehner, Norbert Meyer, Achim Streit and Craig Stewart editors, Euro-Par 2006 Workshops, Vol. 4375:116-125 of LNCS, Springer Berlin/Heidelberg, Dresden, Germany, June 2007.

- Syed Naqvi, Philippe Massonet and Alvaro Arenas. Pragmatic Security Analysis of the Grid - A Requirements Engineering Perspective. Proceedings of the 4th International Conference on Information Systems Security 2006, 2006.
- Domenico Talia, Paolo Trunfio, Salvatore Orlando, Raffaele Perego and Claudio Silvestri. Systems and techniques for distributed and stream data mining. Technical report, TR-0045, Institute on Knowledge and Data Management, CoreGRID - Network of Excellence, July 2006.
- Carlo Mastroianni, Pasquale Cozza, Domenico Talia, Ian Kelley and Ian Taylor. A Scalable Super-Peer Approach for Public Scientific Computation. Future Generation Computer Systems, Elsevier Science, 2008.
- Carlo Mastroianni, Giuseppe Pirrò and Domenico Talia. Data Consistency in a P2P Knowledge Management Platform. Proceedings of the HPDC 2007 Conference & Co-Located Workshops, Pages 17-24, ACM Press, Monterey, California, USA, June 2007.
- Eddie Al-Shakarchi, Pasquale Cozza, Andrew Harrison, Carlo Mastroianni, Matthew S. Shields, Domenico Talia and Ian Taylor. Distributing workflows over a ubiquitous P2P network. Scientific Programming, Vol. 15(4):269-281, IOS Press, Amsterdam, The Netherlands, 2007.
- Massimo Coppola, Yvon Jégou, Brian Matthews, Christine Morin, Luis Pablo Prieto, Oscar David Sanchez, Erika Yang and Haiyan Yu. Virtual Organization Support within a Grid-Wide Operating System. IEEE Internet Computing, 2008.
- Andrea Pugliese, Domenico Talia and Ramin Yahyapour. Modeling and supporting Grid scheduling. Journal of Grid Computing, Vol. 6(2):195-213, 2008.
- Anastasios Gounaris, Christos Yfoulis, Rizos Sakellariou and Marios D. Dikaiakos, A Control Theoretical Approach to Self-optimizing Block Transfer in Web Service Grids. ACM Transactions on Autonomous and Adaptive Systems, Vol. 3(2): 1-30, 2008

CoreGRID White Papers:

- A Marcin Adamski, Alvaro Arenas, Angelos Bilas, Paraskevi Fragopoulou, Vasil Georgiev, Alejandro Hevia, Gracjan Jankowski, Brian Matthews, Norbert Meyer, Jorg Platte, Michael Wilson, **Trust and Security in Grids: A State of the Art**, CoreGRID White Paper – WHP-0001
- lvaro Arenas, Angelos Bilas, Jesus Luna, Manolis Marazakis, Carmela Comito, Domenico Talia, Marios D. Dikaiakos, Anastasios Gounaris, Philippe Massonet, Syed Naqvi, Jim Smith, Paul Watson, Federico Stagni. **Knowledge and Data Management in Grids: Notes on the State of the Art**. CoreGRID White Paper –WHP-0002

2) Programming Model

a. Objectives

The success of Grid computing infrastructures, like any other computing infrastructures, depends mainly on the availability of a large catalogue of applications designed to run on them. However, developing applications for Grid infrastructures remains a complex task due to the low level of abstractions exposed to the programmers. Moreover, applications developed for a specific Grid infrastructure cannot be executed on other ones due to the dependencies upon some specific hardware resources. Therefore, programming the Grids

remains a research challenge and this is why CoreGRID devoted an Institute to gather European experts involved in this research field. The institute on Programming Model has four main objectives in its quest to define a suitable programming model:

- Better programmability and productivity by providing high-level abstraction
- Higher scalability and ability to manage the heterogeneity of both hardware and software resources
- Better deployment to allow transparent remote execution of software codes
- Better efficiency by dynamically adapting software codes to the computing resources provided by the Grid

b. Achievements

The Programming Model Institute brings together 13 full partners and 2 associate partners, with a total of more than 45 researchers and about 30 PhD students. Altogether, these researchers have contributed to an extensive programme of short visits, to the Researcher Exchange Programme, to CoreGRID fellowships often involving other CoreGRID Institutes, and have published a significant number of joint research papers in international journals and presented them at conferences.

To reach the four objectives aforementioned, the activities of the Programming Model Institute, since the very beginning, have been aimed at developing a new component model for the Grid, providing basic composition and management primitives. Since the design and the implementation of such a new component model by far exceeds the budget capacity of the CoreGRID network, we decided to focus mainly on the design while the implementation was assigned to another EU funded project (GridCOMP) as a spin-off project from CoreGRID. After four years of activity, CoreGRID has issued a complete specification of the Grid Component Model (GCM) that integrates several research results provided by the researchers involved in this institute. This has to be considered as the success story of the Programming Model Institute. This area is a very good example of a successful integration of the research coming from European researchers.

Before CoreGRID, there were a dozen national projects aiming at designing programming models for Grids: Assist (U. Pisa / CNR), Fractal (INRIA), GridCCM (INRIA), HOC-SA (U. Muenster), Ibis (Vrije U. Amsterdam), ICENI (IC), MALLBA (UPC), Mocca (Cyfronet), Polytop (U. Passau), POP-C++ (UASF), ProActive (INRIA), and REFLEX (U. Chile). This scattered approach within the EU did not help the European researchers to be highly visible in the research arena at international level. In the area of component model, the most visible project was the Common Component Architecture consortium which is a group of researchers from the US national labs and academic institutions committed to defining a standard component architecture for high performance computing including Grids.

In 2003, this consortium has initiated a series of workshops, called CompFrame. At that time it has to be noticed that most of the Program Committee members were from the USA with very few Europeans although one of the goal of these workshops was to establish cooperation between US and EU researchers. Thanks to CoreGRID, we successfully balanced the participation of EU researchers within this initiative by organizing a series of workshops called HPC-GECO, as of 2006. Thanks to our stronger visibility in the area of Component models for the Grids, we were able to merge the two series of workshops into the HPC-GECO/CompFrame workshops. Without CoreGRID, this merge would not have been

possible. This is a perfect example of the results of the integration we achieved, that brought us stronger visibility at international level.

In term of dissemination, the institute published 24 Technical Reports in the CoreGRID series, and an average of 20 to 30 joint (i.e. signed by more than one single partner) papers on Institute research topics in international conferences and journals. Partners of the Institute regularly contributed to CoreGRID Summer schools, as well as to the Integration workshops and to CoreGRID Symposia. Also, members of the Institute participated to several scientific events and activities due to their "Programming model Institute" affiliation. As an example, Programming model Institute researchers participated to NGG EU meetings, contributed to the NESSI initiatives and research agenda documents, were involved in the organisation of several grid/service related international conferences/workshops/events. Moreover, one book on "Component Models and Systems for Grid Applications" has been published in the CoreGRID Springer series at the beginning of the network. This book is the proceedings of the Workshop on Component Models and Systems for Grid Applications held on June 26, 2004 in Saint Malo, France jointly with the ICS04 conference. A second book "Making Grids Work" was also published following a workshop that was held in Heraklion in June 2007.

Last but not least, a standardisation process has been initiated through ETSI[1] that will eventually result in a complete GCM standard. The first steps have been performed and the GCM ADL is close to being an ETSI standard.

Overall, the institute has contributed to investigating and to solving, among others, the following research problems:

- Optimization Techniques for Implementing Parallel Skeletons in Distributed Environments
- Behaviour Customization of Parallel Components for Grid Application Programming
- Characterization of the performance of ASSIST programs
- Parallel program/component adaptivity management
- Automatic mapping of ASSIST applications using process algebra
- User-Transparent Scheduling of Structured Parallel Applications in Grid Environments
- Specification and Verification of Reconfiguration Protocols in Grid Component Systems
- Fault-Tolerant Data Sharing for High-level Grid Programming:
- Hierarchical Storage Architecture
- Deriving Grid Applications from Abstract Models
- User-Transparent Scheduling for Software Components on the Grid
- Prototyping and reasoning about distributed systems: an Orc based framework
- Cost of Task Re-Scheduling in Fault-Tolerant Task Parallel Computations
- Towards hierarchical management of autonomic components: a case study
- LooPo-HOC: A Grid Component with Embedded Loop Parallelization
- Grid Environment for Real-Time Multiplayer Online Games

[1] ETSI is The European Telecommunications Standards Institute (ETSI) that produces globally-applicable standards for Information and Communications Technologies (ICT), including fixed, mobile, radio, converged, broadcast and internet technologies. See http://www.etsi.org for more information.

- Measurable Component Values and Services: the Economic Planning of Resource Transactions
- Enhancing Grids for Massively Multiplayer Online Computer Games
- Towards Software Component Assembly Language Enhanced with Workflows and Skeletons

c. Future roadmap and vision

The huge activity of the Programming Model Institute resulted in the design of GCM and in the assessment of its main features. Now this result must be consolidated and promoted in all the relevant Grid related contexts. In turn, this identifies some precise challenges that can be understood as the "sustainability roadmap goal" of the Programming Model Institute:

- *Complete integration of GCM with the software service framework.* Although Institute partners already partially demonstrated the feasibility of porting GCM concepts on top of the Service Component Architecture, much work has to be done to guarantee that the advanced concepts introduced in GCM (such as collective communication patterns or autonomic managers) can be migrated to the software service scenario. Because of the increased importance of the service oriented paradigm in Grids, clouds and in general in the distributed architecture scenario, this integration should definitely be considered as being important.
- *Implementation of effective advanced programming models on top of GCM*. There are several kinds of advanced programming models considered to further raise the level of abstraction presented to Grid application programmes, ranging from skeleton-based structured programming paradigms to component paradigms combining spatial (component-like) and temporal (à la workflow) composition of components. The development of effective, advanced programming models will provide the Grid/SOA user/programmer the advantages coming from the results achieved with GCM without exposing the full complexity of GCM program development methodology.
- *Development of a complete set of formal (or semi-formal) tools supporting Grid program design, development and tuning.* The multiple, partial achievements of the Programming Model Institute in this field should be exploited to provide some general purpose, user friendly tool actually supporting the activities of programmers, from design to fine tuning.

d. List of selected publications

- M. Aldinucci, S. Campa, M. Danelutto, M. Vanneschi, P. Kilpatrick, P. Dazzi, D. Laforenza and N. Tonellotto, Behavioural skeletons in GCM: autonomic management of grid components, in: Proc. of Intl. Euromicro PDP 2008: Parallel Distributed and network-based Processing, pages 54-63, IEEE, 2008
- M. Aldinucci, M. Danelutto, H. Lilia Bouziane and C. Pérez, Towards Software Component Assembly Language Enhanced with Workflows and Skeletons, in: Proc. of the ACM SIGPLAN Component-Based High Performance Computing (CBHPC), 2008
- S. Gorlatch, F. Glinka, A. Ploß and J. Müller-Iden (University of Münster), R. Prodan, V. Nae and Thomas Fahringer (University of Innsbruck), Enhancing Grids for Massively Multiplayer Online Computer Games, Institute on Programming Model & Institute on Grid Information, Resource and Workflow Monitoring Services &

Institute on Resource Management and Scheduling, CoreGRID Technical Report, TR-0134, June 17, 2008
- C. Dumitrescu, J. Dünnweber and S. Gorlatch (University of Münster), D.H.J. Epema (Delft University of Technology), User-Transparent Scheduling for Software Components on the Grid, Institute on Programming Model & Institute on Resource Management and Scheduling, CoreGRID Technical Report, TR-0086, 11 May, 2007
- A. Stewart, M. Clint, J. Gabarro, T. Hammer, P. Kilpatrick and R. Perrott. "Managing Grid Computations: An ORC-Based Approach". ISPA-2006, LNCS 4330, 278-291, Dec. 2006..
- J. Gabarró, A. García, M. Clint, P. Kilpatrick, A. Stewart. Bounded Site Failures: An Approach to Unreliable Grid Environments. In *Making Grids Work*, Springer 2008
- F. Baude, D. Caromel, M. Danelutto, V. Getov, L. Henrio, C. Pérez, GCM: A Grid Extension to Fractal for Autonomous Distributed Components, Annals of Telecommunications – 2008
- E. Tejedor, R. Badia, P. Naoumenko, M. Rivera, C. Dalmasso, Orchestrating a safe functional suspension of GCM components, Integrated Research in Grid Computing, Proceedings of the CoreGRID Integration Workshop 2008, Crete University Press, 2008, ISBN 978-960-524-260-2
- H. Bouziane, C. Pères, T. Priol, A Software Component Model with Spatial and Temporal Compositions for Grid Infrastructures, Proceedings of EuroPar 2008, LNCS No. 5168, Springer Verlag, 2008
- Natalia Currle-Linde, Michael Resch (HLRS), Christian Pérez (INRIA/IRISA), Massimo Coppola (University of Pisa/CNR-ISTI), Measurable Component Values and Services: the Economic Planning of Resource Transactions, Institute on Grid Systems, Tools, and Environments & Institute of Programming Model, CoreGRID Technical Report, TR-0125, May 19, 2008
- M. Leyton, D. Caromel, A Transparent non-Invasive File Data Model for Algorithmic Skeletons, IEEE International Symposium on Parallel and Distributed Processing. IPDPS 2008, IEEE Press, 2008

CoreGRID White Paper:
- Marcin Adamski, Alvaro Arenas, Angelos Bilas, Paraskevi Fragopoulou, Vasil Georgiev, Alejandro Hevia, Gracjan Jankowski, Brian Matthews, Norbert Meyer, Jorg Platte, Michael Wilson**, Trust and Security in Grids: A State of the Art**, CoreGRID White Paper – WHP-0001

3) Architectural issues: scalability, dependability, adaptability

a. Objectives

The main goal of the Institute on Architectural Issues (System Architecture - SA) is to provide the techniques that will pave the way towards scalable, adaptable, and dependable Grid architectures and services meeting the mandatory properties of the Next Generation Grids. The main objective of the SA Institute is to significantly improve architectural designs of Next Generation Grids by performing collaborative research on the following key architectural aspects:

- To apply and extend results from research on peer-to-peer systems for enabling higher scalability and self-organisation of Grid infrastructures.
- To investigate the mechanisms for fault-tolerance and robustness of the Grid infrastructure to assure reliable Grid services.
- To study methods for adaptability and self-management in order to establish paradigms for automatic and low-cost Grid management.

To achieve scalability we extend the results from research on peer-to-peer systems and propose scalable approaches for resource discovery. We propose dependability mechanisms for all levels of the Grid including checkpointing and recovery. We devise techniques for fault tolerance and robustness for Grid services, providing a more reliable Grid architecture. In adaptability and self-management, we develop mechanisms for automated adaptation and reconfiguration of Grid infrastructure. Through these objectives, the SA Institute aims to contribute to the mandatory architectural principles of the Next Generation Grids by scalability of services, resilience, straightforward administration and configuration management.

b. Achievements

The partners involved in the SA Institute are the following: FORTH-ICS, INRIA, KTH, SICS, MTA SZTAKI, UNICAL, UCO, UCY, UCL, UoM, UoW UPC and ZIB. Around 60 senior researchers as well as PhD students are involved in this institute.

The scientific achievements of the SA Institute could be briefly summarised as follows:

- The design of a hybrid P2P-based system that works efficiently for multi-attribute queries on static and dynamic attributes and its experimental evaluation on existing Grid platforms (Grid'5000).
- The development of dependability mechanisms based on virtualisation techniques that lead to zero downtime. The proposed solution is applicable to individual servers, clusters, as well as Grid infrastructures.
- Modelling and prediction of workloads and system behaviour with impact to adaptable solutions for self-managing systems that provides support for scheduling.
- Refinement of scalable, dependable and adaptable Grid solutions, and their experimentation of real Grid platforms.
- Investigation of the fault conditions of existing Grid infrastructures (e.g. EGEE) towards the development of a framework that will propose concrete solutions for low-cost automatic fault management in Next Generation Grids.

The SELFMAN FP6 EU project is a spin-off of CoreGRID involving mainly partners from the SA Institute and industrial partners. This project is related to the areas of peer-to-peer computing and adaptability/self-management. This collaboration provides the SA Institute with real-world use cases, especially for dependability and self-managing applications. Other EU spin-off projects were proposed during the last EU call from partners of the SA Institute and industry.

The Institute partners have produced 42 CoreGRID Technical Reports containing their joint research activities and presented their work in workshops organized by CoreGRID. A special issue of the Future Generation Computer Systems Journal with selected papers from the Architectural Issues Institute appeared in February 2007. Recently, a CoreGRID Springer volume, entitled "Making Grids Work" was jointly edited by three other CoreGRID institutes.

This volume includes a selection of papers presented in the CoreGRID Workshop held at FORTH-ICS, Heraklion, Crete, Greece in June 2007. Lastly, a special issue of the Parallel Processing Letters (PPL) Journal will appear in September 2008, containing a selection of the latest research results from the Architectural Issues Institute.

Overall, the institute has contributed to solving the following research problems:

- Scalable Desktop Grid System
- Tree Vector Indexes: Efficient Range Queries for Dynamic Content on Peer-to-Peer Networks
- Self Management of Large-Scale Distributed Systems by Combining Peer-to-Peer Networks and Components
- Classifier-based Capacity Prediction for Desktop Grids
- Scheduling for Fast Turnaround Time on Institutional Desktop grid
- Peer-to-Peer Models for Resource Discovery on Grids
- A Feedback Based Approach to Reduce Duplicate Messages in Unstructured Peer-to-Peer Systems
- Characterizing Result Errors in Internet Desktop Grids
- An Overview of Existing Tools for Fault-Injection and Dependability Benchmarking in Grids
- Deterministic Models of Software Aging and Optimal
- Rejuvenation Schedules
- A DHT-based Peer-to-Peer Framework for Resource Discovery in Grids
- Failure Management in Grids: The Case of the EGEE Infrastructure
- Validating Desktop Grid Results By Comparing Intermediate Checkpoints
- Benchmarking the OGSA-DAI Middleware
- Reputation-based trust management systems and their applicability to grids
- Divide et Impera: Partitioning Unstructured Peer-to-Peer Systems to Improve Resource Location
- A Scalable Multi-Agent Architecture for Remote Failure Detection in Web-Sites
- Peer-To-Peer Techniques for Data Distribution in Desktop Grid Computing Platforms
- A Fault-Injector Tool to Evaluate Failure Detectors in Grid-Services
- Using Micro-Reboots to Improve Software Rejuvenation in Apache Tomcat
- SZTAKI Desktop Grid: Building a scalable, secure platform for Desktop Grid Computing
- Design and Implementation of a Hybrid P2P-based Grid Resource Discovery System
- Use of P2P Overlays for Distributed Data Caching in Public Scientific Computing
- Self-optimizing Block Transfer in Web Service Grids
- A P2P Job Assignment Protocol for Volunteer Computing Systems
- Implementing Dynamic Querying Search in k-ary DHT-based Overlays
- Defeating Colluding Nodes in Desktop Grid Computing Platforms
- Data Consistency and Peer Synchronization in Cooperative P2P Environments
- Optimizing the Data Distribution Layer of BOINC with BitTorrent
- Distributed Data Mining in Desktop Grids
- A Scalable Architecture for Discovery and Planning in P2P Service Networks
- Using Virtualization to Improve Software Rejuvenation
- High-Available Grid Services through the use of Virtualized Clustering
- Dependable Grid Services: A Case Study with OGSA-DAI

- Carrying the Crash-only Software Concept to the Legacy Application Servers
- Towards Self-adaptable monitoring framework for self-healing
- Adaptive Distributed Mechanism Against Flooding Network Attacks Based on Machine Learning
- Metadata Ranking and Pruning for Failure Detection in Grids

c. Future roadmap and vision

The challenges faced by the SA Institute for System Architecture constitute the cornerstone upon which any Grid middleware should be built. During the CoreGRID project, concrete solutions have been proposed and major steps have been made towards the invisible Grid. However, many several obstacles have to be overcome before the objectives of the Next Generation Grid are achieved. Grids have to become more robust and pervasive.

Current large-scale distributed computing systems and infrastructure, such as Grid and P2P systems, ad hoc wireless, sensor, and vehicular networks, have the characteristic of being decentralized, and composed of a large number of autonomous entities. The complexity of these systems is such that human administration is nearly impossible and centralized or hierarchical control is highly inefficient. Moreover, often these systems need to run on highly dynamic environments, where content, network topologies and workloads are continuously changing. These systems are also characterized by the high degree of volatility of their components and the need to handle efficiently large amounts of data. Thus, design for adaptation becomes a key feature.

Novel approaches need to be devised for the construction of scalable and efficient large-scale distributed computing systems that need to have the following properties: “self-organization” (Grid components are autonomous and do not rely on any external supervisor), decentralization (decisions are to be taken only on the basis of local information) and adaptive nature (mechanisms must be provided to cope with the dynamic characteristics of hosts and resources). The need for self-directing and self-managing can be achieved using a number of entities that perform simple activities in an autonomous fashion, a behaviour refereed to as “swarm intelligence”.

In another line of research virtualization technologies have emerged as a promising solution to overcome today’s barriers. The challenge is to devise efficient methods based on virtualization to dynamically shape and adapt large scale Grid infrastructures. Additionally this challenge includes the study of heuristics to adapt the computational capacity of the Grid using Cloud providers, thus providing adaptive and elastic management of the Grid infrastructure.

d. List of selected publications

- Javier Alonso, Luis Moura Silva, Artur Andrzejak, Jordi Torres, “High-Available Grid Services through the use of Virtualized Clustering”, 8th IEEE/ACM Int. Conf. on Grid Computing, GRID 2007, Austin, USA, September 2007.
- Demetrios Zeinalipour-Yazti, Kyriakos Neocleous, Chryssis Georgiou, Marios Dikaiakos, “Identifying Failures in Grids through Monitoring and Ranking”, in Proc. of the 7th IEEE International Symposium on Network Computing and Applications (NCA 2008), Cambridge, MA, 2008.
- Anastasios Gounaris, Christos Yfoulis, Rizos Sakellariou, Marios D. Dikaiakos, “Self-Optimizing Block Transfer in Web Service Grids”, in Proc. of the 9th Annual ACM

International Workshop on Web Information and Data Management (WIDM'07), Lisbon, Portugal, Nov. 2007.

- Fernando Costa, Luis Silva, Gilles Fedak, Ian Kelley, "Optimizing the Data Distribution Layer of BOINC with BitTorrent", Second Workshop on Desktop Grids and Volunteer Computing (PCGRID 2008) held in conjunction with IPDPS 2008, IEEE Computer Society, Miami, Florida, USA, April 2008.
- Gheorghe Cosmin Silaghi, Filipe Araujo, Luis Moura Silva, Patricio Domingues, Alvaro E. Arenas, "Defeating Colluding Nodes in Desktop Grid Computing Platforms", Second Workshop on Desktop Grids and Volunteer Computing Systems (PCGrid 2008), Miami, Florida, USA, April 2008.

- G. Pirrò, M. Ruffolo, D. Talia, "Advanced Semantic Search and Retrieval in a Collaborative Peer-to-Peer System", in Proc. of the HPDC 2008 Conference & Co-Located Workshops, Boston, Massachusset, USA, ACM Press, June 2008.
- Artur Andrzejak, Monika Moser, Luis Silva, "Managing Performance of Aging Applications via Synchronized Replica Rejuvenation", 18th IFIP/IEEE Distributed Systems: Operations and Management (DSOM 2007), Silicon Valley, CA, USA, October 29-31, 2007.
- Artur Andrzejak, Luis Silva, "Using Machine Learning for Non-Intrusive Modeling and Prediction of Software Aging", IEEE/IFIP Network Operations and Management Symposium (NOMS 2008), Salvador de Bahia, Brazil, April 7-11, 2008 (Best paper award).
- Artur Andrzejak, Derrick Kondo, David P. Anderson, "Ensuring Collective Availability in Volatile Resource Pools via Forecasting", 19th IFIP/IEEE Distributed Systems: Operations and Management (DSOM 2008) (part of Manweek 2008), Samos Island, Greece, September 22-26, 2008.
- Luis Moura Silva, Paulo Silva, Javier Alonso, Jordi Torres, Artur Andrzejak, "Using Virtualization to Improve Software Rejuvenation", 7th IEEE International Symposium on Network Computing and Applications (IEEE NCA'07), Cambridge, MA, USA, July 2007 (Best paper award).
- Agostino Forestiero, Carlo Mastroianni, Harris Papadakis, Paraskevi Fragopoulou, Alberto Troisi, Eugenio Zimeo, "A Scalable Architecture for Discovery and Planning in P2P Service Networks", Institute on Knowledge and Data Management & Institute on Architectural Issues: Scalability, Dependability, Adaptability, CoreGRID Technical Report, TR-0152, June 17, 2008.
- Carlo Mastroianni, Giuseppe Pirrò, Domenico Talia, "Data Consistency and Peer Synchronization in Cooperative P2P Environments", Institute on Knowledge and Data Management & Institute on Architectural issues: Scalability, Dependability, Adaptability, CoreGRID Technical Report, TR-0128, February 12, 2008.

- Paolo Trunfio, Domenico Talia, Ali Ghodsi, Seif Haridi, "Implementing Dynamic Querying Search in k-ary DHT-based Overlays", Institute on Architectural issues: scalability, dependability, adaptability, CoreGRID Technical Report, TR-0119, December 28, 2007**.**

- Paulo Silva, Luis Silva, Artur Andrzejak, "Using Micro-Reboots to Improve Software Rejuvenation in Apache Tomcat", Institute on Architectural Issues: Scalability, Dependability, Adaptability, CoreGRID Technical Report, TR-0099, September 17, 2007.

CoreGRID White Papers:

- Marcin Adamski, Alvaro Arenas, Angelos Bilas, Paraskevi Fragopoulou, Vasil Georgiev, Alejandro Hevia, Gracjan Jankowski, Brian Matthews, Norbert Meyer, Jorg Platte, Michael Wilson, **Trust and Security in Grids: A State of the Art**, CoreGRID White Paper – WHP-0001
- A. Andrzejak, A. Reinefeld, F. Schintke1, T. Schütt, C. Mastroianni, P. Fragopoulou, D. Kondo, P. Malecot, G.-C. Silaghi, L.M. Silva, P. Trunfio, D. Zeinalipour-Yazti, E. Zimeo. **Grid Architectural Issues: State-of-the-art and Future Trends**. CoreGRID White paper – WP-0004

4) Grid Information, Resource and Workflow Monitoring Services

a. Objectives

The idea of the Grid Information, Resource and Workflow Monitoring Services architecture is to convey all the data through the information service in order to have a standard interface across different administrative sites and services. The characterisation of the required information service greatly depends on such factors as the demand placed on the source of information (e.g. static versus dynamic, publication rate), its purpose (e.g. discovery, logging, monitoring) and QoS requirements.

The primary objective of the IRWM Services research group is to study and provide general information and services for the underlying Grid management required by the Next Generation Grid. The Grid management services considered here include Grid core services and components. During the project the partners of the IRWM defined several specific objectives which allow us to work out the major goals, among others: providing multi-grain and dynamic monitoring for Grid resources and services, enabling reliable online monitoring of status and performance for a wide range of resources, support for extraction and representation of job workflows from programming models, framework for user management and user and job separation, supporting kernel and application level checkpointing. The Institute is focused on the following major objectives:

- Providing network infrastructure monitoring
- Providing monitoring of the progress of complex job workflows
- Support for extraction and representation of job workflows for programming models
- Realising middleware support for complex job workflow execution
- Framework for user management and user and job separation
- Supporting accounting services in virtual environments
- Providing checkpoint restart functionality in heterogeneous environment supporting dynamic job migration
- Supporting kernel and application level checkpointing.

b. Achievements

The IRWM Institute involves 13 CoreGRID partners (FHG, FORTH, INFN, INRIA, PSNC, SZTAKI, UMUE, UNICAL, UIBK, UMEA, UNI DO, UoS, UOW) including 44 researchers and PhD students. The work is mostly organised into research groups, units of two or more CoreGRID partners collaborating closely together on common goals. Joint technical reports, publications and prototype implementations are the major measurable outcomes. One of the

new researches is devoted to modelling dynamic workflow structures in higher order chemical languages.

One of the main achievements of the Institute in the area of resource monitoring is no doubt an integrated and unifying framework architecture for Grid monitoring. This framework integrates several components such as a workflow analyzer, a checkpointing manager, a Grid information system, a networking monitoring service and an account and user management system. The main scientific contribution is the layout of the internal structure of its components as well as their mutual interactions.

Another important contribution is the definition of the Grid Checkpointing Architecture (GCA) that allows the Grid middleware to use a variety of existing and future low-level checkpointing mechanisms in a conscious way. The current GCA proposal provides cooperation features with Virtual Machine Monitor technology that allow for suspending and later resuming the whole computing environment (together with the OS).

Research activities related to workflow have produced many novel ideas and, as a whole, is one of the success stories associated with the Institute. Workflow management in Grid computing is emerging as a programming paradigm for complex scientific applications. The institute has made progress on the modelling of workflow, the expression of workflow using either High-Level Petri-Net or unconventional paradigms such as the chemical metaphor. The institute has also proposed future research directions introducing fault tolerance techniques for Grid workflow systems. Since there are many workflow systems, with their own languages and low-level representations, the institute has proposed a technique to design translators able to convert legacy workflows.

The Institute partners have produced 23 CoreGRID Technical Reports containing their joint research activities and presented their work in workshops organized by the Institute.

Overall, the institute has contributed to solving the following research problems:

- User Management for Virtual Organizations
- Scalable multilevel checkpointing for distributed applications - on the integration possibility of TCKPT and psncLibCkpt
- A Grid Workflow Language Using High-Level Petri Nets
- Architecture of a Network Monitoring Element
- Scalable multilevel checkpointing for distributed applications - on the possibility of integrating Total Checkpoint and AltixC/R
- Grid Checkpointing Architecture - a revised proposal
- Virtual Environments - Framework for Virtualized Resource Access in the Grid
- Grid Checkpointing Architecture - Integration of low-level checkpointing capabilites with GRID
- Network Monitoring Session Description
- Grid Infrastructure Architecture: a modular approach from CoreGRID
- Fault-tolerant behavior in state-of-the-art Grid Worklow Management Systems
- Highly Dynamic Workflow Orchestration for Scientific Applications
- Benchmarking Grid Applications
- Taxonomies of the Multi-criteria Grid Workflow Scheduling Problem
- End-to-end Network Monitoring Infrastructure

- On the Characteristics of Grid Workflows
- A Grid Environment for Real-Time Multiplayer Online Games
- Enhancing Grids for Massively Multiplayer Online Computer Games
- Authorizing Grid Resource Access and Consumption

c. Future roadmap and vision

Besides the integration of research already done by individual partners within the IRWM, these services (information and network monitoring, checkpointing, workflow, accounting and user management) are essential for production Grid environments and all the current activities in Europe which is focusing on sustainable distributed computing environments as they provide: data for evaluation of the efficiency of systems and tools resulting from their research and support core functionality necessary for production Grid environments.
We will focus on proving the correctness of concepts worked out in each research group and disseminate the collected knowledge outside the CoreGRID consortium. This will result in further deployments of the integration work in national and international projects. One of the example is checkpointing, which was included in the concept of the National Grid Initiative (NGI) in Poland – PL-GRID and will be implemented in 2009. But we can find also further examples, like Austrian Grid or the D-GRID in Germany.
There are also plans of activities after the project will finish to keep alive the research activities worked out, e.g. in workflow services area (workflow forum defined by FhG, http://www.gridworkflow.org), Grid Checkpointing Architecture (PSNC web site is available, partners of the research group are interested to continue, http://checkpointing.psnc.pl).

The whole group expressed its interest to continue the work by having further regular meetings in 2009 and participating in common workshops. The IRWM proposed two research topics for ERCIM working group on following subjects:

- **Network Monitoring** (http://network-monitoring-rp.di.unipi.it/)
  The Network Monitoring research group, whose prominent participants were INFN and FORTH, reached the commitment of the group, a fully functional prototype of the monitoring infrastructure we had in mind, associating extreme scalability, security and flexibility. These results should be used as a basis for a more advanced investigation, aimed at deploying the infrastructure in a real scale environment, or re-engineering parts of the existing prototype to match new requirements and new environments, or starting from its limits to design a better one. The network monitoring infrastructures have a relevant role in industrial applications (telecom and grids inclusive), with a number of open issues that deserve our interest as researchers.
- **Virtual Environments – towards new generation e-Infrastructure**. The research topics are covering several tasks, including deployment of virtual environments on demand, workflow of virtual machines and environments, reliability and trust, accounting in virtual environments and real-time and scalability aspects in multi-user virtual environments.

A preliminary interest to co-operate in the mentioned subjects were raised by PSNC, University of Dortmund, University of Muenster, FORTH, UIBK, UMEA and UoS.

d. List of selected publications

- Ciuffoletti, Augusto and Polychronakis, Michalis, *Architecture of a Network Monitoring Element*, CoreGRID workshop at EURO-Par 2006, August 2006, Dresden (Germany), Springer
- Augusto Ciuffoletti, Antonio Congiusta, Gracjan Jankowski, Michal Jankowski, Ondrej Krajicek and Norbert Meyer. *GRID INFRASTRUCTURE ARCHITECTURE: A Modular Approach from CoreGRID*. In 3rd International Conference on Web Information Systems and Technologies (WEBIST), Barcelona (Spain), March 2007
- J. Kovacs, R. Mikolajczak, R. Januszewski, G. Jankowski: "Application and middleware transparent checkpointing with TCKPT on Clustergrid", Proceedings of 6th Austrian-Hungarian Workshop on Distributed And Parallel Systems, DAPSYS 2006, Innsbruck, Austria, September 21-23, 2006, pp. 179-189.
- Andreas Hoheisel and Martin Alt: Petri Nets. In: Workflows for e-Science - Scientific Workflows for Grids, Ian J. Taylor, Dennis Gannon, Ewa Deelman, and Matthew S. Shields (Eds.), Springer, 2006
- E. Elmroth, M. Jankowski, and N. Meyer. Authorizing Grid Resource Access and Consumption., 3rd CoreGRID Workshop on Grid Middleware, Barcelona (Spain), June 2008, CoreGrid Series Volume No. 11, Springer-Verlag, 2008.
- Justin Ferris, Mike Surridge, E. Rowland Watkins, Thomas Fahringer, Radu Prodan, Frank Glinka, Sergei Gorlatch, Christoph Anthes, Alexis Arragon, Chris Rawlings, and Arton Lipaj, Edutain@Grid: A Business Grid Infrastructure for Real-Time On-Line Interactive Applications in Workshop on Grid Economics and Business Models, August 26, 2008, The 5th International Workshop on Grid Economics and Business Models.
- Augusto Ciuffoletti, Yari Marchetti, Antonis Papadogiannakis, Michalis Polychronakis, Prototype Implementation of a Demand-Driven Network Monitoring Architecture, CoreGRID Integration Workshop, Heraklion-Crete, Greece, April 2008. Springer Verlag.
- Gracjan Jankowski, Radosław Januszewski, Jozsef Kovacs, Grid Checkpointing Service – Integration of Low Level Checkpointing Packages with the Grid Environment, 3rd CoreGRID Workshop on Grid Middleware, June 5–6, 2008, Barcelona, Spain.
- Zsolt Nemeth, Christian Perez and Thierry Priol: Chemical coordination: an abstract enactment model for workflows. Submitted for Submitted for PARA08 Workshop on State-of-the-Art in Scientific and Parallel Computing, Trondheim, 2008
- Justin Ferris, Mike Surridge, E. Rowland Watkins, Thomas Fahringer, Radu Prodan, Frank Glinka, Sergei Gorlatch, Christoph Anthes, Alexis Arragon, Chris Rawlings, Arton Lipaj, Edutain@Grid: A Business Grid Infrastructure for Real-Time On-line Interactive Applications, 5th International Workshop on Grid Economics and Business Models, Las Palmas, Gran Canaria, Spain, August 26, 2008, Springer Verlag
- Kassian Plankensteiner, Radu Prodan, Thomas Fahringer, Attila Kertesz, Peter Kacsuk, Fault-tolerant behavior in state-of-the-art Grid Workflow Management systems, CoreGRID Integration Workshop, Hersonisson, Crete, Greece, April 2-4, 2008, Crete University Press
- G. Kecskemeti, G. Terstyanszky, P. Kacsuk, T. Kiss, T. Delaitre: Automatic Service Deployment Using Virtualisation 16th EuroMicro Conference on Parallel, Distributed and Network-Based Processing, Toulouse, France, February 13-15, 2008
- Y. Zetuny, G. Terstyanszky, S. Winter, P. Kacsuk: Reputation-Policy Trust Model for Grid Resource Selection DAPSYS 2008, 7th Internal Conference on Distributed and Parallel Systems, Debrecen, Hungary, September 3-5, 2008

CoreGRID White Papers:

- Marcin Adamski, Alvaro Arenas, Angelos Bilas, Paraskevi Fragopoulou, Vasil Georgiev, Alejandro Hevia, Gracjan Jankowski, Brian Matthews, Norbert Meyer, Jorg Platte, Michael Wilson, **Trust and Security in Grids: A State of the Art**, CoreGRID White Paper – WHP-0001.
- Gracjan Jankowski, Radoslaw Januszewski, Rafal Mikołajczak, Jozsef Kovacs, **The Grid Checkpointing Architecture**, CoreGRID White Paper - WHP-0003.
- Stefan Freitag, Ramin Yahyapour, Gracjan Jankowski, Radoslaw Januszewski, **Virtualization Management for Grids and SOA**, CoreGRID White Paper, WHP-0005.

5) Resource Management and Scheduling

a. Objectives

The Institute works on several research aspects to provide complete solutions in this area. This includes the architectural perspective on how a suitable Grid scheduling architecture should be structured and implemented, but also on the algorithmic side towards efficient scheduling algorithms for different application use cases. These main objectives are achieved in several research tasks that focus on specific aspects. In order to provide a common and generic solution for Grid scheduling and management in Grids and Service-oriented Architectures (SOA), the Institute addresses the following research tasks:

- Definition of the components of a Grid scheduling architecture and their interaction
- Multi-level scheduling strategies with interaction between local resource management systems and higher-level Grid scheduling
- Workflow Grid scheduling strategies for jobs with temporal dependencies between different resource requirements
- Evaluation and benchmarking of Grid scheduling systems
- Model for mapping and scheduling of high performance parallel applications
- Coordinating Grid scheduling with data management
- Performance prediction for improving advance scheduling of resource allocations
- Service Level Agreements
- Virtualisation Management

b. Achievements

The Resource Management and Scheduling Institute brings together 21 full partners and 3 associate partners with a total of 97 researchers and PhD students.

One of the main achievements of the Institute has been its capability to build a community of users and developers in the area of Scheduling. Based on the definition of a set of Grid Scheduling Use Cases, a joint activity with the OGF, the institute identified a set of common requirements that helped to define a generic architecture for Grid Scheduling. The institute has also contributed to the study of several novel strategies for scheduling jobs on Grid resources, taking into account their heterogeneity or extracting information from workflows.

The Institute made progress in the definition of a Grid meta-broker providing a layer of abstraction between the users and various non-interoperable existing resource brokers.

The Institute has carried out research activities to define efficient scheduling algorithms able to cope with dynamic workloads, parallel applications, advance reservation and data dependencies.

The Institute also made progress in understanding the Service Level Agreement issues by surveying what has been done in several other EU-funded projects such as Akogrimo, AssessGRID, BEinGRID, BREIN, NextGRID and TrustCOM. Based on these findings, the institute investigated several models for the negotiation of SLAs

All these research activities generated an impressive number of publications in the best journals, conferences and workshop. 51 CoreGRID Technical Reports were published over the last four years. Moreover, the institute participated in the organization of scientific events within and outside CoreGRID. It was also involved in several demonstrations such as at IST'2006 or the CoreGRID Industrial Conference. Last but not least, it is strongly involved in several working or research groups within the Open Grid Forum.

Overall, the institute has contributed to solving the following research problems:

- Resource Management for Future Generation Grids
- Characterization of the performance of ASSIST programs
- Bringing Knowledge to Middleware – Grid Scheduling Ontology
- A Meta-Scheduling Service for Co-allocating Arbitrary Types of Resources
- A Proposal for a Generic Grid Scheduling Architecture
- Automatic mapping of ASSIST applications using process algebra
- Grid Environments Traces: Analysis and Implications
- Integration of ISS into the VIOLA Meta-Scheduling Environment
- Semantic Support for Meta-Scheduling in Grids
- User-Transparent Scheduling of Structured Parallel Applications in Grid Environments
- On Grid Performance Evaluation using Synthetic Workloads
- Reusable Cost-based Scheduling of Grid Workflows
- Operating on Higher-Order Components
- Load Balancing: Toward the Infinite Network
- Reliable Orchestration of Resources using WS-Agreement
- Co-Allocation of Compute and Network Resources in the VIOLA Testbed
- Reliable Orchestration of distributed MPI-Applications in a
- UNICORE-based Grid with MetaMPICH and MetaScheduling
- Improving Workflow Execution through SLA-based Advance Reservation
- Modeling and Supporting Grid Scheduling
- Simulating Grid Schedulers with Deadlines and Co-Allocation
- Towards ServMark, an Architecture for Testing Grids
- Application-oriented scheduling for HPC Grids
- BPDL: A Data Model for Grid Resource Broker Capabilities
- Advanced Techniques for Scheduling, Reservation and Access Management for Remote Laboratories and Instruments
- Using the eNANOS Low-Level Support in the GRMS Framework
- Co-allocation of MPI Jobs with the VIOLA Grid MetaScheduling Framework

- Dynamic SLA-negotiation based on WS-Agreement
- User-Transparent Scheduling for Software Components on the Grid
- Scheduling Malleable Applications in Multicluster Systems
- QoS-constrained List Scheduling Heuristics for Parallel Applications on Grids
- Using SLA for resource management and scheduling - a survey
- Attributes and VOs: Extending the UNICORE authorisation capabilities
- Benchmarking Grid Applications
- Taxonomies of the Multi-criteria Grid Workflow Scheduling Problem
- Multi-Cluster Text Mining on the Grid using the D-Grid UNICORE environment
- IANOS: An Intelligent Application Oriented Scheduling
- Middleware for a HPC Grid
- Meta-Brokering requirements and research directions in state-of-the-art Grid Resource Management
- Comparative Evaluation of the Robustness of DAG Scheduling Heuristics
- Novel Approaches for Scheduling in D-Grid – Towards an interoperable Scheduling Framework
- Towards a standards-based Grid Scheduling Architecture
- Optimization of Application Execution in the GridSpace Environment
- A Framework for Resource Availability Characterization and On-Line Prediction in Large Scale Computational Grids
- A Comparison of SLA Use in Six of the European Commissions FP6 Projects
- A Grid Environment for Real-Time Multiplayer Online Games
- Enhancing Grids for Massively Multiplayer Online Computer Games
- Towards SLA Based Software License Management in Grid Computing
- IANOS: Intelligent Application Oriented Scheduling for HPC Grids

C. Future roadmap and vision

There are still major challenges in the general adoption of management and scheduling features in production environments. The providers of infrastructures and software solutions are typically focused on proprietary solutions. However, there is trend to move to an open and agile environment which supports dynamic adaption to actual demand. Therefore, there is an existing gap to open and interoperable management solutions for diverse application environment. The Institute fosters this effort in its strong involvement in standardization activities like e.g. the Open Grid Forum, which links commercial and academic stakeholders. Since the start of the CoreGRID network the partners of the Institute for Resource Management and Scheduling have entered major collaborations in different contexts. This includes transferring and applying results to different application environments in several projects. This includes different technological scenarios like virtualization and cloud computing. At the same time, different application scenarios like e.g. multimedia, gaming, financial sector, telecommunication, and managing licenses in IPR have been considered. It is planned to continue such activities to disseminate project results beyond the scope of the network. It is envisioned to continue these collaborations within the Institute as a think tank to foster the research exchange. The regular meetings and joint discussions have led to a natural exchange of ideas and work on common topics of mutual benefit. The partners are eager and highly committed to continue these collaborations beyond the current funding period of the network. There will be joint efforts to keep this existing momentum of the Institute in order to master the continuous thread of new research fragmentation in this field.

d. List of selected publications

- Pugliese, D. Talia, R. Yayhapour, Modeling and Supporting Grid Scheduling. Journal of Grid Computing, Springer, 6(2), pp. 195-213, 2008.
- R. Baraglia, R. Ferrini, N. Tonellotto, L. Ricci, R. Yahyapour., A Launch-time Scheduling Heuristics for Parallel Applications on Wide Area Grids, Journal of Grid Computing, DOI 10.1007/s10723-006-9061-52007. Springer Verlag, 2007.
- L. Dumitrescu, D. H. Epema, J. Dünnweber, and S. Gorlatch. Reusable costbased scheduling of grid workflows operating on higher-order components. In 2nd IEEE International Conference on e-Science and Grid Computing (e-Science 2006). IEEE Computer Society Press, December 2006
- M. Parkin, R. M. Badia, and J. Martrat.. A Comparison of SLA Use in Six of the European Commissions FP6 Projects. CoreGRID TR-0129, Apr 14, 2008.
- Klusáček, L. Matyska, H. Rudová, R. Baraglia, and G. Capannini, Local Search for Grid Scheduling. In the Doctoral Consortium at the International Conference on Automated Planning and Scheduling (ICAPS 2007), Providence, RI, USA, 2007.
- Iosup, D.H.J. Epema, T. Tannenbaum, M. Farallee, and M. Livny, "Inter-Operating Grids through Delegated Matchmaking," *Supercomputing 2007 (SC2007)*, November 2007.
- Nae, A. Iosup, S. Podlipnig, R. Prodan, D.H.J. Epema, and T. Fahringer, "Efficient Management of Data Center Resources for Massively Multiplayer Online Games" *Supercomputing 2008*, November 2008.
- Pichot; P. Wieder; O. Wäldrich; W. Ziegler: Dynamic SLA Negotiation based on WS-Agreement. Proceedings of the 4th International conference on Web Information Systems and Technologies (WEBIST 2008), Mai 2008, Funchal, Portugal, ISBN: 978-989-8111-29-6.
- P. Hasselmeyer, H. Mersch, H. Quyen, B. Koller, L. Schubert, P. Wieder: Implementing an SLA Negotiation Framework Proceedings of the eChallenges Conference (e-2007), 2007.
- H. Rasheed, R. Gruber, V. Keller, O. Wäldrich, W. Ziegler, P. Wieder, P. Kuonen, M-C. Sawley, S. Maffioletti, P. Kunszt: IANOS - An Intelligent Application Oriented Scheduling Middleware for a HPC Grid ,TR-0110, Dec 28, 2007
- Kertész, P. Kacsuk, I. Rodero, F. Guim, J. Corbalan: Meta-Brokering requirements and research directions in state-of-the-art Grid Resource Management, TR-0116, Nov 8, 2007.
- Grimme, J. Lepping, A. Papaspyrou, P. Wieder, R. Yahyapour, A. Oleksiak, O. Wäldrich, W. Ziegler: Towards a standards-based Grid Scheduling Architecture, TR-0123, Dec 31, 2007

CoreGRID White Papers:

- Marcin Adamski, Alvaro Arenas, Angelos Bilas, Paraskevi Fragopoulou, Vasil Georgiev, Alejandro Hevia, Gracjan Jankowski, Brian Matthews, Norbert Meyer, Jorg Platte, Michael Wilson**, Trust and Security in Grids: A State of the Art**, CoreGRID White Paper – WHP-0001

- Stefan Freitag, Ramin Yahyapour, Gracjan Jankowski, Radoslaw Januszewski. **Virtualization Management for Grids and SOA**. CoreGRID White Paper – WHP-0005.

6) Grid Systems, Tools, and Environments

a. Objectives

The research activities of the STE Institute are organised in four tasks with the following objectives:

- **Generic Platform**: This task addresses the development of a generic, component-based platform, with focus on its architecture, interoperability, security, and design methodology. This platform is specifying the basic "glue" from which higher-level components and services can be constructed. Challenges arise from heterogeneous software libraries, toolkits, operating systems and existing Grid frameworks. The specification of this platform underpins the research areas of the other three tasks.
- **Mediator Components**: This task is responsible for designing a component suite that mediates between the applications and system layers, providing core services according to the architecture of the generic platform. These include such capabilities as dynamic application steering, meta-data retrieval, and service discovery. The components delivering these capabilities are integrated and exposed via a runtime environment that acts as a gateway to system components, thus addressing the integration aspects of the generic component platform.
- **Integrated Toolkit**: This task is specifying and designing a toolkit for simplifying the deployment of Grid-unaware applications while also optimising application performance. More specifically, this involves defining and mapping application requirements to the component-based generic platform as well as providing application interfaces to the mediator components. The runtime environment of such an integrated toolkit is able to run applications in a Grid and optimise their performance dynamically in a way transparent for the user.
- **Advanced Tools and Environments for Problem Solving**: This task involves designing PSEs or portals that take legacy software and automatically deploy it as a service that conforms to a standard service model. The aim is to integrate our generic platform into PSEs and portals allowing users to compose, steer, monitor and visualise job execution in a transparent and simple way. This work is based on identified deployment scenarios, using technologies for wrapping legacy code and mechanisms for deploying and managing services and jobs in peer-to-peer environments.

b. Achievements

The STE Institute brings together twelve partners – IPP-BAS, USTUTT, ICS-FORTH, IC, INRIA, SZTAKI, UWC, UNIPI, UOW, UPC, VUA, and CYFRONET – from ten different European countries. In addition two new associated partners – CATNETS and UNIVIE – joined the Institute. The STE Institute involves more than 80 senior researchers and PhD students who participate in eleven research groups for implementing the Institute's roadmap.

The main contribution of the STE Institute is the definition of a generic platform. This research task aimed at developing the common aspects of a generic, component-based platform. This platform specifies the basic "glue" from which higher-level components and services can be constructed. The work focused on the challenges arising from the need to support heterogeneous software libraries, toolkits, operating systems and existing Grid frameworks in its approach. The second contribution is the definition of mediator

components, a suite of components that mediate between applications and system components, providing core services according to the component framework. These include such capabilities as application steering, meta-data retrieval and service discovery. The third contribution is the development of an integrated toolkit to simplify the deployment of Grid-unaware applications while optimizing the performance of the application. The approach follows a component-based strategy. The fourth and last contribution was the specification of advanced tools and environments for problem solving. It allows component-based PSEs and portals to compose, expose, and monitor legacy and non-legacy applications.

The Institute produced 28 CoreGRID Technical Reports during the four years of the CoreGRID NoE project.

Overall, the institute has contributed to solving the following research problems:

- Grid Application Programming Environments
- GEMLCA: Running Legacy Code Applications as Grid Services
- Legacy Code Support for Production Grids
- Lightweight Grid Platform: Design Methodology
- Information Sources and Sinks in a Grid Environment
- Security Models for Lightweight Grid Architectures
- Mapping “Heavy” Scientific Applications on a Lightweight Grid Infrastructure
- Componentising a Scientific Application for the Grid
- A Software Component-based Description of the SEGL Runtime Architecture
- A Super-Peer Model for Multiple Job Submission on a Grid
- Domain-Specific Metadata for Model Validation and Performance Optimisation
- Reliability and Trust Based Workflows’ Job Mapping on the Grid
- Grid Superscalar and GriCoL: Integrating Different Programming Approaches
- Highly Dynamic Workflow Orchestration for Scientific Applications
- Problem Solving Environment for Distributed Interactive Applications
- Performance Monitoring of GRID superscalar with OCM-G/G-PM: integration issues
- Use of P2P Overlays for Distributed Data Caching in Public Scientific Computing
- Towards a scientific workflow-oriented computational World Wide Grid
- Dynamic Service-Based Integration of Mobile Clusters in Grids
- HLA Component Based Environment for Distributed Multiscale Simulations
- Measurable Component Values and Services: the Economic Planning of Resource Transactions
- Autonomic Behaviour of Grid Applications using Component Platforms
- Generic Component Platform Architecture and Design Methodology
- Dynamic Service Aggregation – Virtual Clusters and WSPeer
- Security Issues in Component-Based Grid Systems
- Use of GCM for Grid System Software Design and Development
- Interoperability at Component Level and at Workflow Level
- Grid Integrated Development Environment – GIDE
- Mediator Component Framework
- Software Licensing Issues for Complex Distributed Applications
- Automation of Application Deployment on Grids
- Methodology for Component Application Development - Use Cases

c. Future roadmap and vision

The current research profile of the STE Institute includes strong expertise and recent research results in component-based grid infrastructures and platforms. This includes support for the management of a large number of services and the ability to dynamically adapt these services if needed through runtime steering techniques. We have been investigating the design methodology for component-based services in a dynamically reconfigurable distributed platform supporting flexible and fault-tolerant composition and execution of workflows. Our approach, objectives, methodology, and existing tools and environments contribute directly to Objective ICT-2007.1.2: Service and Software Architectures, Infrastructures and Engineering. Also, the increased activity in the Institute related to services and service oriented architectures, will address several problems and objectives recently stated in EU research guidelines such as the NESSI Technology Platform.

Recent results, publications, and deliverables by the STE Institute represent the culmination of the research efforts of all partners during the 4-year period of the CoreGRID NoE grant with particular emphasis on the integration of each approach into a unified generic Grid architecture. Most of our recent joint research is still in progress and will require some more time for completion. One of the major outcomes of this research, however, is that it helps to highlight the future challenges and open issues in a number of topics central to the design methodology of our generic Grid services platform. One important aspect of these efforts is to enable the maximum adaptation of the Grid platform and high level of automation in the design, development and execution of complex application. Automation is also vital in addressing scalability and robustness and is a key area of future investigation.

As previously stated, the lack of longer-term experience, because of the very rapid developments in the field, and the complexity of the target systems demand more research and results. An immediate priority in our future research efforts is to give higher attention to the service-oriented approach towards the design and development of component-based services computing platform. Important open topics are relating the existing and recent work of the STE institute to further developing our “Invisible Grid” concepts and actively applying them into complex services computing systems and the upcoming area of cloud computing in particular. Future opportunities stem also from the new concepts of Infrastructure as a Service (IaaS) and Software as a Service (SaaS).

d. List of selected publications

- N. Parlavantzas, V. Getov, M. Morel, F. Baude, and D. Caromel, *Design Support for Componentising and Grid-enabling Scientific Applications*, Proc. ACM CBHPC Symposium, pp. 31-38, ACM Press, 2007.
- E. Tejedor, R.M. Badia, T. Kielmann, and V. Getov, *A Component-based Integrated Toolkit*, In: Making Grids Work, pp. 139-151, Springer, 2008.
- S. Isaiadis, V. Getov, I. Kelley, I. Taylor, *Dynamic Service-based Integration of Mobile Clusters in Grids*, In: Grid Computing: Achievements and Prospects, pp. 159-171, Springer, 2008.
- M. Aldinucci, M. Danelutto, H.L. Bouziane, and C. Pérez, *Towards Software Component Assembly Language Enhanced with Workflows and Skeletons*, Proc. ACM CBHPC Symposium, pp. 1-11, ACM Press, 2008.

- R.M. Badia, R. Sirvent, M. Bubak, W. Funika, and P. Machner, *Performance Monitoring of Grid Superscalar with OCM-G/G-PM: Integration Issues*, In: Achievements in European Research on Grid Systems, pp. 193-205, Springer, 2008.
- P. Kacsuk and T. Kiss, *Towards a Scientific Workflow-oriented Computational World Wide Grid*, CoreGRID TR 0115, Institute on Grid Systems, Tools and Environments, CoreGRID NoE, December 2007.
- A.-M. Oprescu, T. Kielmann, M. Danelutto, and M. Aldinucci, *Autonomic Behaviour of Grid Applications using Component Platforms*, CoreGRID TR-0156, Institute on Grid Systems, Tools and Environments, CoreGRID NoE, July 2008.
- T. Kiss, P. Kacsuk, G. Terstyanszky, S. Winter, *Workflow Level Interoperation of Grid Data Resources*, Proc. IEEE CCGrid, pp. 194-201, IEEE CS Press, 2008.
- M. Malawski, T. Gubala, M. Kasztelnik, T. Bartynski, M. Bubak, F. Baude, and L. Henrio, *High-level Scripting Approach for Building Component-based Applications on the Grid*, In: Making Grids Work, pp. 309-321, Springer, 2008.
- M. Ejdys, U. Herman-Izycka, N. Lal, T. Kielmann, E. Tejedor, and R. Badia, Integrating Application and System Components with GCM, In: From Grids to Service and Pervasive Computing, pp. 47-59, Springer, 2008.
- T. Kielmann, G. Wrzesinska, N. Currle-Linde, and M. Resch, *Redesigning the SEGL Problem Solving Environment: A Case Study of Using Mediator Components*, In: Integrated Research in Grid Computing, pp. 255-269, Springer, 2007.
- J. Thiyagalingam, V. Getov, S. Panagiotidi, O. Beckmann, J. Darlington, *Domain-specific Metadata for Model Validation and Performance Optimisation*, In: Achievements in European Research on Grid Systems, pp. 165-178, Springer, 2007.
- M. Malawski, M. Bubak, F. Baude, D. Caromel, L. Henrio, and M. Morel, *Interoperability of Grid Component Models: GCM and CCA Case Study*, In: Towards Next Generation Grids, pp. 95-106, Springer 2007.
- H.L. Bouziane, C. Pérez, N. Currle-Linde, and M. Resch, *Analysis of Component Model Extensions to Support the GriCoL Language*, In: Making Grids Work, pp. 45-55, Springer, 2008.
- A. Anciaux-Sedrakian, R.M. Badia, J.M. Pérez, R. Sirvent, T. Kielmann, A. Merzky, *Reliability and Trust Based Workflows' Job Mapping on the Grid*, CoreGRID TR-0069, Institute on Grid Systems, Tools and Environments, CoreGRID NoE, Jan. 2007.

CoreGRID White Paper:

- Marcin Adamski, Alvaro Arenas, Angelos Bilas, Paraskevi Fragopoulou, Vasil Georgiev, Alejandro Hevia, Gracjan Jankowski, Brian Matthews, Norbert Meyer, Jorg Platte, Michael Wilson**, Trust and Security in Grids: A State of the Art**, CoreGRID White Paper – WHP-0001

VI. Integration Activities

1) Preparation of project proposals

With regards to success stories related to the integration activities, we have demonstrated our ability, as a single entity, to help identifying the main research challenges in Grid. Some of them were taken into account by the FP6 work programme in Grid technologies. CoreGRID has also played a role in facilitating the emergence of proposals that were submitted to several

FP6 and FP7 calls. Some of the proposals were mainly driven by CoreGRID partners (XtreemOS, GridCOMP, Selfman, GridTrust, EchoGRID, Phosphorus, SmartLM) some others included partners from CoreGRID because of their involvement in the NoE (S-Cube, OGF-Europe). This activity was also a side effect of the level of integration we have reached within CoreGRID: it is easier to prepare a proposal when you have a clear picture of what partners are doing, and when you are already involved in a joint research activity that you would like to pursue with a more ambitious agenda. To illustrate this latter aspect, one can cite the GridCOMP project for which CoreGRID contributed to the design of the Grid Component Model whereas GridCOMP has allowed some of the CoreGRID researchers to propose a reference implementation of GCM through its use within applications from the Industry.

2) Annual Integration Workshop

The annual integration workshops have been also quite successful in providing the dissemination tool to publicize the best integrated research carried out within CoreGRID. The motivation for the integration workshops was to analyze overlapping areas and synergies existing between different work packages and between different research institutions within the same work package. The workshops focused on the interdisciplinary topics within the Network and identified and promoted durable activities that guaranteed a high level of collaboration between partners. These workshops also stimulated the competition among CoreGRID partners since not all submitted papers were accepted, with the exception of the first workshop. Figure 3 gives the list of workshops we organized during the last four years. We intended to organize one workshop per year, but we realised that October and November were not suitable since they are usually busy months. This is why we decided to shift the last one to April 2008. Moreover, since it was the last year of the Network, it was more convenient to organize it 6 months before the end of the contract, in order to have a more complete overview of the research work carried out by CoreGRID researchers.

	Location	Date	Number of attendees	Number of submitted papers	Percentage of positive impression
CoreGRID Integration Workshop 2005	Pisa (Italy)	28.11.2005 - 30.11.2005	124	46	85%
CoreGRID Summer School 2006	Cracow (Poland)	19.10.2006 - 20.10.2006	89	42	89%
CoreGRID Integration Workshop 2008	Hernonissos (Greece)	2.4.2008 - 4.4.2008	87	49	97%

Figure 3: List of Integration Workshops

3) Trust and Security

Another successful activity is related to the training of CoreGRID researchers on Trust and Security issues. We must say that we took the risk in the proposal not to have a dedicated Institute on Trust & Security. We were a bit anxious to see the reaction of the reviewers on this intricate choice. After four years running, we think we took the right decision. The risk of having a dedicated institute was to isolate researchers from those who participated in the other institutes. Thanks to its leader, Alvaro Arenas from STFC, the six CoreGRID institutes have made the necessary effort to identify Trust and Security issues related to the research topics they covered. This activity lead also to the making of a White Paper on Trust & Security that is available from the CoreGRID web site. Another indicator showing the effectiveness of this

horizontal activity is the number of CoreGRID technical reports that addressed Trust & Security issues. So far, 10 technical reports, from four CoreGRID institutes, have addressed trust or security issues in various aspects. Thanks to this horizontal activity, we have been successful in training and increasing the knowledge of CoreGRID researchers on these particular and important issues.

4) Testbed

The network does not have its own Grid testbed to allow researchers to carry out their experiments. Instead, several CoreGRID partners that have already deployed testbeds, agreed to give CoreGRID researchers access to these. DAS-3[2], Grid'5000[3] and Phosphorus[4] were used by the network and several successful experiments were carried out during the last four years that lead to several publications.

5) Durable Mobility of Researchers

The Network implemented a specific mobility programme, and allocated substantial funding to it, with a goal to encourage researchers, post-docs and PhD students to stay, for a given period, with research teams of Network participants. This mobility programme fosters better relationships between network researchers and increases the effectiveness and the level of integration of the Network. It is mainly based on two mobility schemes: fellowship for long stay (several months) and researcher exchange (REP) for shorter stays (several weeks). Figure 4 gives the number of fellowships and REPs allocated to the CoreGRID partners. During the four years, we provided funding for 45 REPs (Researcher Exchange Programme). A total of 33 CoreGRID partners among 48 benefited from this REP.

[2] http://www.cs.vu.nl/das3/
[3] https://www.grid5000.fr/mediawiki/index.php/Grid5000:Home
[4] http://www.ist-phosphorus.eu/

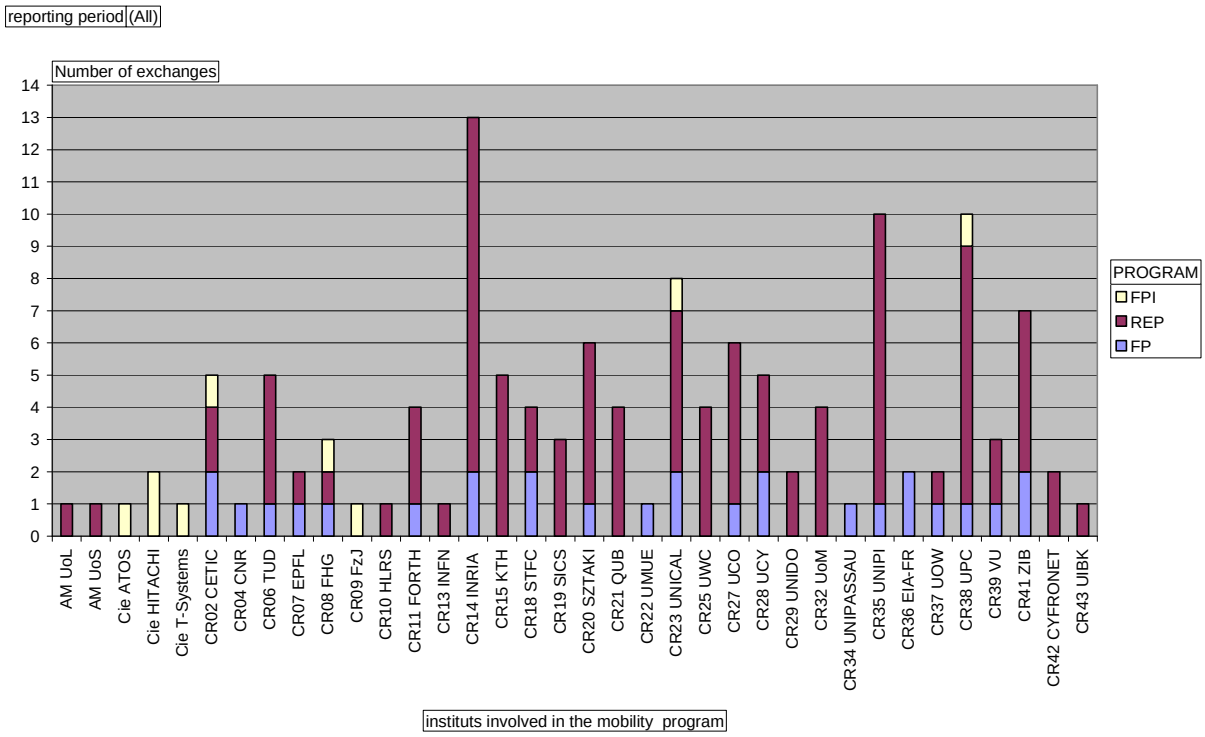


Figure 4: Number of REP per partner

Figure 5 shows the distribution of REPs per Institute (a) and per year (b). The latter figure is a clear indicator of the level of integration we reached at the end of the Network. Whereas in the first year and even in the second year we observed a low number of REPs, the last 2 years showed a significant increase of REPs. It took almost 2 years to the CoreGRID researchers to gain a good knowledge and understanding of what their colleagues were doing in terms of research and to adapt their research agenda to be able to start joint collaborations.

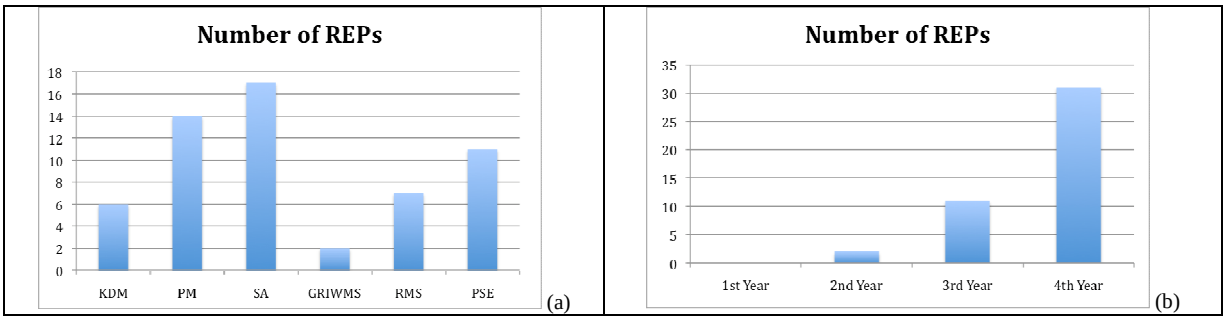


Figure 5: Number of REPs

Concerning the fellowship programme, we used the entire budget that was associated with this initiative. 19 fellows were funded, out of which 4 fellowships were attributed to foster the relationship with the industry. Figure 6-a shows that 21 partners benefited from the fellowship programs and 3 industrial partners were involved in it. All six institutes were granted between 1 and 8 fellowships (Figure 6-b). In general, one or two institutes can be involved in a single fellowship.

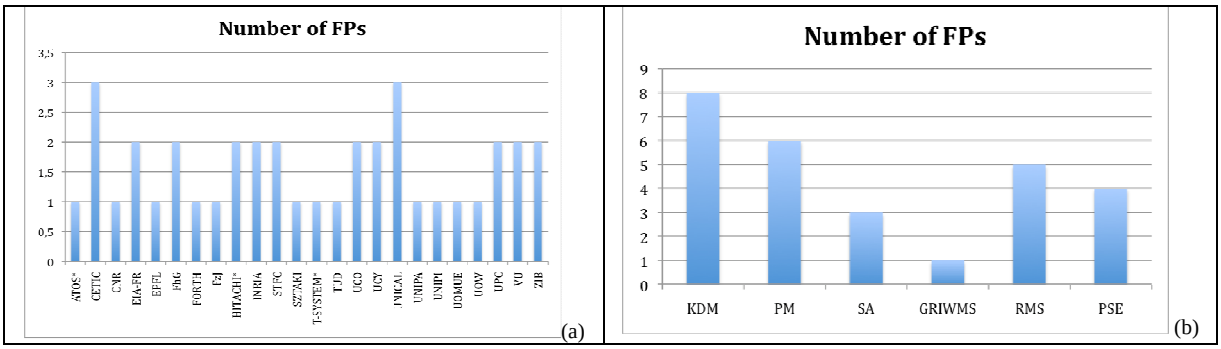


Figure 6: Number of FPs

6) Meetings and Workshops

To support its integration work, the network has organized approximately 60 meetings and 17 thematic workshops, providing forum for discussions between CoreGRID researchers.

7) Researchers database

CoreGRID has built a researchers database to help identifying the European experts in a given research topic related to Grids. It contains 270 researchers and searches can be performed using 45 key words from the Grid domain.

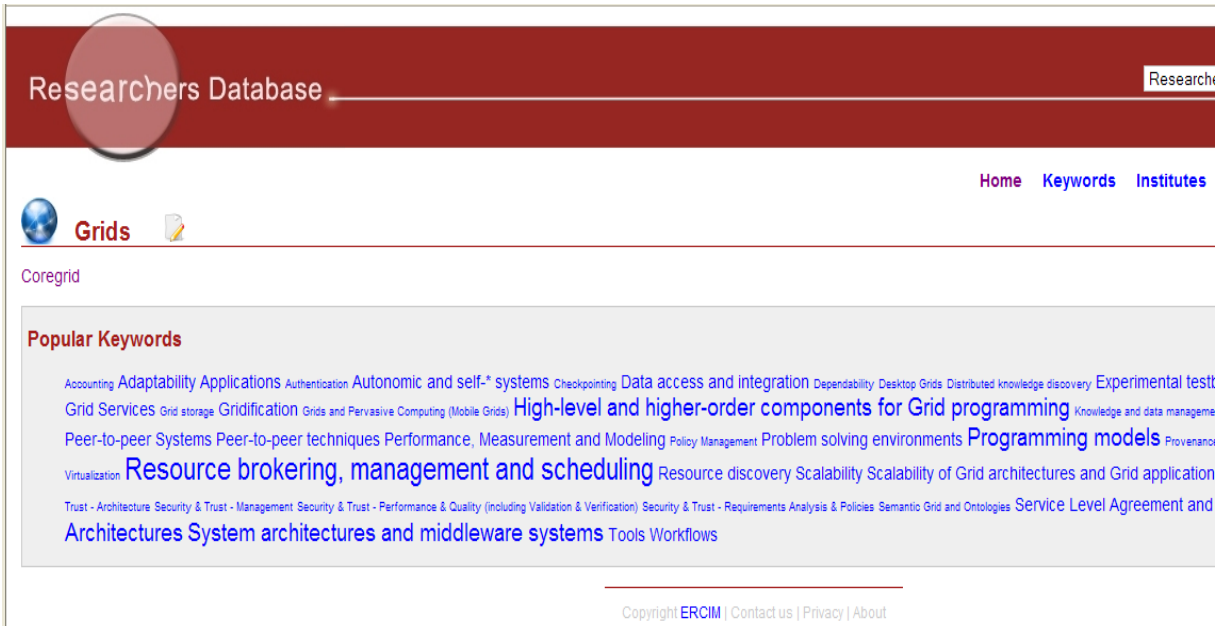


Figure 7: researchers database

8) Measuring the level of integration

The network implemented a database of all Grid related publications made by the network members. This database was made publicly available through the CoreGRID web site. This

database contained around 1050 entries in September 2008 and features among the most accessed web pages of the CoreGRID web site. It has also been an excellent tool to assess the level of integration within the network. Figure 9 shows the increase in the level of integration thought joint publications published from 2004 to 2008. It shows an increasing number of collaborations that led to an ever-higher number of publications during the four years. This left no isolated member in the network although some partners had more collaborations than others (the link colour shows the number of joint publications involving two given partners). This evaluation is a clear indicator of the network impact on integration.

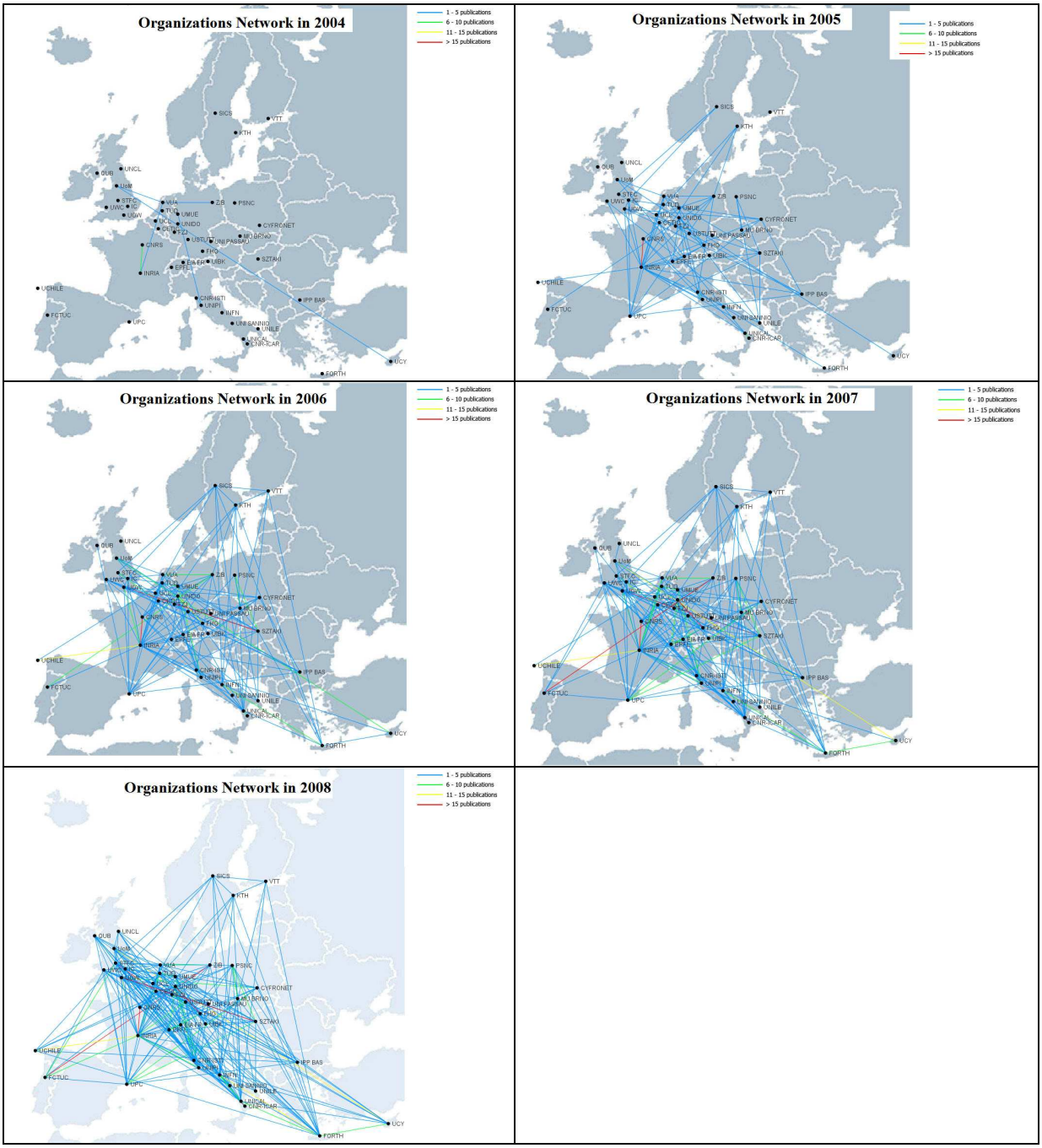
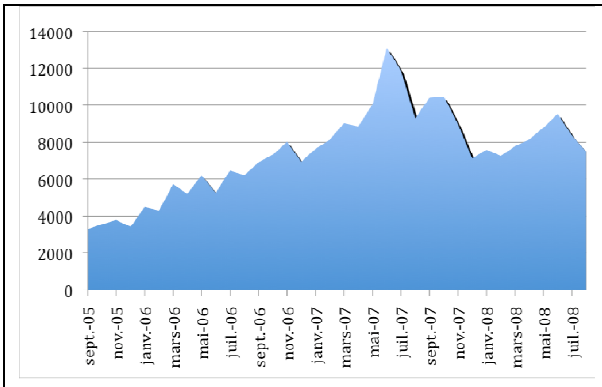


Figure 8: Joint publications by CoreGRID partners.

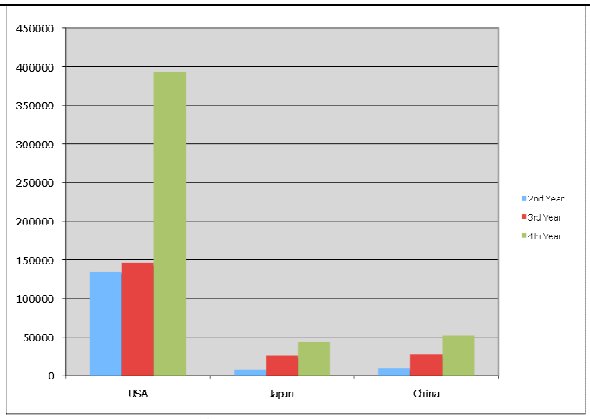
VII. Spreading Excellence

1) Public Website

The CoreGRID web site has been set up immediately after the launch of the Network with the objective to disseminate the activities of the consortium. In fact, it has done more than just showcasing our own activities since we received many requests from other researchers and institutions to advertise events they organized, new positions available in academia or the industry, etc. As shown in Figure 9, the web site audience (a) has increased rapidly during the first two years (we observed a decrease during the third year, probably due to a lesser hype about Grid computing, in line with the famous Gartner hype curve). In (b) we show an increasing number of visitors from the USA, Japan and China during the last three years.



(a) Number of unique visitors per month during the last three years



(b) Number of times a page is viewed by a visitor from a given country

Figure 9: CoreGRID web site access statistics.

CoreGRID became the main European portal in Grid research and filled one important gap: the lack of a Web portal showing European research activities in the Grid domain. This statement can be illustrated by the Google ranking as shown in Figure 10.

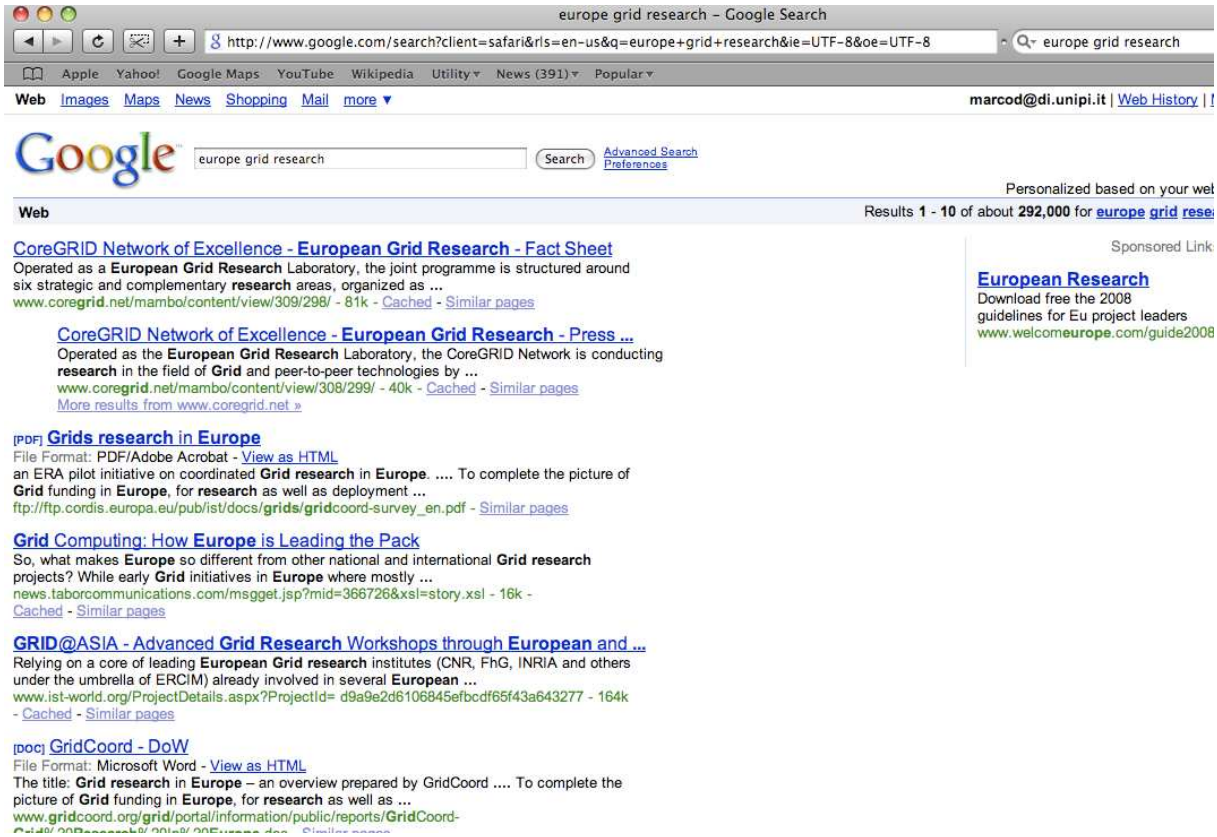


Figure 10: CoreGRID web site visibility through Google

We have had many discussions with our foreign colleagues involved in Grid computing. Most of them have difficulties to build an overall picture of what is going on in Grid computing. They are exposed to a large number of projects funded by different Units. Moreover, each member state has its own national programme on Grid. We think that a NoE, in a particular research domain, should also play the role of being the main portal to expose all the best research results in its domain of expertise. This is what we tried to do with our web site. However, we have probably underestimated the cost of playing this role since a lot of efforts had to be invested in constantly updating the web site.

The Google page rank for the CoreGRID web site is 8/10 and the site appears in the first position when issuing a query to Google to look for web addresses of sites using the following keywords: “European Grid Research” and “Excellence Grid Research”.

2) Publications

The network has set up four initiatives to disseminate its research results through publications:

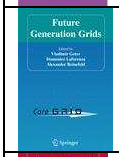
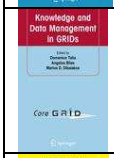
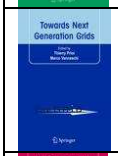
- A series of Technical Reports to disseminate the results
- A series of White Papers
- A series of CoreGRID Edited Book published by Springer
- A database of scientific papers published by the CoreGRID researchers

Except for the white papers, these activities were not mentioned in the contract when we started the network but were added after a couple of months when we gained a better understanding of how to foster integration.

In October 2004, we established the concept of Technical Reports to foster collaboration between researchers involved in the network. All CoreGRID partners were requested to publish their research results as CoreGRID Technical Reports whenever they were the result of joint research activities, prior to (or soon after) submission of the paper in a conference or in a journal. To foster integration, only reports with authors from at least two different CoreGRID partners are eligible for publication as CoreGRID Technical Reports. Although this activity was added lately, it has been quite successful since at the end of contract, the network has published 176 Technical Reports (10 the first year, 39 the second year, 54 the third year and 73 to-date in the final year). These technical reports are all available on the web site.

The publication of White Papers has however not been so successful since only 5 were published and mainly during the fourth year of the network. It has been quite difficult to convince researchers to write such reports showing the state of the art in a particular aspect of Grid research. So far, we did it only for Knowledge and Data Management, Checkpointing Architecture and Trust & Security.

As the network organized workshops and symposiums, we negotiated a series of CoreGRID Edited Books with Springer, in order to publish proceedings associated with these events. We have edited 11 books as shown in Figure 11.

Cover page	Title	Editors
	Component Models and Systems for Grid Applications Proceedings of the Workshop on Component Models and Systems for Grid Applications held June 26, 2004 in Saint Malo, France ISBN 978-0-387-23351-2	Vladimir Getov (University of Westminster) & Thilo Kielmann (Vrije Universiteit)
	Future Generation Grids Proceedings of the Workshop on Future Generation Grids, November 1-5, 2004, Dagstuhl, Germany ISBN 978-0-387-27935-0	Vladimir Getov (University of Westminster), Domenico Laforenza (CNR-ISTI) & Alexander Reinefeld (ZIB)
	Knowledge and Data Management in GRIDs ISBN 978-0-387-37830-5	Domenico Talia (University of Calabria), Angelos Bilas (FORTH) & Marios D. Dikaiakos (University of Cyprus)
	Integrated Research in Grid Computing ISBN: 978-0-387-47656-8	Sergei Gorlatch (University of Münster) & Marco Danelutto (University of Pisa)
	Towards Next Generation Grids ISBN 978-0-387-72497-3	Thierry Priol (INRIA) & Marco Vanneschi (University of Pisa)
	Achievements in European Research on Grid Systems ISBN: 978-0-387-72811-7	Sergei Gorlatch (University of Münster), Marian Bubak (CYFRONET) & Thierry Priol (INRIA)

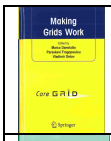
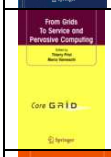
	Making Grids Work ISBN: 978-0-387-78447-2	Marco Danelutto (University of Pisa), Paraskevi Fragopoulou (Forth) & Vladimir Getov (University of Westminster)
	Grid Middleware and Services – Challenges and Solutions ISBN: 978-0-387-78445-8	Domenico Talia (University of Calabria), Ramin Yahyapour (University of Dortmund) & Wolfgang Ziegler (Fraunhofer)
	Grid Computing – Achievements and Prospects ISBN: 978-0-387-09456-4	Sergei Gorlatch (University of Münster), Paraskevi Fragopoulou (Forth) & Thierry Priol (INRIA)
	From Grids to Service and Pervasive Computing ISBN: 978-0-387-09454-0	Thierry Priol (INRIA) & Marco Vanneschi (University of Pisa)
	CoreGRID workshop on Grid Middleware	N. Meyer (PSNC), D. Talia (University of Calabria), and R. Yahyapour (University of Dortmund)

Figure 11: CoreGRID edited books.

As suggested by the Scientific Advisory Board in May 2005 to allow all members of a Network of Excellence to be aware of the research activities of other members, we implemented a database of all Grid related publications by the network members. This database was made publicly available through the CoreGRID web site. This database contained around 900 entries in August 2008 and features among the most accessed web pages of the CoreGRID web site. Additionally, CoreGRID researchers have published more than 450 joint scientific papers accepted in peer-reviewed conferences, workshops and journals. We should highlight two of them that received a best paper award associated with specific joint research work carried out within CoreGRID:

- Artur Andrzejak (Zuse Institute Berlin) and Luis Silva (University of Coimbra) received the Best Paper Award of the NOMS 2008 Conference for their paper "Using Machine Learning for Non-Intrusive Modeling and Prediction of Software Aging"
- Luis Moura Silva (University of Coimbra), Javier Alonso (Universitat Politècnica de Catalunya), Paulo Silva (University of Coimbra), Jordi Torres (Universitat Politècnica de Catalunya) and Artur Andrzejak (Zuse Institute Berlin) received a Best Application Paper Award for their paper "Using Virtualization to Improve Software Rejuvenation", during the 6th IEEE International Symposium on Network Computing and Applications (July 2007).

3) Visibility

Since our strategy was to appear as a single, yet geographically distributed laboratory, we decided in 2005 to publish an annual report like any other research laboratories. Over the duration of the contract, CoreGRID produced three annual reports (2005, 2006 and 2007) as shown in Figure 10.

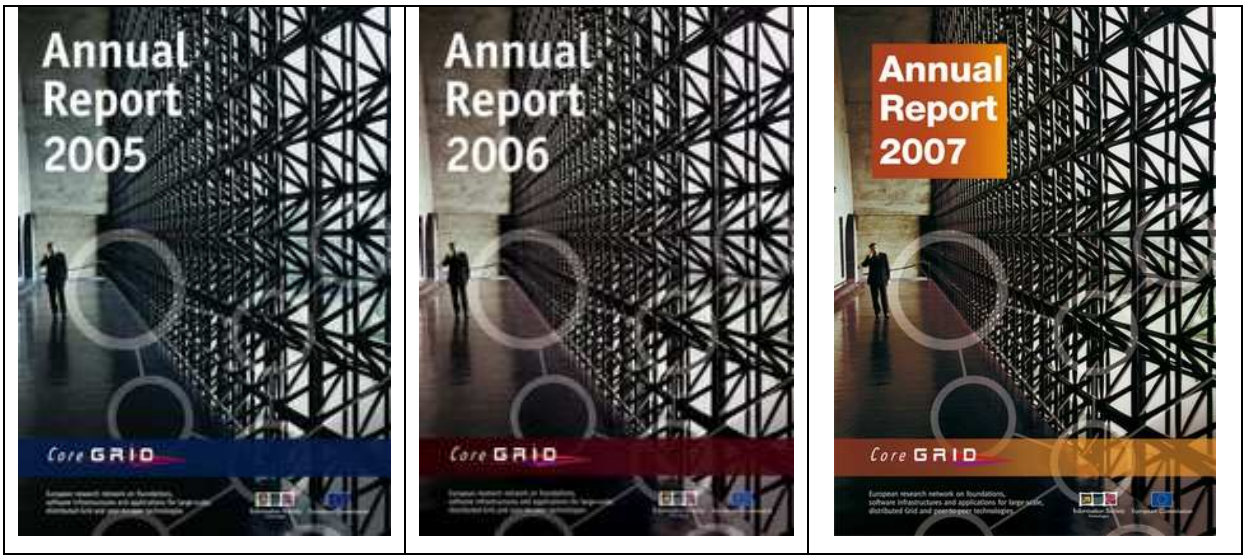


Figure 12: Series of Annual Reports

These reports follow high-quality editing standards with a particular effort on attractiveness and readability. They have been disseminated during various events organized either by the Network, by EU funded projects or at relevant meetings organized by the European Commission. They can also be downloaded from the CoreGRID web site. We think that these three reports helped us to establish CoreGRID as a Grid research laboratory worldwide. We also published a brochure that gives a short description of the network activities. Moreover, every executive member in the network was given a set of CoreGRID business cards to be used when they interacting with non-affiliated researchers.

To complement this effort towards an increased visibility, we established a CoreGRID sponsorship strategy. First of all, we have been Silver Organizational Member of the Open Grid Forum since 2006. We also gave the "CoreGRID label" to several scientific events such as high-quality conferences and workshops. Among them, we have established a closer collaboration with the Euro-Par organization that organizes a very well known conference on parallel and distributed computing in Europe. We were tempted to organize our own series of conferences but this would have been funded at the expenses of existing well-established European conferences covering the Grid research area. This is why we preferred to set up a strong link with Euro-Par. We have organized all our network annual meetings (Members General Assembly) jointly with Euro-Par since 2005. We also set up a symposium in 2007 and 2008 with the goal of being the premiere European event on Grid Computing for the dissemination of the results obtained by European and member states initiatives, as well as other international projects in Grid research and technologies. As part of the negotiation with Euro-Par, we gave them the CoreGRID label each year. For the other conferences or workshops, the decision to grant the CoreGRID label was taken each year by the Executive Committee. The following conferences were given the CoreGRID label: Grid@Works'05, HPC-GECO/CompFrame'05, Grid'06, HPDC'06, Grid'2007, PCGrid'08, CCGRID'08. The label was given on the basis of the reputation of the event and the participation of CoreGRID researchers to programme committees. Once the conference got the CoreGRID label, CoreGRID researchers were allowed to use their CoreGRID budget to attend these events.

4) Training and education activities

Training and education are important when running a network of excellence, especially when we have a large number of PhD students like in CoreGRID. The network has successfully organised four summer schools allowing participants to attend lectures made by well-known Grid researchers but also to practice some existing Grid middleware technologies.

	Location	Theme	Date	Number of attendees
CoreGRID Summer School 2005	Lausanne (Switzerland)	Introduction to GRID technologies	5.9.2005 - 9.9.2005	42
CoreGRID Summer School 2006	Bonn (Germany)	Current and Future Generation Grid Technology	24.7.2006 - 28.7.2006	55
CoreGRID Summer school 2007	Budapest (Hungary)	Current and Future Generation Grid Technology	3.9.2007 - 7.9.2007	40
CoreGRID Summer School 2008	Dortmund (Germany)	Linking Grids and SOA	7.7.2008 - 11.7.2008	31

Figure 13: list of CoreGRID summer schools

VIII. Network Coordination

1) Administrative & Financial Coordination

The main tasks of the CoreGRID Administrative and Financial coordinator throughout the four years of the project consisted mostly in :

- Setting up and maintaining the various communication tools, both internal (BSCW) and external (public website),
- Ensuring quality of the deliverable submission process and submitting 200 deliverables to the European Commission,
- Preparing the management reports and cost statements for the consortium,
- Implementing the contractual changes during the lifetime of the contract,
- Creating and maintaining a CoreGRID researchers list,
- Acting as the Network point of contact for the European Commission,
- Organising the annual EC review meetings.

In terms of financial coordination, the Administrative and Financial coordinator was trusted with :

- Coordinating the annual updates of the Joint Programmes of Activities (JPA),
- Managing the financial side of the various exchange programmes (fellows, researchers, industrial collaboration)
- Monitoring levels of spending and transferring EC funds to the partners according to their previous usage of advance payments,
- Recovering EC funds from under-spending or withdrawing partners, for re-allocation to more active members of the consortium,
- Collecting, validating and submitting all annual cost statements from the 42 partners.

Effort and Budget summary.

Over the four years of the CoreGRID Network of Excellence, the consortium has invested a total effort of **3,000 person-months**. As a concrete comparison, this figure represents the equivalent of 62 persons working full-time on the project for 4 years.

Translated in financial terms, the global cost of this effort as reported by the CoreGRID partners represents a total amount of **17 000 000 Euros**, of which the European Commission has contributed to the tune of **8 200 000 Euros**.

The following table reflects the Network activities that were financed by the EC grant over the full duration of the Network. In a nutshell, close to 60% of the funding was dedicated to research, while a rounded-up 10% were used to finance dissemination, mobility, scientific coordination (SCO) and administrative management (AFC) respectively. The share invested in Mobility went into the Fellowship Programme for three quarters and into the Research Exchange Programme (REP) for another quarter.

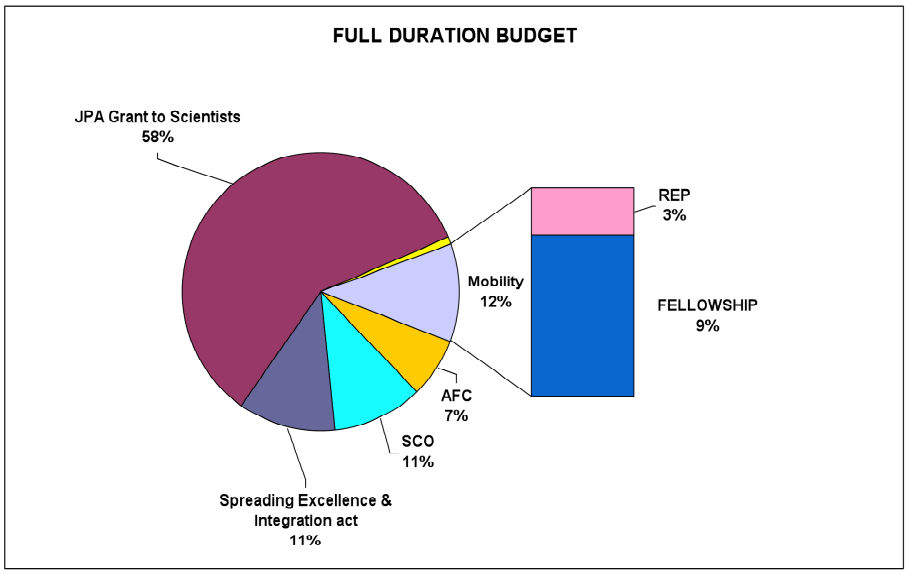


Figure 14: distribution of the CoreGRID budget

2) Scientific Coordination

Scientific coordination of the network is carried out by the Scientific Coordinator (SCO) who heads the Network for all issues related to scientific and technical matters. The SCO is responsible for ensuring self-assessment and scientific quality assurance, coordinating between and beyond scientific and technical work packages as well as for an overall follow-up of the performance of all the tasks, activities and deliverables of the Network. Thus, the SCO is also responsible for reacting to conflict situations or to negligence of any of the tasks the partners are expected to perform. His main activities during the lifetime of the network were:

- Strategic planning of all scientific and dissemination events
- Raising public participation and awareness of CoreGRID
    - Annual reports, update of the CoreGRID web site
- Representing the Network in meetings with the European Commission on technical and scientific matters
- Monitoring of integration with the help of the Integration Monitoring Committee
- Assessing and evaluating the research activities carried out by the CoreGRID partners
- Heading of the Executive Committee and organizing face to face meeting and phone conference
- Organizing the self-assessment and scientific assurance
- Coordinating the preparation of the Joint Program of Activities (JPA) on an annual basis
- Organizing the assessment of the network by the Scientific Advisory Board on an annual basis

IX. Conclusion and roadmap

After four year of existence, CoreGRID has reshaped the research landscape in Grid computing in Europe. This has been done thanks to a large number of actions that encouraged and fostered collaboration between researchers. The CoreGRID management team has been driven by a unique goal: establish CoreGRID as a highly visible and sustainable research Lab in Grid computing, a more ambitious vision than just only a EU-funded project with a limited lifetime. The idea to set up CoreGRID came in 2002 and since that date, computing technologies have evolved and Grid is merging with some prominent concepts such as Service and Cloud computing. CoreGRID is playing an important role in the transition from Grid to Service computing. It already features a set of activities targeting service infrastructures, in particular in the area of Trust & Security, Service Level Agreement and Middleware Systems. However, the initial idea remains the same and our vision is still valid: how to set up a fully distributed, dynamically reconfigurable, scalable and autonomous infrastructure to provide location independent, pervasive, reliable, secure and efficient access to a coordinated set of services encapsulating and virtualising resources (computing power, storage, instruments, data, etc.) in order to generate knowledge.

Grid computing has paved the way towards new computing systems that consider the Internet as a computing infrastructure per se. The Grid research community can be proud of what it has achieved over the last ten years: it has shown that a large scale distributed computing infrastructure can be implemented and deployed over the Internet to support the execution of e-Science applications and to serve a wide spectrum of users, thereby answering their needs for advances in their own research field.