

Next Generation Grids 2

Requirements and Options for European Grids

Research 2005-2010 and Beyond

Expert Group Report
July 2004

Commission Disclaimer

This document contains information provided by a group of independent experts convened by the European Commission with the objective to identify potential European Research priorities for Next Generation Grid(s) 2005 – 2010 and Beyond. The document does not necessarily reflect the view of the European Commission.

Authors' Institutional Disclaimer

The views represented here are those of the individuals forming the group and do not necessarily represent those of the organizations to which the individuals belong.

1. EXECUTIVE SUMMARY	4
2. RATIONALE.....	7
2.1. EUROPEAN CONTEXT	7
2.2. OBJECTIVES.....	7
3. BACKGROUND, GAP ANALYSIS AND SWOT ANALYSIS.....	9
3.1. BACKGROUND.....	9
3.2. GAP ANALYSIS.....	12
3.3. SWOT ANALYSIS.....	12
3.4. THE NEEDS AND OPPORTUNITIES FOR A JOINT PUBLIC-PRIVATE PARTNERSHIP.....	15
4. SCENARIOS AND NEW REQUIREMENTS	17
4.1. DISASTER SCENARIO.....	17
4.1.1. <i>Description</i>	17
4.1.2. <i>Challenging aspects</i>	18
4.2. PROACTIVE PDA SCENARIO	19
4.2.1. <i>Description</i>	19
4.2.2. <i>Challenging Aspects</i>	20
5. RESEARCH AGENDA AND TOPICS.....	22
5.1. INTRODUCTION.....	22
5.1.1. <i>End-user vision</i>	23
5.1.2. <i>Architectural Vision</i>	23
5.1.3. <i>Software Vision</i>	24
5.2. RESEARCH PRIORITIES	25
5.2.1. <i>Properties</i>	25
5.2.2. <i>Facilities</i>	26
5.2.3. <i>Models</i>	27
6. REQUIREMENTS FOR EU-LEVEL ACTION	29
7. ROADMAP TO FP7 AND BEYOND	30
8. CONCLUSIONS	31
9. ANNEX NGG2 MEMBERSHIP.....	32
9.1. PARTICIPANTS TO NGG AND NGG2.....	32

1. EXECUTIVE SUMMARY

The popularity of Grids has been growing very rapidly driven by the promise that they will change dramatically the life of individuals, organisations and the society as much as the Internet has done in the past decade. By enabling knowledge and computing resources to be transparently delivered to and used by citizens and organisations as traditional utilities, Grids have the potential to give new impetus to the IT market and boost growth and competitiveness in many industrial and business sectors.

The ongoing convergence between Grids, Web Services and the Semantic Web is a fundamental step towards the realisation of a common service-oriented architecture empowering people to create, provide, access and use a variety of intelligent services, anywhere, anytime, in a secure, cost-effective and trustworthy way.

In the first half of 2003 a group of high level experts was convened by the European Commission, Unit INFSO/F2, in order to produce a report titled "Next Generation Grids, European Grid Research 2005-2010". In this report, known as the NGG report, the experts have pioneered the vision of the 'Invisible Grid', i.e. whereby the complexity of the Grid is fully hidden to users and developers through the complete virtualisation of resources, and have sketched the research priorities underpinning the realisation of the Next Generation Grids.

The NGG report 2003 has complemented the description of the research scope and objectives of the FP6 2nd Call for the Strategic Objective "Grid-based systems for complex problem solving" reflected in the IST work programme 2003-2004. Following the results of the evaluation of this Call, a number of projects which are being launched in the course of 2004 will address several of the research challenges identified by the NGG report. It is expected that such actions will make significant progress beyond the state-of-the art in Grids in terms of new architectures, middleware and generic service components necessary to make the Grid an economically viable utility for industry, business and the society.

The NGG report 2004 has identified new additional requirements that have arisen in the light of one more year of experience of the experts working in the Grids domain. In particular, the shortcomings of existing Grids middleware (see section 3) are much better understood, and despite the development directions concerning OGSA providing some integration with services-oriented architectures, it is becoming clearer that applications in the Grids environment require a greater range of services than can be provided by the combination of currently evolving Grids middleware and existing operating systems. This is particularly apparent in the area of knowledge and semantics; there is a need for semantically-rich knowledge-based services in both Grids Foundations Middleware and Grids Services Middleware both to improve the functionality but also to support applications on the Grids surface which require semantic and / or knowledge-based support. This leads to the proposed architectural stack (see section 5.1, Figure 2) where the development of applications is supported by three abstraction layers consisting of Grids Service Middleware, Grids Foundations Middleware and Operating system.

As the Grids concept matures from metacomputing towards a pervasive Ambient Intelligence surface for wealth creation and improvement of the quality of life, the number of nodes increases dramatically and the range of node capability widens considerably. Addressing individual components as well as assemblies (composed atomic components) becomes commonplace, allowing ad-hoc dynamic creation of IT organisations to support business or societal applications. This has revolutionary implications for security, scheduling, data movement, code movement and concepts of 'state', transactions and messaging. New demands for performance, reliability, self-healing and self-management become apparent.

The Grids environment must behave intelligently using knowledge-based techniques, including semantic-rich operation and interoperation. Long-cherished computer science principles are re-examined in the light of the new requirements. There are particular implications in managing the software complexity demanded by the requirements derived from the applications envisaged; this has aspects of software theory, design, construction practice and also tools and environments to assist software development.

The NGG2 expert group has considered in-depth these issues and concludes that future calls should include:

- (a) development of a design for a new operating system that provides a fault-tolerant, scalable, self-healing, self-managing environment upon which Grids service middleware may 'sit';
- (b) development of Grids foundations middleware suitable both for enhancing existing operating systems and for inclusion within (a);
- (c) development of Grids service middleware in a modular fashion allowing applications to utilise those services they require;
- (d) research and development in computer science and information technology required to accomplish (c), (b) and (a), notably new models and software for transactions and messaging; for scheduling, resource management and optimisation; for trust, security and privacy; for data, information and knowledge management; for software development and deployment including mobile code; and for intelligent and appropriate user interfaces and device interfaces;
- (e) development of novel applications that are wealth-creating or improve the quality of life, particularly in the e-business domain, but also in e-health, e-environment, e-culture, e-science, e-government;

Each of these areas should be accomplished in two phases: preliminary investigations, prototypes, demonstrators should be requested in FP6 IST later calls; full developments and deployments should be pursued in early FP7 IST calls.

The development of a coherent programme of research activities in the above areas is considered essential to consolidate and further strengthen the European competitiveness in Grid technologies and in other related areas such as software and service technologies, and applications.

In view of its broad scope and cross-disciplinary character, this research should be developed through different types of funding interventions in order to build a critical mass at the pan-European level.

In addition to the FP6 instruments of IPs (Integrated Projects), NoEs (Networks of Excellence) and STREPs (Specific Targeted REsearch ProjectS), the creation of an industry-driven public-private partnership on service-oriented architectures, to be implemented through mechanisms such as the European technology platform concept (ETP) described in the Communication of the European Commission (COM(2004) 353), could be envisaged to help to streamline research to engage industry to develop a shared, more integrated vision, and to elaborate routes for more effective implementation.

This report, in conjunction with the NGG report 2003, establishes a European vision and technological requirements towards the realisation of the “invisible Grid”, offering key features for a service-orientated knowledge utility, a new paradigm for software and service delivery, for the next decade.

2. RATIONALE

2.1. European Context

In the first half of 2003 the European Commission DG INFSO took the initiative to launch a high-level group of experts to produce a report titled "Next Generation Grids, European Grid Research 2005-2010" (see <http://www.cordis.lu/ist/grids>). In the report, widely known as the NGG report, the group has identified numerous research challenges that need to be faced to enhance the capabilities of current Grids and transform the current Grid into an economically viable tool for industry, business and the society.

The European research context is evolving very rapidly. While there is a general commitment to achieve the Lisbon objectives of increasing the investment in research up to 3% of GDP by 2010, discussions are ongoing: which issues are new or more relevant since the launch of FP6; what is the socio-economic impact of European research; what are the motivations for research investments; how to co-ordinate and structure research efforts across Europe beyond 2006.

A new Commission for the enlarged Europe will be established in the course of 2004; a new financial envelope will open in 2007; and a new EU research Framework Programme may cover the period 2007-2010.

Within this changing context there is a paramount need for some fresh and radical thinking on the future of Grids research within the broader context of ICT research and on the socio-economic rationale for action. Now is the time to make reflections more concrete.

After almost one year from the release of the NGG report the Commission Services, DG Information Society is now calling upon an enlarged Expert Group (NGG2) to complement, enhance, and where necessary extend the NGG report along the following objectives:

2.2. Objectives

Short/medium-term horizon

- To structure and prioritise Grids research themes, building on and possibly updating the NGG vision in view of the preparation of IST-FP6 Work Programme 2005-2006.
- To propose a research agenda and a roadmap for implementation taking into account (to the extent possible) the research addressed by the projects being launched as a result of the IST Call 2 evaluation.

Long-term horizon

- To further consolidate and expand the present Grid research vision (2005-2010) beyond 2010 within the context of the evolution of the whole ICT area and its link to wider socio-economic policies.

- To anticipate the possible scientific and technological (S&T) challenges in Grid research beyond FP6 taking into account cross- and inter-disciplinary aspects to explore ideas and theories that go beyond the capabilities of the Next Generation Grid.

Implementation issues

- To envision possible mechanisms to address both technical and non-technical barriers to and requirements for the optimal development, standardisation, deployment and use of Grid technologies, while taking into account relevant Community policies.
- To identify the main structural elements of research supporting the above visions (e.g. basic research, exploratory research, applied research, etc.) in view of the preparation of the Framework Programme 7.

3. BACKGROUND, GAP ANALYSIS AND SWOT ANALYSIS

3.1. Background

Today's trend in the IT market in shifting revenues from the sales of products towards the provision of on-demand services is expected to continue, driven by the increasing need of cutting costs of operations management and making business processes more effective. Several large IT multinationals have developed or adapted their concepts and strategies towards the emerging paradigm of providing IT services as a set of utility services, in a fashion similar to traditional utilities. Recent market forecasts made by highly renowned business analysts stated that the worldwide market for IT services is expected to increase considerably in the future; in particular, process management services will grow fastest, as demand for outsourcing IT management and applications rises. Grid technologies have the potential to drive the market evolution of the IT industrial sector "toward IT services".

In parallel with this trend, IT infrastructures are experiencing a fundamental paradigm shift in moving away from scalable client/server approaches towards decentralised, scale-free, service-oriented approaches. They may consist of millions of heterogeneous networked components and billions of "non-functional" dependencies. The unprecedented level of complexity, instability and pervasiveness reached by today's IT systems very often leads to situation where local or component failures can be propagated without control across the IT infrastructure and degenerate into global crashes or misbehaviors.

When confronted with such a level of complexity, the traditional computing models show limitations as they lack the capabilities, constructs and associated semantics necessary to express emergent and non-functional properties and behaviors and ensure fundamental properties such as composability, consistency, completeness. Furthermore, the existing implementation models (mostly based on the inter-leaving of several middleware layers) offer very little support for managing, adapting, and reacting to complex contextual changes and failures.

Clearly, new theoretical foundations are required to allow each component to react to changing circumstances dynamically and adapt to failures without compromising the system function or performance as a whole. Key research issues are, for example the development of new flexible and adaptable architectures; new software engineering methodologies, languages and tools, new approaches to operational semantics, new computational models to express emergent and non-functional behaviours; new approaches to resource virtualization, interoperability and integration; new models for trust and security; complemented by horizontal issues such as economic models for IT service provision, societal and commercial acceptance.

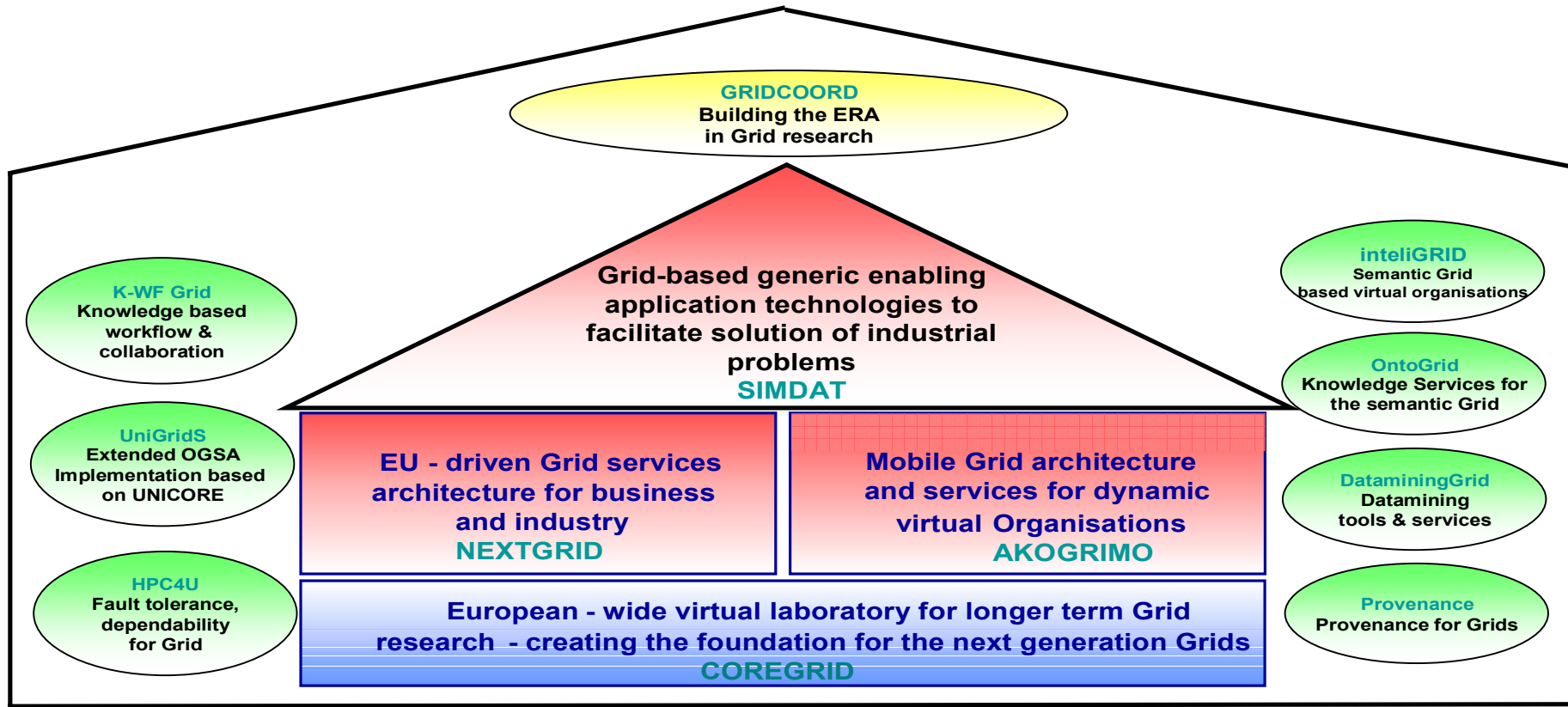


Figure 1: The Call 2 Projects as a 'house'

3.2. Gap Analysis

The NGG2 expert group has analysed the research scope of the Call 2 projects under negotiation and assessed the level of coverage of the research priorities required for realising the Next Generation Grid, i.e. properties, facilities and models (see Section 5.2). For simplicity reasons only large Grids projects within the project portfolio of the unit INFISO/F2 were considered, however it should be noted that other Grid-related projects are also running or being considered for funding within other areas of the IST priority that could address one or more of the research issues identified in the analysis.

The Call 2 projects can be represented conveniently as a ‘house’ (Figure 1); this diagram has been provided by the unit INFISO/F2 and is utilised with their permission.

The results of this analysis are summarised in the following table:

NGG1 Priorities \ Coverage	Weak coverage	Good coverage
Properties	transparency, reliability, pervasiveness, persistence	open-ness, security, usability, standardisation, person-centric, scalability, configurability
Facilities	information representation system management	coordination, virtual organisations
Models	Grid economics	business models, user interface

Table1: Summary of Gap analysis of Call 2 proposals vs. NGG research priorities

3.3. SWOT analysis

The NGG2 expert group has carried out a SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis of European Grid capabilities and research. The SWOT analysis highlights the strategic importance of Grids both as primary ICT technology on which Europe can build its strengths, and as a key enabling tool for boosting productivity, competitiveness and growth of European economy.

The Commission has already sought advice on vision and research challenges in several ICT-related fields. For example, the IST Advisory Group Report of June 2002 has produced an extensive report in the area of software technologies, embedded and distributed systems. Many considerations identified in this report, such as the for example the effects of OS vendor monopolies and skills shortages, obviously apply to Grids software development as well. Our SWOT analysis focuses mainly on those issues which have direct impact on Grids technology, or where Grid technologies may have a wider impact on ICT policies. The complete SWOT analysis is described in the following table.

Strengths	Weaknesses
<p>Ontologies and web semantics: EU member states have a strong presence in the research and business communities working on the theory, tools and applications of ontologies and the semantic web. These technologies will be crucial if Grids are to provide scalable support for complex, heterogeneous applications.</p> <p>Collaborative approach to data: Member states benefit from national and European programmes for the collection and curation of research data and support for networking between researchers. This has led to a significant number of world-leading data collections and an open, collaborative approach to knowledge sharing, which is already forming a valuable underpinning for Grid based research. European research activities in the area of knowledge Grids represents a valuable contribution that can enforce the EU technology in data and knowledge management.</p> <p>Mobile devices and embedded systems: The strength of the European telecommunications industry and the diversity of its market for electronic control systems has given Europe a leading position in some areas of mobile and embedded technology. This is of particular relevance to the vision of Grids as a pervasive, user-centred utility.</p> <p>Grids Middleware: Europe has established a strong position on higher-level Grids middleware research. Most European developments have their particular strength in offering high-level services and tools for building applications, targeting the special needs of user and service provider communities such as industrial engineering, business applications, and application services provision. In this research area of higher level Grids services, European players seem to have gained a worldwide leading role.</p> <p>Semantic Grids: In Europe there is a long-standing expertise in Semantic Grids, agent and cognition technologies, which could be better capitalized and integrated towards Grids developments.</p>	<p>Hardware: Europe lacks research and commercial leadership in desktop and server scale computer system design. This may hamper development of efficient, scalable Grids components in favour of more traditional server-based architectures.</p> <p>Operating Systems: Historically, support for adopted middleware technology has eventually migrated to the operating system level. Middleware and Grids technology choices made by the major OS vendors may therefore have a disproportionate effect on the uptake of particular Grids technologies.</p> <p>Programming Languages: A significant proportion of Grids and web service research and application development is dependent on language platforms which lie outside the control of European user communities. Despite excellent European academic R&D there is little pull-through to wealth creation. There is a problem overcoming the inertia of existing widely-used languages.</p>

Opportunities	Threats
<p>Paradigm shift toward IT services: Today's trend in the IT market in shifting revenues from the sales of products towards the provision of on-demand services creates unprecedented opportunities to develop a European competitiveness in IT services, which could have a catalysing, positive effect on adjacent sectors both upstream and downstream the user-supplier value chain.</p> <p>Operating system virtualisation: The adoption of Web Service technology by the major OS vendors allows the development of distributed applications that are independent of the underlying operating system and language technologies. The convergence between Grids and Web Services therefore provides a significant opportunity to move to a model of software development and service provision where the market dominance of particular OS vendors is no longer a major economic issue.</p> <p>Existing infrastructure for collaboration: FP6 already has instruments in place to allow for the effective exploitation of Europe's diverse Grids-related research community, and can rapidly build collaborative projects to allow researchers and businesses to exploit Grids applications.</p> <p>Service Model for Industry: The Grid today is used most heavily in particle physics, environmental science, life science applications, genomic research, protein folding, and medical applications, in advanced engineering R&D, in chemistry and materials science. It is expected that business, like finance or media, and many industries, such as aerospace, automotive, or entertainment will seize the opportunity to use existing IT resources more efficiently. Grids offer a new range of service models for the IT service industry.</p> <p>Standardisation for European Advantage: a window of opportunity exists for a pan-European effort to develop the Grids foundations / middleware and have a workforce familiar with Grids so that Europe (business, industry, science, healthcare, environment, culture, education...) may have an interoperating advantage.</p> <p>Next generation Grids: The distinctive European vision of Grids operating from the level of devices to supercomputers, to serve communities ranging from individuals to whole industries, could have a significant economic and social impact far beyond the scope of existing compute and data Grids.</p>	<p>Dependency on development tool support: Support for interoperable messaging protocols (such as SOAP) depends on the tools provided by the various language and OS platform owners. While at the moment there is agreement on the overall direction of Grids middleware and Web Service evolution, disputes or changes in policy over supported technologies could have a rapid impact on the ability of Grids developers and service providers to support particular language and operating system combinations.</p> <p>Standards evolution: As Grids technologies mature and become more complex, the adoption of standards (official or de facto) will be a requirement for sustained development of Grids, and only applications compatible with those standards will gain widespread adoption. It is vital that any European vision for the evolution of Grids is accompanied by a clear representation of that vision to the key standards bodies and technology providers worldwide.</p> <p>Non-acceptance and Lack of Use: Industry may not accept the developed / developing European Grids Foundations / Middleware leading to a non-interoperable environment thus reducing the potential market size and the advancement of the knowledge society.</p>

Table 1: SWOT analysis of European ICT capabilities and research as they relate to Grids technologies.

The main conclusions of the SWOT analysis can be summarised as follow:

- Ontologies and semantic web technologies will be crucial to provide scalable support for complex, heterogeneous Grids middleware and applications.
- The strengths of the European telecommunications industry and the diversity of its market for electronic control systems have given Europe a leading position in the areas of mobile and embedded technology. This is of particular relevance for the realization of the vision of a Grid as a pervasive, user-centered utility.
- The weakness in hardware and primary software products (e.g. commodity processors, server and desktop Operating systems, Programming Languages, etc.) may hamper the development of a European leadership in Grids Technologies.
- The convergence between Grids and Web Services provides a significant opportunity to move to a model of software development and service provision where the market dominance of particular OS vendors is no longer a major economic issue.
- The distinctive European vision of a Grids environment that operates from the level of devices to supercomputers, to serve communities ranging from individuals to whole industries, including data, information and knowledge and emphasizing resilience and scalability could have a significant economic and social impact far beyond the scope of existing compute and data Grids. This should be contrasted with the North American Grid vision of programmer-level metacomputing.
- It is vital that any European vision for the evolution of Grids is accompanied by a clear representation of that vision to the key standards bodies and technology providers worldwide.

3.4. The needs and opportunities for a joint public-private partnership

In order to exploit the strengths and opportunities described in the SWOT analysis and to capitalise from the benefits of the paradigm shift towards IT services, it is suggested to investigate the possibility to establish an industry-driven public-private partnership on service-oriented architectures to help to streamline research and engage industry to develop a shared, more integrated vision, and to elaborate routes for more effective implementation. Such an initiative could be implemented through mechanisms such as the technology platform concept described in the Communication of the European Commission (COM(2004) 353).

An industry-driven public-private partnership on service-oriented architectures could exert a catalysing effect on a number of adjacent sectors, both downstream and upstream the user-supplier value chain and should help reducing the competitive gap of Europe in those IT areas that are very critical for boosting competitiveness but where Europe has demonstrated chronically weakness (for example general-purpose Operating Systems, Databases, Development Environments, etc.). At the same time, the proposed partnerships could

represent an essential building block accelerating innovation in several industrial, business and society application sectors, therefore underpinning the Lisbon's objective of making Europe the most competitive knowledge-based society in the world.

4. SCENARIOS AND NEW REQUIREMENTS

Following on from the NGG1 Report perspectives (end-user, architectural, software) and analysis, the group considered scenarios as motivational examples: the 'Disaster Scenario' requiring crisis management and the 'Proactive PDA' requiring services for the citizen in both business and leisure roles in their widest senses. From these scenarios, and from the previous analyses (Context and SWOT Analyses) the new requirements (Research Agenda and Topics) are derived.

4.1. Disaster Scenario

4.1.1. Description

We live in an era with both sudden natural disasters (earthquake, volcanic eruption, landslide, avalanche, tsunami, flood, drought, forest fire, hurricane, tornado...) and long-term climatic change with consequences for sea-level rise, local climatic extremes and differentiation, increased storms. Furthermore, human activity is affecting strongly - and generally adversely - the natural environment. Human error can also cause fires, floods and other disasters. In addition there exist wars and terrorism (and social unrest, demonstrations and revolution) with consequent damage to the environment and to people.

This scenario involves situations where various crisis or disasters should be handled correctly and efficiently by public safety work forces and other mission critical mobile workers. This scenario depicts proper handling of crisis situations, but not preventing such situations from happening. Nevertheless, proper handling of a crisis that has already occurred is of great importance in order to reduce the probability of inflicting damage to people and material involved in the crisis, and in order to prevent a crisis turning into a tragedy. Crisis prevention scenarios were not handled by the NGG expert group because the group felt such scenarios involved issues that were out of scope for the mandate that is set for the group (e.g. privacy issues, intelligence issues).

Crisis management scenario requires that mobile workers (police, fire fighters paramedics, environmental monitors, military, etc.) collaborate in time critical and dangerous situations, and have real-time access to information and knowledge in order to improve their decision-making processes under demanding circumstances. Such cases involve large numbers of endangered people, and can happen in:

- big sport events, big concerts
- airports, railway stations, ports,
- rescue forces in disaster areas

Modern security personnel deal with a number of difficult missions where technological and scientific tools can improve the resulting security. The scenarios are extended to the whole duration of events (before and after these events have taken place) and incorporate a variety

of devices and infrastructure facilities that are needed for this (including mobile phones and telecommunications networks, PDAs, cameras, notebooks, wireless hotspots and others).

4.1.2. *Challenging aspects*

In many crisis management situations the commanders and the security personnel (along with medical team) have to operate outside their facilities and all the relative built-in equipment. These scenarios have some specific requirements that have to be taken under consideration and go beyond those identified in NGG1 Report.

Enhancing of senses for the mobile users

Accurate, relevant information has to be presented in the optimal way to support the end-user. Mobile devices are rather weak in performing computing intensive actions. This can be done through the use of the high performance computing capabilities that are assigned mainly to base stations. The enhancement of senses mainly corresponds to optimized:

- Pictures, videos, and stereo sound analysis: Victims in a crisis situation often carry mobile devices that can be incorporated into the crisis management infrastructure (e.g. providing images and video clips of the surrounding area, being used to instruct victims or rescue personnel) that will optimise communication in stressful environments
- Dynamic access to information from co-operating mobile devices of the team members or victims: Mobile devices of the victims, as well as the rescue personnel, can be used as information source for crucial real-time information, e.g. spatial and temporal coordinates of the victims and team members.

Intelligent decision support for both the individuals and the control centre

Mobile units collect and process information locally. However, they are also capable of providing information to a central unit in order to post-process it, perform simulation of the situation, extract useful information and statistics, and in the sequel to predict actions so as to support decisions for the pro-active handling of the emergency situation. Orientation of the crisis team on terrain, positioning of key-personnel in specific points, population of victims in different areas of the crisis field, as well as classification of the risk are some of the decisions that will be supported by the system.

Distributed mobile ad hoc network

These critical situations require that the portable devices should operate independently from the central station. In cases where the central station has been damaged or is temporarily unavailable, the mobile devices should be able to co-operate so as to perform a set of services, providing thus tolerance in some emergency cases. This is also necessary in cases where small teams have to operate independently for some time, and local data can be cached, avoiding causing overhead to communication links and computational capacity of the central station.

Access to remote databases

This aspect refers to the ability of the mobile staff to have access to remote information distributed over the world. For instance, in case of a sports event involving tens of thousands of people, access to the medical record of each victim (regardless of where the victim comes from) will have crucial impact on the results of the healthcare operation. This ability is especially needed in cases where time-critical situations should be handled and any delay through information mediators may influence the result of the actions.

Within the framework of one of the funded FP6 Call 2 IPs, the Crisis management scenario will be based, as much as the situation can allow it, on the realistic scenario that will rise from the Olympic Games of Athens. The Olympic Games will provide an ideal candidate for requirement collections that will be used for the scenario definition and layout. The information that will be gathered during the Games will mainly be used to identify potential instances of a Crisis Management test-bed. It will provide realistic figures for size and kind of the sport centre, size and hierarchy of the security team, degrees of freedom for the team, population that is inside the sport centre, kind of disaster etc. Such scenario can involve the evacuation procedure of a big stadium due to fire alarm, terrorist action, etc.

4.2. Proactive PDA Scenario

4.2.1. *Description*

The key to most human activities (e.g. business, agriculture, extraction, manufacturing, logistics, retailing, science & technology R&D, environment, health, culture, entertainment, government, social services) is efficient provision of information channels to/from the user. Increasingly this will be pervasive and continuous to a PDA (personal digital assistant) or wearable computer. For this to be effective there are requirements for:

- Excellent networking
- Excellent user interface for minimal effort and maximum utility
- Intelligent push as well as pull facilities for access to information, computation, sensors or other people
- Facilities for data collection including spatio-temporal coordinates
- Appropriate security and trust

Today's Personal Digital Assistant (for example, the Pocket PC, Palm, Psion or 'smart' mobile phone) is an interactive handheld device widely used to access personal and public information, to store personal information and to communicate with others. These devices are an example of the increasing number and variety of pervasive or ubiquitous computing devices which are set to pervade our everyday environment, as parts of 'wearables', within everyday artefacts and embedded in the fabric of our living environment – this is part of the Ambient Intelligence vision.

What we need from PDAs – and future information devices – is the right information displayed at the right time on the right device in the right format, and with the right level of intrusiveness upon the user's current tasks. This requires context-awareness: the location of

the device, the profile of the user(s), their current schedule and tasks, the capabilities of the interface, aspects of security and trust. In fact this rich notion of context becomes the query – using context knowledge, devices can act proactively and autonomously in order to provide the user with the information they require. In fact there is no need for the user to make an explicit query, just being there is enough (being who you are, where you are and doing what you are doing). This is our notion of the **Proactive PDA**. Furthermore, PDAs can also be used to capture information – as tools to support the users in a creative process of collecting and working with information and knowledge.

Consider for example the user who walks into a business meeting and looks at his/her PDA for background information about others in the meeting, together with up-to-the-second stock market data, and then seeks a prediction about the economic impact of a particular decision. Add to this the creation of content through taking notes, and the need for rapid access to the relevant parts of previous meetings. Or the example of several people travelling separately into a foreign town to meet for a meal, their geographically distributed PDAs suggesting a restaurant with an appropriate menu and availability given the various requirements and constraints – then someone taking a photograph of the menu to obtain a translation (which, behind the scenes, involves some local and remote processing).

This vision of the proactive PDA makes significant demands on our information processing capabilities and computational infrastructure. It requires the automatic synthesis of information from multiple sources, pre-emptive or on-demand, customised for delivery to the user's device. It may require computation for search and synthesis but also for purposes of modelling and prediction – which in turn may inform the proactive behaviour of the devices. These examples are very much focused on the user, on dynamic behaviour and timeliness. Some of the 'back end' of this functionality draws on the vision of the Semantic Web, the extension to the current Web in which information is given well-defined meaning to facilitate automated processing. These technologies are also used to enable the description, discovery and automated composition of the services in order to meet the user's requirements.

These scenarios are Grids applications. They require the assembly of data and computational resources to meet application requirements, they involve the creation and support of virtual organizations and they need an infrastructure, which provides interoperability but also security and appropriate models for service negotiation and charging. They also require a user-centric perspective, working with context and user information, and new qualities of Grids service to address the dynamic aspects such as timeliness and change.

4.2.2. Challenging Aspects

In summary, the key additional research challenges, arising since the publication of NGG1, are:

- An increased user-centric focus on how applications of next generation Grids are manifested to the user via pervasive computing devices.
- The additional information representation requirements for context- and person-awareness, supporting proactive behaviour. This includes ontologies for aspects of situation, user and task.

- On-demand and timely presentation of information, requiring dynamic composition and negotiation of services. This creates challenges for negotiation, orchestration and scheduling.
- Pre-emptive behaviour by Grids, which is related to autonomous behaviour, and contrasts the traditional view of Grids as a batch processing system.
- Valuable information representation on small devices, synthesis of knowledge models on wireless devices for ubiquitous use.

5. RESEARCH AGENDA AND TOPICS

5.1. Introduction

The Grids concept demands novel approaches to system design and management – and thus to operating system behaviour, middleware requirements and services offered to applications. There is emerging a three-level conceptual architecture. The application requirements provide a specification for the required services in the Grids Service Middleware layer, and this in turn drives requirements of the Operating System layer including the Grids Foundation Middleware required to elevate the interface of each Operating system to that required for the Grids Service Middleware. This is particularly apparent in the area of knowledge and semantics; there is a need for semantically-rich knowledge-based services in both Grids Foundations Middleware and Grids Services Middleware both to improve the functionality but also to support applications on the Grids surface which require semantic and / or knowledge-based support. It is also apparent in Grids scheduling, management, durability and dynamic reconfiguration.

It is becoming increasingly clear that existing Grids Foundations Middleware such as GLOBUS and associated software does not provide the required support to enable easy and reliable application construction and execution in a Grids environment. Not least because of their underlying use of a client-server model, they fail to scale to a Grids environment - with billions of heterogeneous nodes - which is self-healing and self-managing.

Furthermore, current operating systems (dominantly LINUX in work to date but increasingly Symbian and others used in embedded systems) do not provide the necessary services for effective operation of the Grids Service Middleware layer. Again, the model of the operating system controlling a single node and managing its resources exclusively (security, scheduling) is at variance with the philosophy of Grids. This leads to the concept of augmenting existing operating systems with Grids Foundations Middleware to provide the required functionality. The key here is the definition of the interfaces above (to the application) and below (to the ideal Grids Operating System) of Grids Service Middleware. Any gap between the Grids Service Middleware and the Operating system is filled by the Grids Foundation Middleware. These interfaces determine respectively the capabilities and features of applications and the performance, reliability, security and other aspects of the operating system(s).

The visions and research priorities below are intended to extend those of NGG1 Report to cover requirements emerging since that time. The overall Grids architecture envisaged is sketched below (Figure 2). It should be noted that the Grids Service Middleware is envisaged as a library of interoperating services and that any application on the Grids surface utilises those components it requires. Similarly, Grids Foundations Middleware is envisaged as a library of interoperating components and the appropriate components are utilised to augment any particular operating system in order to meet the interface (and thus provide the facilities) expected at the base of the Grids Service Middleware.

Application A	Application B	Application C
Grids Middleware Services Needed for A	Grids Middleware Services Needed for B	Grids Middleware Services Needed for C
Grids Foundations for Operating System X	Grids Foundations For Operating System Y	Grids Operating System (including Foundations)
Operating System X	Operating System Y	

Figure 2: Overall Grids Architecture

5.1.1. End-user vision

The end-user should interact with a Grids environment using whatever mode and device is most convenient at any time. The Grids environment should accept the user request, propose a 'deal' to satisfy the request (which may involve financial or rights trading) and then – assuming the user agrees to the 'deal', execute the response to the request. The key aspects of the end-user vision include pre-emptive and proactive action by the Grids environment but also a Grids environment that performs adequately and provides continuity of service. The end-user or user application in a Grids environment in fact interfaces to the Grids Service Middleware - except for any requirements for direct invocation of Ideal OS primitives - and thus many of the facilities required by the end-user are provided in the Grids Foundations Middleware layer compensating for the inadequacies of Operating systems.

5.1.2. Architectural Vision

The Architectural Vision of NGG1 was somewhat vague about the locus of services between the middleware and the operating system. With NGG2 we are making this more precise and assuming an architecture where the Grids Service Middleware

- provides a target interface for applications such that the applications request and receive the services they require,
- provides the services themselves,

while the Grids Foundations Middleware provides mappings to operating systems capabilities to provide those services or components required of them where unavailable.

This allows for multiple operating systems to be utilised within nodes in the Grids environment and thus extends the range of entities (devices, services) available in the Grids environment including embedded systems and mobile phones or PDAs. It predicates the requirement for well-defined interfaces above and below the Grids Service Middleware layer and the Grids Foundations Middleware layer. The common interfaces are above (to applications, i.e. the Grids surface) and below (to Grids Foundations Middleware i.e. the defined Grids OS interface) the Grids Service Middleware layer.

The motivational examples also emphasise the need for an architectural vision which has resilience, self-healing and self-managing properties. There is also a further need for enhanced semantic reconciliation between interoperating nodes.

A Grids environment, based on Grids Service Middleware plus an OS (whether a modified existing OS, a meta-OS bridging above existing different OSs, OSs extended by Grids Foundations Middleware or a novel OS) must be able to handle dynamic resources that are spread over multiple administrative domains, each with its own policy. Several important assumptions used in traditional OS design are no longer valid, thereby requiring a re-thinking of the traditional OS principles:

- The notion of a global state is only of conceptual use, because a global state cannot be observed due to the lack of a physical clock. This raises difficulties in replication control, synchronization, check-pointing, etc.
- The four ACID properties (atomicity, consistency, isolation, durability) are difficult or impossible to implement.
- Service response time can hardly be guaranteed – even with QoS networks and real time OS services. Many services are just best effort. Services and resources may be relocated or taken out of service without prior announcement.
- Implementing security and trust requires new conceptual thinking.
- Scalability is given on a conceptual level, but can only be guaranteed on a statistical basis.

5.1.3. Software Vision

NGG2 scenarios, which emphasise aspects of Grids beyond a metacomputing research Grid especially in the areas of knowledge engineering, semantics, reliability and scalability lead to further software vision elements building upon those of NGG1 Report. In particular, the use of metadata in the form of resource or service descriptions, functional signatures of processes, characterisation of a user, restrictions (security, privacy, charging) predicates a new approach to software especially in the Grids Foundations layer. It is likely that metadata-configured intelligent agents, acting on behalf of Grids entities, will interact through brokers to achieve the service provision required by the applications and / or the end-user in the Grids Service Middleware Layer.

Software productivity implies that software solutions will be high level, declarative, functional, with signatures describing (as metadata) functional capabilities and appropriate

interfaces for data and messaging. These software solutions will require support for tools to facilitate the design, construction and deployment of services and applications.

5.2. Research Priorities

5.2.1. Properties

In addition to the properties mentioned in NGG1 Report, a Next Generation Grids environment should have the following properties in order to satisfy the requirements of the scenarios considered:

- (a) **pervasive**, with mobility as the cornerstone enhanced with more advanced pervasive computing facilities;
- (b) **self-managing** with the ability to handle highly dynamic and unpredictable configuration of demanders and suppliers;
- (c) **resilient** with the ability to handle highly dynamic and unpredictable configuration of the network connecting the computing nodes;
- (d) **flexible** to handle various types of computing nodes and highly dynamic distribution of computation tasks among involved resources;
- (e) **resilient** with the ability to handle intermittent connectivity and associated synchronisation of information sources;
- (f) **easy to program** with a high-level, functional programming interface reusing existing software modules;
- (g) **flexible in trust** to allow business operations to work effectively and efficiently as virtual organisation and distributed collaborations;
- (h) **secure** to assure confidence in its use for business purposes.

NGG1 produced a list of properties that are desirable in a Grids environment. These are all at least partly supported by the OS (augmented by Grids Foundations) layer:

- (a) reliability;
- (b) security & trust across multiple administrative domains;
- (c) persistence;
- (d) scalability;
- (e) open to wide user communities;
- (f) pervasive and ubiquitous;
- (g) transparent, easy to use and program;

- (h) person-centric;
- (i) based on standards for software and protocols.

We now wish to stress for the OS / Foundations environment additional aspects on each item of this list:

- (a) self-adaptive, self-healing , self-managing and self-reconfiguring;
- (b) more sophisticated role-based security and trust between operating system instances or components;
- (c) extended in the sense of business continuity;
- (d) scale-independent;
- (e) open for interoperation – cooperating operating systems or components;
- (f) extended with the concept that the OS should be modular so that minimal configurations can be used without sacrificing interoperability;
- (g) a clear and open interface for Grids Foundations Middleware to Grids Service Middleware;
- (h) extended in the sense of context-aware geographically, temporally and role-based;
- (i) re-use of standards in operating system components to encourage interoperability and to provide a consistent interface to Grids foundations;
- (j) appropriate power consumption and code-size for the Grids entity (e.g. nano device);

5.2.2. Facilities

The motivating examples indicate that a Grids environment (particularly the Grids Service Middleware) should offer (in addition to those characterised in NGG1 Report) the following:

- (a) flexible, dynamic, reconfigurable resources available on demand and to the level required for the application and / or end-user;
- (b) accurate, relevant information presented in the optimal way for the end-user which implies reconciliation across heterogeneous distributed information systems;
- (c) context awareness, task awareness and service negotiation capability driven from the user interface;
- (d) pre-emptive and proactive services as seen by the end-user;

In addition to the facilities provided in a Grids environment listed in NGG1, in the case of Grids operating systems / Foundations / Service Middleware the facilities to be provided include:

- (a) a platform for cooperative and collaborative working;
- (b) a mechanism for managing service and participation-level agreements;
- (c) Grids file systems;
- (d) scalable naming services;
- (e) data consistency / replication;
- (f) process management and scheduling;
- (g) synchronisation;
- (h) mobile services including synchronisation, communication with an intermittent connection;
- (i) messaging support;
- (j) external interface for Quality of Service support;
- (k) external interfaces for resource usage in support of management, optimization, usage tracking and accounting, e.g. for costing and charging;

5.2.3. Models

The Grids environment will provide facilities to support the initiation of new modes of participating in business, science, education, healthcare, culture, leisure and social life, etc. These new opportunities are characterised by requirements for appropriate information presentation and process handling (workflow) to satisfy the business models of the (virtual) organisation.

Furthermore, the Grids environment will then need to support such applications with the properties and facilities already described.

As already indicated, there is emerging the concept of a 3-layer architecture below applications with Grids Service Middleware, Grids Foundations Middleware and Operating system. The specification of the desirable interfaces above and below the Grids Service Middleware will be key. This will place demands on the Operating system layer. For the foreseeable future, it will be necessary to interface the Grids Service Middleware Layer to existing operating systems; this can be accomplished by modifications to the operating systems (where permitted) or by the Grids Foundations Middleware providing a layer of 'squeezeware' between the operating systems below and the Grids Service Middleware layer above, thus preserving both the interfaces and doing mappings / transformations between. Alternatively, a more radical approach is to build an operating system that can operate across

the various Grids environment nodes – from RFID tags through sensors, mobile phones, PDAs, laptops and on to servers and supercomputers with massive data stores. This new OS should meet the lower interface of the Grids Service Middleware layer and should include the Grids Foundations Middleware, presenting the applications interface at the top of what is now the Grids Service Middleware layer. Such an operating system should be modular so that only the required components are loaded into a particular device / node in the Grids environment: the operating system for a simple mobile phone should be different from that for a supercomputer – but both should share interface specifications.

6. REQUIREMENTS FOR EU-LEVEL ACTION

The SWOT analysis indicates an opportunity for Europe to gain a leading position in this technology and thus assist European industry in wealth creation and assist European agencies in improving the quality of life.

The size and complexity of the projects envisaged requires supra-national effort at the European level. No one nation has a monopoly on advanced knowledge to build the appropriate technology. Furthermore, success depends on wide acceptability and again this requires cooperative European action to ensure wide understanding, acceptability and take-up of the technology.

Using European coordination of national activities, and added European resources, Europe could and should position itself optimally for a roll-out of Grids technology that is suitable for e-business, e-government, e-health, e-culture, e-learning etc. for both wealth creation and improvement of the quality of life. Such an R&D programme will deliver the knowledge society in Europe.

The instruments of NoEs (Networks of Excellence), IPs (Integrated Projects), STREPs (Specific Targeted REsearch ProjectS) and ETPs (European Technology Platforms) should be used as appropriate to achieve the objectives.

7. ROADMAP TO FP7 AND BEYOND

The report maps out requirements and suggests fruitful directions for R&D. The difficulty is to scale activities in these directions along a time axis. In general, FP6 IST Call2 projects are pushing the envelope of capability of the latest developed Grids Foundations software (middleware) as realised in GLOBUS with OGSA. CoreGrid, as a NoE, has the remit to consider novel paradigms, including programming languages.

Thus, the more radical proposed directions (like building a new Grids Operating System) are likely to need to be informed by the difficulties of the FP6 IST Call2 projects encountered during their lifetime. Preparatory work on these radical directions could be undertaken in the later part of FP6 (probably through STREPs closely linked to CoreGrid) but projects (IPs) to implement production-quality systems would almost certainly be postponed to FP7.

It is particularly important that the design work for any new Grids Foundations Middleware and operating system be done in the later stages of FP6 when experience of Grids technology that exists - and that has been further developed - over the last 5 years or so can be harnessed to provide the specifications for the new Grids Foundations Middleware and Grids Operating System, but to allow time for Europe to take advantage of a new standard design and interfaces developed in the earlier part of FP7.

The direction of research activities should take into account current computing, communication and storage devices and novel devices that will be developed in the near future as a result of hardware, software and communication R&D activities. All these devices should be considered as active resources in dispersed and pervasive Grids that must be managed by new Grids Foundation middleware and Grids Operating systems.

A schedule might be:

Later Stages of FP6:

- (1) initial R&D for Grids Foundations Middleware and Operating system
- (2) initial R&D for Grids software development technology
- (3) development of Grids Service Middleware as prototype
- (4) development of applications supported by (3) as demonstrators

Early Stages of FP7:

- (5) development of Grids Foundations Middleware for production use
- (6) development of Grids Operating system as prototype and then production quality
- (7) construction of interoperating software stacks from Operating system to the upper interface of Grids Service Middleware to support Grids applications as prototypes then production quality

8. CONCLUSIONS

The NGG2 expert group has further substantiated and widened the Next Generation Grid vision and research priorities, to support more general application requirements as well as to master the increasing complexity of the resources orchestrated by the Grid.

The main results of the NGG2 expert group are:

- Two new application scenarios confirming the need to shift the application focus of Grid from large science to business, industry and societal applications, and highlighting new research challenges to be addressed at both architectural and generic Grid application middleware levels.
- A gap analysis of the research scope of the Call 2 projects in order to assess the level of coverage of the research priorities required for realising the Next Generation Grid.
- A SWOT analysis of European Grid capabilities and research highlighting the strategic importance of Grids both as primary IT technology on which Europe can build its strengths, and as a key enabling tool for boosting productivity, competitiveness and growth of European economy.
- A new research focus on Network-Centric Grid Operating Systems as a new essential fabric layer to fully realise the Next Generation Grid vision.

The development of a coherent programme of research activities, complemented by industry-driven initiatives such as public-private partnerships in the areas described in this report, is considered essential to consolidate and further strengthen the European position in Grid technologies and improve European competitiveness.

9. ANNEX NGG2 MEMBERSHIP

9.1. Participants to NGG and NGG2

Expert	Country	Affiliation	NGG	NGG2
H. Bal	NL	Dept. Of Computer Science, Vrije Universiteit Amsterdam	Y	
S. Campadello	FI	NOKIA		Y
C. de Laat	NL	Faculty of Science, Informatics, University of Amsterdam	Y	
D. De Roure	UK	University of Southampton		Y
B. Farshchian	NO	TELENOR		Y
M. Fehse	DE	T-Systems		Y
C. Goble	UK	Dept. of Computer Science, University of Manchester		Y
Y. Guo	UK	Inforsense Ltd		Y
S. Haridi	SE	SICS	Y	
M. Hermenegildo	ES	Universidad Politecnica de Madrid		Y
K. Jeffery	UK	CCLRC-Rutherford Appleton Laboratory	Y	Y
J. Labarta	ES	Universitat Politecnica Catalunya	Y	
D. Laforenza	IT	ISTI-CNR	Y	Y
P. Maccallum	UK	Lion Bioscience AG	Y	Y
J. Masso	ES	Gridsystems	Y	Y
L. Matyska	CZ	Faculty of Informatics, Masaryk University	Y	
T. Priol	FR	INRIA-IRISA	Y	Y
A. Reinefeld	DE	Konrad-Zuse-Zentrum für Informationstechnik	Y	Y
A. Reuter	DE	European Media Lab GmbH	Y	
M. Riguiedel	FR	ENST	Y	Y
W.Schröder-Preikschat	DE	Friedrich-Alexander University Erlangen-Nuremberg		Y
D. Snelling	UK	Fujitsu Laboratories of Europe	Y	Y
D. Talia	IT	University of Calabria		Y
M. van Steen	NL	Dept. of Computer Science, Vrije Universiteit Amsterdam	Y	
T. A. Varvarigou	GR	National Technical University of Athens		Y

Chair: David Snelling; Editor-in-Chief: Keith Jeffery