Project no. FP6-004265

CoreGRID

European Research Network on Foundations, Software Infrastructures and Applications for large scale distributed, GRID and Peer-to-Peer Technologies

Network of Excellence

GRID-based Systems for solving complex problems

D.IRWM.03 – Roadmap version 2 on Grid Information, Resource and Workflow Monitoring Services

Due date of deliverable: February 28, 2006
Actual submission date: May 3rd, 2006

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

Organisation name of lead contractor for this deliverable:
Poznań Supercomputing and Networking Center

Ver. 1.2

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

Keyword List: Grid information systems, network monitoring, workflow, checkpointing, Grid accounting, virtual accounts, account management, Grid service

Table of content

1. Executive Summary

The CoreGRID – Network of Excellence – aims at building a virtual European-wide Research Laboratory that will achieve scientific and technological excellence in the domain of large scale distributed, Grid and Peer-to-Peer (P2P) technologies. The primary objective of the CoreGRID Network of Excellence is to build solid foundations for Grid and Peer-to-Peer both on a methodological basis and a technological basis, and to stay at the forefront of Excellence. This will be achieved by structuring research activities, leading to integrated research among experts from the relevant fields and, more specifically, distributed systems and middleware, programming models, algorithms, tools and environments.
The three main project objectives of the CoreGRID Network of Excellence are:

1) excellence,
2) integration & sustainability,
3) dissemination.

The Institute on Grid Information and Monitoring Services (WP5) has changed the name to **Grid Information, Resource and Workflow Monitoring Services (IRWM).** In addition the Institute left a few partners. The second version of the Joint Programme of Activities (JPA2) is defined by 11 partners from the following institutions :

- FHG - Fraunhofer Gesellschaft (Germany)
- FORTH - Institute of Computer Science, Foundation for Research and Technology (Greece)
- INFN - National Institute for Research in Nuclear Physics (Italy)
- MU - Masaryk University Brno (Czech R.)
- PSNC - Poznan Supercomputing and Networking Center (Poland)
- SZTAKI - Computer and Automation Research Institute, Hungarian Academy of Sciences (Hungary)
- UMUE - University of Muenster (Germany)
- UNICAL - University of Calabria (Italy)
- UCAM - University of Cambridge (UK)
- UNI DO - University of Dortmund (Germany)
- UOW - University of Westminster (UK).

The Institute on **Grid Information, Resource and Workflow Monitoring Services** was created to integrate a group of experts in the following Grid areas:

- **Information and Monitoring Services (task 5.1)**
- **Checkpointing Services (task 5.2)**
- **Workflow Services (task 5.3)**
- **Accounting and User Management Services. (task 5.4).**

The first version of the roadmap covered the first 18 months of the project, giving an overall description of the work to be done.

The IRWM Institute will focus on the following objectives:

- Providing multi-grain and dynamic monitoring for Grid resources and services
- Developing scalable Grid monitoring architecture with enhanced robustness and QoS guaranties
- Enabling reliable online monitoring of status and performance for a large range of resources
- Providing monitoring the progress of complex job workflows
- Support for extraction and representation of job workflows from programming models
- Realizing middleware support for complex job workflow execution
- Framework for user management and user and job separation
- Supporting accounting services
- Providing checkpoint restart functionality in heterogeneous environment supporting dynamic job migration
- Supporting kernel and application level checkpointing.

The roadmap also implements suggestions given by the experts during the annual review in December 2005 (Brussels) as well as the remarks of the second SAB (Scientific Advisory Board) in January 2006.

Main recommendations given in the annual review follow:

1. *Prototyping and work on mapping use cases and work-package results onto existing test-beds is crucial and should be intensified, obviously to be supported by the integration of additional projects' results linked to CoreGRID at a European and national level.*
2. *The vertical aspects of trust and security have to be networked with the different work-packages in an even closer way, either by adding a respective task to each WP or by having a more visible cross-activity in WP 1. The outcome in this field so far looks very promising; however, the lack of participation by some work-packages has to be resolved.*
3. *For each major technical deliverable, the relevant technical/research achievements have to be summarised in 3-5 pages.*
4. *The technical results of the projects should be made public (as white papers, for instance) as soon as they reach sufficient maturity in order to increase the project visibility and influence.*

The future work plan described in the second version of the roadmap provides for new prototype versions, e.g. in terms of checkpointing or user account management and accounting. An extra section devoted totrust and security has been added. A broadened summary of technical results will be put in the next deliverables.
Finally we plan a common workshop of the KDM, IRMW and RMS Institutes at the EuroPar 2006 conference, with internal CoreGRID presentations and some external presentations. The aim of the workshop is to achieve a critical mass in the areas we are analysing in the mentioned Institute [GMW, 2006].

2. Introduction

Context

The purpose of the CoreGRID Network of Excellence is to promote first class joint research on Grids and Peer to Peer systems. This broad research topic has been structured into six complementary and mutually interdependent research areas, each characterized by an important topic. These areas are structured by six CoreGRID Institutes, parts of the CoreGRID Research Laboratory:

- WP2: Knowledge & Data Management (KDM)
- WP3: Programming Model (PM)
- WP4: System Architecture (SA)
- WP5: Grid Information, Resource and Workflow Monitoring Services (IRWM)
- WP6: Resource Management and Scheduling (RMS)
- WP7: Systems, Tools, and Environments (STE).

As mentioned, the primary objective of the CoreGRID Network of Excellence (NoE) is to build solid foundations for Grid and Peer-to-Peer both on a methodological basis and a technological basis, and to stay at the forefront of Excellence. This will be achieved by structuring research activities, leading to integrated research among experts from the relevant fields and, more specifically, distributed systems and middleware, programming models, algorithms, tools and environments.

The roadmap of the IRWM Institute is presented in this document to help coordinate research and promote scientific expertise among its members. The services gather, transport and provide vital information about the Grid, its individual components and their combination, provides the workflow mechanism and information, supports the end user and administrator by defining multi-level checkpointing, and user account management. Finally, it delivers accounting information about the used resources.

These services are used by other Grid components and also by Grid users to make qualified decisions about the management and use of the Grid, and make it as efficient as possible. While often hidden and only indirectly accessed, the information and monitoring services provide an information infrastructure that is indispensable for all the other Grid components and users. Also, this infrastructure creates a specific Grid, which must be built and maintained in a similar way as the Grid directly accessible by end users.
Besides the integration of research already done by individual partners within this Institute, these services (information and monitoring, checkpointing, workflow, accounting and user management) are also essential to other Institutes 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. Strong collaboration is thus essential and we foresee mutual benefits resulting from the common work within the CoreGRID NoE.
This synergy is to be especially expected between the IRWM Institute and KDM, SA, RMS and STE Institutes, as all the developed tools and environments must become part of the monitored infrastructure and will provide data to be evaluated and fed into the resource management systems.

According to the general remarks after the first annual review in Brussels (December 2005), the Institute should focus on having real prototypes, and define its demands in terms of trust and security. A limited release prototypes are planned to be built, including integrated services used in IRWM. However the major implementation work will be done in other R&D projects both national and international. The IRWM results will not be deployed in the CoreGRID testbed, as primarily planned, but will be used in the national or international testbed provided by other projects. Some of the projects have been mentioned in the following section.

The work is organised similarly to the former roadmap proposition into Research Groups, units of two or more CoreGRID partners collaborating closely together on common goals.

The CoreGRID NoE has been integrating the research results in each Institute. The IRWM Institute defined several research groups in each of the four tasks.
The added value of IRWM Institute is a common architecture in terms of monitoring and information infrastructure, checkpointing, workflow services, user account management and accounting. The IRWM

partners are engaged in national and international projects, the major R&D results will be taken from the following projects:

- SZTAKI:
  - "Hungarian Supercomputing GRID" (No.: IKTA4-075),
  - "Cluster Programming Technology and Its Application in Meteorology" (No.: IKTA3-029),
  - "Chemistry Grid and its Application for Air Pollution Forecast" sponsored by the Ministry of Education (No.: IKTA5-137,
  - "Hungarian SuperCluster project" (No.:IKTA-00064/2003)
- PSNC:
  - Pionier national programme (www.pionier.gov.pl),
  - Progress ("Polish Research on Grid Environment for SUN Servers" project no. 6.T11.069.2001C/5688),
  - SGIGrid ("High Performance Computing and Visualization with the SGI Grid for Virtual Laboratory Applications", no. 6 T11 0052 2002 C/05836),
  - Clusterix (National Cluster of Linux Systems, 2003-2006)
  - General Architecture for Virtual Laboratories (national project funded by the Ministry of Education and Science, http://vlab.psnc.pl)
  - Baltic Grid (6 FP project, http://www.balticgrid.org)
  - EGEE (European Grid for E-Science, 6 FP project, http://www.egee.org)
- FHG:
  - K-Wf Grid ("Knowledge-based Workflow System for Grid Applications", STREP project),
  - Fraunhofer Resource Grid,
  - D-Grid (German Grid Initiative)
- FZJ:
  - UNICORE
  - OpenMolGRID
  - NextGRID
- MU:
  - MediGRID (national project),
- UoW:
  - UK EPSRC funded OGSA testbed project.
- INFN:
  - Grid.IT-WP3 (http://www.grid.it), a national project for the development of an advanced network infrastructure
  - DATATAG-WP4 (http://datatag.web.cern.ch), an European project to promote interoperability between Grids, supported the development of the GlueDomains prototype, a network monitoring architecture that incorporates active sensors and a configuration database that is used to control monitoring sessions.

The research results and experience taken from the national and international projects are a base of discussion within the research groups. The IRWM Institute partners have plans for releasing first prototypes and demonstrators. A detailed plan is described in the following section. An example of such a prototype is the checkpointing system of low level kernel and higher level checkpointing of PVM applications. The Network Monitoring research group is also planning some experimental activities.

Problem(s)

A large-scale heterogeneous Grid is subject to frequent changes and disruptions in the service of a huge number of components it is built of. To manage such a dynamic system and its resources, online monitoring of the resources to determine their availability is required. Without this information, the challenging goals set up in the report "Next Generation Grid(s), European Grid Research 2005-2010" (also known as the NGG report) [NGG2, NGG1] can not be achieved. While the monitoring is not extensively mentioned in the NGG report, it is an essential part of any Grid infrastructure, helping to hide its complexity. Current Grid information and monitoring frameworks have identifiable drawbacks, as they are either too focused on specific aspects or do not scale enough. The same information is often used in many places, at a different aggregation level, and allowing each component needing such information to collect it independently will impose on high overhead in the whole Grid. The centralized approach used currently for the monitoring and management of large scale Grids does not really scale beyond tens or at most hundreds of nodes/resources and is not sufficient when dealing with problems spanning several administrative domains. Despite the recent progress in this area, the level of architectures and

concepts (like the Grid Monitoring Architecture [SF,2001])is still either too high or there are too specific implementations, like MDS [CFF,2001], R-GMA [SF,2001], Mercury [BG,2003], or iGrid [ACE,2005], which are not yet able to provide a reliable information and monitoring system for a general heterogeneous large scale Grid. The most important design issues are shared with the SA Institute—namely the scalability, reliability and robustness of such a system. Additional problems lie in a blur line between information and monitoring systems (Is it possible to merge the two into one infrastructure?) and also in further processing of monitoring information it is not possible to collect all the data, so powerful filters, aggregation functions and distributed storage for most important logs must be developed. Also, in a web service-oriented world, what are the "natural" interfaces to the monitoring information and to which extent an active monitoring system based on probes and sensors could be complemented with a passive system based on the service and resource instrumentation.
The throughput of network lines between Grid nodes adds another complexity layer, as there may not even be enough processing power to work in the real time. In the network layer of the Grid, we would like to investigate services like accurate traffic monitoring and classification, scalability properties, performance aspects, provision of fault detection, dependability and security services for Grids.
Monitoring is not used only to detect failures or other types of problems. The collected data could be further processed to give an understanding of the overall Grid performance. We plan to develop new models and approaches that would provide metrics for Grid performance evaluation. These will be used, among others, to compare the influence of different Grid components on the overall behaviour of the Grid.

Monitoring the network infrastructure of a Grid has a vital role in the management and the utilization of the Grid itself. Network monitoring, as an instance of resource monitoring, is not targeted to system maintenance, but should produce observations that are used by brokers and other Grid agents in order to optimize Grid applications.
The challenge of network monitoring is that it must face a task which, in principle, grows with the square of the dimension of the system. Many subtasks inherit the same adverse characteristics: network load induced by active monitoring, query processing time, database size, etc. In addition, Network Monitoring is characterized by peculiar security problems, other than those connected with data protection. In fact, many monitoring tools require cooperation between two "peer" testing applications, which requires the existence of an appropriate authentication mechanism [AAC,2005], [TPP,2006], [AC,2005].
The are ongoing activities in a number of next generation network projects that intend to provide interfaces between the network layer and the Grid middleware, e.g. EU projects GEANT2, MUPBED [GN2], [MUPBED].

The checkpointing is needed to minimize effects of hardware and software failures on user jobs and to allow dynamic jobs migration. The contemporary operating systems were not designed to directly support checkpointing even on a single processor in network environments. As a result, the higher level models that harness the checkpointing mechanism are not well developed either. However, it is a very important research activity, as checkpointing not only increases efficiency, but in some cases it is the only way to get results in a prescribed timeframe even in the presence of failures. Despite numerous technological issues the new checkpointing implementations for different platforms still emerge, but their functionality is diverse and is not directly usable in the heterogeneous Grid environment. Additionally, due to technological and semantic issues, some limitations according to applications that can use checkpointing are imposed. To obtain the ability to utilize both existing and developed in the future low level checkpointing implementations, there is a need to define the alignment of that technology in the context of Grid environment. Additionally, the interfaces to the surrounding Grid environment and low level checkpointing mechanisms must be defined, including possibly direct interfaces with the monitoring infrastructure. To solve these issues, as a result of the work that has already been done within the CoreGRID NoE, the **Grid Checkpointing Architecture (GCA)** has been proposed [JKM,2005] [JJK,2005]. In the next eighteen months we will work on improving the architecture with the interface to the Broker and other external (from the point of view of GCA) services.

Complex job workflows represent another challenge, as the monitoring information must be almost synchronously gathered from many different sources and appropriately processed to provide a coherent view (state information) of the whole workflow and its components. The job workflow itself must be extracted from programming models, the monitoring and information services must be tightly coupled with job checkpointing and migration support to provide an environment where even complex job workflows could be easily deployed, executed, and monitored. Models and methods to provide a virtualized end user account system are a specific part of combined job flow support and information services.

The problem of managing user accounts becomes a non-trivial one in a distributed environment (especially in a Grid environment) that includes many independent sites and virtual organizations with hundreds or even

thousands of user accounts. The complexity rises from the point of view of time required for administration tasks and automatization of these tasks. The accounting issue becomes impossible in the distributed environment, but still possible on a single computer system [DMW,1999], [KLM,2001].

The next important problem is accounting in distributed systems. The existing solutions on the market allow the accounting of resources used for only one system or at most for local and homogeneous clusters [SGAS],[SNUPI],[DJMM,2005]. The problem arises in Grids when the environment is more complex by adding virtual organizations and dynamically assigned accounts. In the existing testbed installations the problem is most often neglected. However, the demand for a solution will significantly increase in the near future, especially in context of production Grids and Grid economy. The GGF created the Accounting Models Research Group, whose goal is to work out the rules of data exchange and a general communication interface between sites, allowing information to be collected about resources used in Grids. The problem becomes more complex the greater the environment is, i.e. the number of VOs and the number of users accessing resources. Obviously, the accounting mechanism should provide us with some information independently of the service scale and range (a single system, local cluster or Grid environment). The next important feature is the scalability and low time overheads generated by this mechanism. This is significant in a production environment characterized by high dynamism of changes. To ensure a real practical accounting mechanism for the Grid, it should have a decentralized, scalable and flexible structure. It should interfere with local domain policies as little as possible. The accounting should allow the data for a single user, group or the whole VO to be processed. The remote system tracks all information about resource usage by the user. The information is sent after the completion of the computations to the VO the remote user belongs to. To meet these requirements, distributed resource allocation accounting models, which would properly work at various sites with different administrations and resource management policies, would be necessary. As mentioned above, the possibility of accounting the resources used in Grids forms the basis for introducing the Grid economy concept.

Objectives

Several services are necessary for establishing a complex Grid architecture, including support for a resource management and scheduling system. All services must be designed to establish a fault-tolerant and flexible behaviour in a large-scale heterogeneous environment. This is impossible to achieve without relevant information about the state of services, properly collected, merged and filtered if necessary. Current models do not scale to the Grid level or are focused on specific aspects. The primary objective of the IRWM Institute is to study and provide general information and services for the underlying Grid management required by the “Next Generation Grid” [NGG2,NGG1]. The Grid management services considered here include Grid core services and components. The main goal of the project months 13-30 is to continue the integration of research results of several established IRWM research groups (RG) described in the chapter 4 (Vision, Strategy and Roadmap).

The IRWM Institute as defined in the JPA2 focuses on the following objectives:

- Providing multi-grain and dynamic monitoring for Grid resources and services
- Developing a scalable Grid monitoring architecture with enhanced robustness and QoS guaranties
- Enabling reliable online monitoring of status and performance for a large range of resources
- Providing monitoring of the progress of complex job workflows
- Support for extraction and representation of job workflows from programming models
- Realizing middleware support for complex job workflow execution
- Framework for user management and user and job separation
- Supporting accounting services
- User account management in production Grid environment
- Decreasing management overheads in production environment
- Providing checkpoint restart functionality in a heterogeneous environment supporting dynamic job migration
- Supporting kernel and application level checkpointing.

The **task 5.1** is focusing and its network monitoring RG is intented to provide a scalable architecture for network monitoring, trying to cope with its quadratic complexity: although we understand that this is intrinsic in its nature, we try to keep its growth under control, and provide ways to avoid resource saturation. The architecture should provide a clear interface to the outside, taking into consideration that the environment is unstable, due to progress in the design of Grid systems. Flexibility should be a primary concern in the design of the interface of the Network Monitoring Architecture with the rest of the Grid [CFK,2006], [CP,2006].

The main goal of **task 5.2** is to define the position of the checkpointing technology within the Grid environment and work out a concept and interfaces of the appropriate services. Up to now, as a result of the CoreGRID R&D effort, PSNC together with SZTAKI proposed a general view of the Grid Checkpointing Architecture (GCA). The proposition of interfaces is included in the GCA definition. The services, interfaces and location of the particular components of the GCA within the Grid environment are presented in "Towards Checkpointing Grid Architecture" (published in PPAM2005 conference proceedings) and "Grid Checkpointing Architecture - a revised proposal" (presented at the CoreGRID Integration Workshop 2005) papers.

The CoreGRID Network of Excellence has become a great opportunity for us to exchange knowledge, up-to-date achievements and proprietary (from each institution's point of view) projects results with other CoreGRID-involved institutes. Thanks to that, PSNC and SZTAKI established the Research Group which aims to integrate their checkpointing-related products. That integration will end up with the tool for checkpointing the PVM-based applications on a new platform (the actual target platform will be chosen after all analysis tasks finish, which has almost been done).

According to the main goal of the task 5.2, in the forthcoming eighteen months we plan to work on further development of GCA. The current state of the GCA is quite mature; nevertheless we want to make it even more integrated with the other Grid-related services (and especially with the metascheduler or broker). Therefore we are going to establish closer cooperation with the members of RMS Institute. As a result of that cooperation we expect to achieve the two following goals: (1) GCA with changes required integrating it with the scheduler, (2) and the concept of such integration (in the form of an appropriately described scenario). There is a possibility that such integration will end up with the necessity of adapting the GCA to other Grid subsystems. If such subsystems are recognized, then we will be forced to perform further changes in the GCA. When the process of refining the GCA finally finishes, the work on providing the proof-of-concept implementation of that architecture should begin. At the moment we cannot declare that the implementation effort will start within the nearest eighteen months but it is our intention. In the initial documents it was declared that the implementation of invented services will be based on OGSA specification. Because in the meantime new and widely acceptable specifications emerged, the implementation will probably be based on WSRF and WSN technologies instead of the planned OGSA. The final decision depends on the results of the aforementioned cooperation with other Grid subsystems providers (it will almost certainly be based on some Web Service-based service).

Simultaneously, the next months is the time when we want to finish all the analyses required to perform the integration of the SZTAKI's TCKPT and one of PSNC's checkpointing tools (the work is advanced) [JKM,2005],[TNC,2005]. Based on the results of those analyses (presented in technical reports) we will decide on what the target platform of the actual integration will be. The time required for that will also depend on the platform that we will choose but in general we plan to finish this integration by the next twelve months. When we finish the integration, the paper concluding the results will be released. As mentioned the GCA architecture will introduce the checkpoint restart functionality on kernel and application level. To show the usability of such an approach SZTAKI and PSNC will integrate products developed in their national projects, i.e. Intel IA-64 kernel checkpointing with TCKPT, which supports PVM applications. However PVM is not not used as frequent as the MPI library, the proposed Grid Checkpointing Architecture approach still remains general. It means the interfaces are universal and does not matter whether we use MPI or PVM for building up an pilot installation.

Thanks to CoreGRID Network of Excellence we also established a horizontal activity between WP5 and WP4 (specifically with UCO). The common parts of our interest are storage issues in context of requirements imposed by checkpointing services. Therefore in the nearest future we plan to define required storage functionality for distributed checkpointing purposes [SC,2005].

Another activity we plan is to have a closer collaboration with the RMS Institute on ressource management and scheduling, where checkpointing plays a crucial role in effective usage of Grid resources.

Another aspect of the research provided by this Institute is the study and development of services able to coordinate the reliable execution of vastly complex compound Grid jobs and realize middleware support for complex job workflow execution (**task 5.3**). This will also include an adequate description and modelling of workflows, mapping abstract onto concrete workflows and providing services for the monitoring of workflows, thus adding support for dynamic workflows on non-reliable Grid resources. Current state-of-the-art Grid workflow management solutions still show a lack of interoperability with other workflow systems, due to incompatible and informal workflow description languages. Therefore an important objective is to study and to overcome compatibility and conversion issues between commonly available workflow description languages, such as BPEL, Directed Acyclic Graphs, and Petri Nets.

Future Grids must support the management of complex jobs and service-level agreements. Those jobs have workflows with co-allocation constraints and dependencies that must be considered for a wide range of diverse resources. Present systems have architectural and design limitations that make them usable in a productive

manner only for simple workflows. This is not sufficient for highly complex Grid applications to be expected in important application domains such as industrial design, engineering, drug design and bioinformatics. The current limitations must be overcome by a common set of job management and execution services based on a powerful model.

The main aim of user management system is controlled, secure access to Grid resources (**task 5.4**). The considerations we try to introduce must take into account the fact that we are dealing with a production Grid environment which has the ability to change its configuration dynamically. Security requires authentication of the user and authorization based on combined security policy from the resource provider and virtual organization of the user. The second important thing is the possibility of logging user activities for accounting and security reasons and then gathering these data both by the resource provider and virtual organization of the user. From the user point of view, an important feature is single sign-on.

The problem of user management is a non-trivial one in an environment that includes a bulk number of computing resources, data, and hundreds or even thousands of users participating in lots of virtual organizations. The complexity rises from the point of view of time required for administration tasks and automation of these tasks. There are many solutions that attempt to fulfil these basic requirements and solve the mentioned problem, but none of them, according to our best knowledge, solve the problem in a complex and satisfactory way.

In this task we will focus on supporting full virtualization of user accounts on the heterogeneous Grid. Investigating approaches for real-time on-demand user account creation and management, supporting hierarchical VOs, with a possibility of user and job separation, correct data protection (including failure recovery) and accountability. The system must be neutral with respect to the actual job submission and authorization service used.

The overall objective of this Institute is to provide a critical mass of researchers to achieve excellent research results. This will also include continuous identification of weak areas and gaps in knowledge, where new research will be pointed out. . During the first period of the project the partners presented their previous work and explained concepts and approaches to each other. The results are a technical report and a paper describing best practices and requirements regarding user management and accounting issues in the Grid environment. This will be used as a base for further work – preparing specification of a set of services that will fulfill the requirements and a design of a model for full user accounts virtualization that will be described in a technical report and related publications. The proof of concept implementation of the services will be done. The implementation will base on integrated and enhanced existing systems developed by the partners. Finally, the implementation will be deployed in the pilot installation.

The forthcoming research and work will take into account the general and specific comments provided by the group of experts during the annual project review (December 2005, Brussels) and during the SAB meeting (January 2006).

Tasks

The IRWM Institute defined the following tasks in the JPA1 and JPA2:

- **Information and Monitoring Services (5.1)**
- **Checkpointing Services (5.2)**
- **Workflow Services (5.3)**
- **Accounting and User Management Services (5.4).**

The structure of the tasks did not change, but we changed the focus after the first project year. The first 12 months were dedicated for studying R&D work done by each partner. Research groups were defined within each tasks, delivering best practises, and the concept of integrated services architecture.
In general the forthcoming phase described by this roadmap will focus on prototyping and scalability. The general remark will also be taken into account to make the CoreGRID results more visible on the international arena.

Task 5.1 Information and Monitoring Services

The Information and Monitoring Services will concentrate on the following specific subtasks:

- Designing of the Grid modules that are in charge of providing network monitoring
- Describing the network monitoring Interfaces with the Grid environment
- Performing experimental activities.

Task 5.2 Checkpointing Services

The Checkpointing Service defined a set of the following specific subtasks:

- Preparing a paper describing the TCKPT (its architecture, assumptions, and the way it utilizes the third party low level checkpointer). The paper will help in understanding our future papers and reports concerning the planned TCKPT and PSNC's checkpointer integration. Thanks to that the potential end users will be able to utilize the new mechanism more consciously.
- Integration of SZTAKI's TCKPT with one of PSNC's checkpointers. Based on the results of anaysies performed in the previous months we will choose the target platform of our work and begin the implementation stage. The task is considered as a goal of the Research Group that is defined in section four.
- When the integration of TCKPT and PSNC's checkpointer finishes, then the next task will be to prepare and publish a paper describing the results. The intent of preparing this paper is also mentioned in the list of planned publications.
- Making GCA even more integrated with external Grid Services and especially with the Broker. By means of short visits, e-meetings and other forms of cooperation with the CoreGRID participants dealing with scheduling-related services, the GCA will be even more fitted into the overall Grid environment.
- When the interfaces to the Broker and external Grid services have already become more precisely specified, a paper describing the integration will be created.
- Definition of required storage functionality for distributed checkpointing purposes. This task will be performed in cooperation with UCO from WP4.

Task 5.3 Workflow Services

The main objectives of the Workflow Services task are the research and collaboration on the following topics:

- Services able to coordinate the reliable execution of vastly complex compound Grid jobs and realizing middleware support for complex job workflow execution
- User interfaces and Web portals for Grid workflow management
- Adequate description and modelling of workflows
- Services for monitoring of workflows
- Executing dynamic workflows on non-reliable Grid resources
- Mapping abstract onto concrete workflows
- Relation between Business and Grid workflows.

Therefore, task 5.3 will focus on the following specific subtasks:

- Maintenance of the Grid Workflow Forum (http://www.gridworkflow.org/) as a collaborative platform for public information exchange and discussions about scientific and commercial approaches in the domain of Grid Workflows.
- Compatibility and conversion of different Grid workflow description languages
- Formalisms for workflow description languages based on Petri Nets. Development of generic tools for visualization, monitoring, and composition of workflows, able to be used in various Grid environments among the CoreGRID partners (independent of Grid middleware).
- Joint research on workflow model checking. The objective is to study tools for automatically checking workflow models for consistency and specific properties, such as deadlocks, conflicts, etc. If necessary, a research group dedicated to find new ways of workflow model checking will be established.
- Joint research on workflow fault management, especially related to transaction safety in Grids (e.g. compensation transaction).
- Collaborative workflow-oriented portals related to collaborative workflow management.
- Research on the extension of different workflow developer and management tools with support for collaborative problem solving.
- Research on integrating Grid workflow and monitoring frameworks.

Parts of the roadmap of task 5.3 are strongly related to the other tasks of this Institute, e.g.:

- Information and monitoring services (Task 5.1) are required for mapping abstract workflows and for monitoring workflows.
- Checkpointing services (Task 5.2) can be used in order to achieve workflow checkpointing for workflow rollback (transactional safety) or migration of sub workflows.
- Accounting and User management Services (Task 5.4) are also important on the workflow level.

The activities of task 5.3 will have a strong interaction with Task 6.3 of the RMS Institute which covers scheduling strategies for Grid workflows.

Task 5.4 Accounting and User Management Services

The **Accounting and User Management Services** task focuses on the following short and long term activities:

- Preparing a detailed architecture of the proposed user management framework.
- Implementation of the above system.
- Analysis of requirements for an accounting gathering system.
- Definition of a minimum set of resources that the accounting system in a production Grid must gather and process.
- Definition of a database structure that will allow storing other and nonstandard accounting data.
- Design of a model architecture for the accounting system that will comply with different resources from different Grids and will take into account local and VOs policies.
- Implementation of a model system.

Drivers

The driving forces of the planned IRWM work is to introduce necessary services which would allow to release a production Grid environment. The vision described by a group of experts in NGG and NGG2 reports gives [NGG1, NGG2] key research priorities in terms of properties, facility and models, cit.:

> *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 de****mand 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;*

The main goal of R&D to be done in the future is to have a **dynamic, scalable and reconfigurable** environment which would meet the requirements of industry and scientific communities.

One of the most promising techniques to attain the planned results is passive network monitoring. FORTH is a leading research institute in the field, and will contribute with its specific know how. INFN has designed and deployed a Network Monitoring system, and can contribute with a clear vision of the overall problem.

The checkpointing mechanism is a crucial one to attain the proper quality of fault-tolerane and load-balancing features in distributed computing environment (especially in production based systems). Nevertheless the nowadays computing environments and platforms are lacking in a wide availability of that technology. The checkpointing technology is available only for a few hardware and OS platforms and if it is then it can be used in limited way (therefore PSNC since a few years performs R&D effort to extend the availability and functionality of checkpointing technology). The conscious and coherent utilization of checkpointing capabilities in Grid environments is even less developed. The GGF woks out on the model of self-checkpointed, Grid-aware applications but there is still lack of general architecture that could allow abstracting the idea of checkpointing utilization (especially legacy implementations of that technology) in Grids. Then the task 5.2 is working on the Grid Checkpointing Architecture that will allow on integration of different checkpointing implementations with Grid environment. Additionally thanks to that architecture the checkpointing will be utilized in conscious and coherent way. The broker or scheduler systems will be aware of the possibility of emploing the checkpointing functionality. The interface to that functionality will be unified from the scheduler / broker viepoint and the low-level checkpointing system implementations will be seen as resources, so that they will be searchable and able to expose their functionality to upper layers of the Grid environment. The final result of the work done by the task 5.2 will be proof-of-concept implementation of the Grid Checkpointing Architecture. As the architecture has been invented as modular one the third party develpers will have the opportunity to integrate their own (already existing and future) checkpointing implementations with the proof-of-concept Grid Checkpointing Architecture.

There are number of tools supporting user management, authorization and accounting, like e.g. PSNC Virtual User System. The knowledge and expirience in the area will rsult in proposition of system that wil fulfil the modern requirements.

3. Positioning

State of the art

This section outlines the existing approaches and their problems, tradeoffs, and limitations. It is structured according to the topics covered by the IRWM Institute.

Monitoring and Information Services

Scalability issues are seldom taken into account when considering Network Monitoring. There is evidence that they can compromise the availability of Network Resources observations. Our primary concern is scalability: to this end, we propose a Network Monitoring Architecture which combines several known concepts: passive monitoring techniques, a domain-oriented overlay network, and attitude for demand-driven monitoring sessions. In order to keep into account the demand of extreme scalability, we introduce solutions to two problems that are inherent to the proposed approach, and well documented in the literature: security and group membership maintenance [BKT, 2006], [BKK,2005].

Checkpointing Services

When the first roadmap was written, the most acceptable specification defining how to write Grid Services was OGSA. In the meantime new and widely acceptable specifications emerged. The new solutions are based on WSRF and WSN technologies instead of OGSA.

Low-level Checkpointing Mechanism

There are three methods of implementing the checkpoint mechanism: the kernel level (operating system), the user level, and the application level. The mechanism included in the operating system takes care of storing the job image. In this case checkpointing is transparent for the applications, which means that the user does not need to modify or extend his/her applications. Using checkpointing on the user level is less convenient. The user must recompile his/her application and link it with additional checkpointing libraries. Most of the libraries periodically store the job's state. On this level the checkpointing is transparent for the applications but it is onerous for the user because of the necessity to link with additional libraries. The third method (the application level checkpointing) is even more complicated. The user must implement the checkpoint and the restart mechanism in his/her application by himself/herself.

Due to historical reasons the checkpointing mechanism is not widely available. When modern operating systems were designed, the checkpointing facility was not considered. Consequently it is hard to provide these operating systems with the checkpointing mechanism. Only few operating systems have been built in this checkpointing (IRIX, UNICOS MK). Additionally there are a few projects that provide checkpointing packages for other platforms. The links to these projects are collected on the http://www.checkpointing.org site. In the aftermath of the aforementioned historical reasons, all available checkpointing mechanisms impose some limitations on programs that are to be checkpointed. However, the new checkpointing packages for new platforms with improved functionality are still developed. For instance, PSNC has developed three checkpointing packages. These packages can be downloaded from the [CKPT] page.

Checkpointing in the context of resource management software and clusters

Some of clustering or resource management systems are able to utilize the low level checkpointing mechanisms. For example, the Sun Grid Engine (SGE) has an interface that allows using the checkpointing facility on nodes which posses this functionality. Further, the Condor project is shipped with a proprietary user-level checkpointing library and allows to submit jobs that are to be checkpointable.

An additional challenge is to checkpoint distributed applications based on the PVM/MPI communication model. The package that addresses that area is the PGRADE development framework developed by MTA SZTAKI which allows to checkpoint the PVM and MPI application based on the low level checkpointing mechanism like e.g. Condor [JJK,2005].

Checkpointing in the context of Grid environment

Currently the Grid virtually lacks in well-developed checkpointing-related services and interfaces. As this problem has been noticed by the Global Grid Forum, they established the Grid Checkpoint Recovery (GridCPR) Working Group. The group aims to define the user-level API and associated layer of services. The home page of that group is [GRIDCPR].

Regarding the checkpointing technology itself, comparing with the state of the art presented in the previous roadmap and in the above section, there are no new facts expect the Grid Checkpointing Architecture (GCA) that has been worked out within the CoreGRID NoE. The GCA defines the high-level checkpoint-restart Grid Service and locates it among other Grid Services. It encompasses both, the abstract model of that service and the lower layer interface that will allow the service to cooperate with the diverse existing and future checkpoint-restart tools. The next task is to make the GCA even more integrated with the existing Broker and other external Grid Services from the point of view of GCA .

Workflow Services

Using workflow has gained interest in the Grid community because it provides a high-level abstraction for application composition. The application of workflow for the composition of Web and Grid services has resulted in several frameworks. In most frameworks the user has to describe his or her workflows by means of a more or less formalized workflow description language, which is then passed to a workflow engine. The workflow engine parses the workflow description and maps it onto the available lower-level Grid middleware, synchronizing the remote method invocations and coordinating the data transfer from one service to the other. One drawback of commonly available Grid workflow systems is that they do not support interaction between the workflow orchestration, composition and execution processes.
Due to the dynamic nature of the Grid, workflow composition is a challenging task because the system has to deal with resource unreliability and unpredictability, which are closely related to fault tolerance and scheduling. Unfortunately, these issues have not been addressed to by most of the related work on Grid-based workflows. Most of this work assumes no faults on the Grid and poor or no information provided to the system to deal with Grid dynamics.

Workflow Description Languages

Currently there exists a broad spectrum of different approaches for describing workflows in Grid computing environments, without any established standard that is commonly accepted by a majority of the communities. The existing workflow description languages can roughly be grouped into two classes: Script-like workflow descriptions specify the workflow by means of a textual “programming language” that often possesses complex semantics and an extensive syntax, while graph-based workflow description languages specify the workflow with only few basic graph elements. Examples of script-based workflow descriptions are GRIDAnt and BPEL4WS. These languages explicitly contain a set of specific workflow constructs, such as a sequence or while/do, in order to build up the workflow. Purely graph-based workflow descriptions have been proposed (e.g. for Symphony, Condor’s DAGman tool, and Grid Workflow Execution Service – GWES) which are mostly based on Directed Acyclic Graphs (DAGs) or Petri Nets. Compared to script-based descriptions, graphs are easier to use and more intuitive for the unskilled user: communications between different services are represented as arcs going from one service to another. Another commonly used script-based approach to describe workflows is the Business Process Execution Language (BPEL) and its recent version for Web Services (BPEL4WS) that builds on IBM’s WSFL (Web Services Flow Language) and Microsoft’s XLANG (Web Services for Business Process Design). BPEL [BPEL] is intended mainly for modelling and implementing business processes and possesses complex and rather informal semantics, which makes it more difficult to use formal analysis methods and to model scientific workflows especially for the unskilled end user. Further comparisons of workflow description formalisms are available at the Grid Workflow Forum (http://www.gridworkflow.org/), which has been set up by CoreGRID partners.

Fault tolerance in workflow execution

Fault tolerant workflow execution needs exact and correct information that describes the current states of Grid resources. Based on this information intelligent workflow management services can make predictions on the performance and availability of resources, and perform optimal and automatic resource allocation, de-allocation and job migration operations. Unfortunately the information systems of current production Grids poorly satisfy this need. Resource providers publish only very basic information about their capabilities and refresh such information with low frequencies. In order to provide workflow manager and scheduler systems with up to date information, dynamic resource testing and monitoring systems are required. These systems should gather and aggregate resource availability information and forward them to appropriate consumers. There are several research efforts worldwide to develop specific or more generic Grid monitoring and resource testing solutions. At the same time there are no workflow manager or scheduler systems that would be capable of using information coming from these tools.

The Monitoring and Discovery System (MDS) is part of the Globus distribution. The latest version (4.x) of the Globus Toolkit based on the Web Services Resource Framework (WSRF) has introduced MDS4 which provides extensible query and subscription/notification interfaces. These interfaces allow to query WSRF services resource properties, execute external scripts and applications to acquire data about resources, and trigger actions in response to certain user-defined conditions. The MDS is capable of integrating different monitoring solutions under the same indexing service.

Inca is a generic framework for automated testing, verification, and monitoring of functionality common to a set of Grid systems. The server acts as a collector of information regarding the monitored nodes, while the clients are capable of executing predefined tests at configurable scheduled times reporting the results to the server. The main advantage of Inca is undoubtedly its fairly complete test coverage. However, current Inca tests target GT2 and GT3 but do not provide support for GT4-based Grid systems [GLOBUS].

GRASP (Grid Assessment Probes) is a set of three composite tests which verify and measure performance of basic Grid functions including file transfers, remote execution, and Grid information services response [GRASP]. The probes involve moving different files of various sizes to one or more machines, perform some sort of computation on them and move the results back. The scripts perform very good assessment of the available bandwidth, besides verifying the functionalities of three fundamental Globus Toolkit services. However, the probes are designed and built to be run as standalone scripts which return their output to a console, and there is no integration with the Globus Toolkit at the moment. Moreover, the scripts target version 3 of the Globus Toolkit only. Integration with the MDS4, after some modification to the test scripts code, is a possibility using an aggregator execution source.

Semantically enhanced service-oriented Workflows

As the complexity of the Grid applications grows, growing attention is being paid to the automatic methods of workflow composition, management and monitoring. As workflows may be complex structures composed of various processes (services), for automated (or at least computer aided) workflow composition, more than a syntactic description of data passed in the workflow is required. One of the possible solutions explored recently is the introduction of a semantic Grid concept. The semantic Grid extends the notion of a Grid service with a description of the service semantics, similarly to the semantic description of web pages in a semantic web. Semantic descriptions of services may cover various aspects of their behavior, such as QoS-related characteristics, transactional behavior (as studied by πt-calculus), security related characteristics, execution contracts (preconditions and postconditions), etc. Such a semantic description is useful for implementing automated workflow composition, assuming that a (scalable) repository of services containing their syntactic and semantic description is available. The goal of the SEAGRIN project is to explore the possibilities of implementing service-oriented workflow infrastructure based on the semantic Grid concept, with features such as automated workflow composition and management. Development of such infrastructure requires careful evaluation of semantic description techniques, possibly devising a new technique suitable for Grid services and workflows. The implementation mechanisms for the abovementioned semantic repository must also be studied. Also implementation of the existing workflow infrastructures for the Grid must be studied (P-GRADE) to examine whether they could be extended with the described features.

User account management

The problem of the single sign is a very important one and two kinds of solutions can be considered. Information connected with user accounts can be replicated to other hosts or user accounts can be assigned dynamically. The first solution is more popular, because it has been more suitable to the local clusters and local networks of computers.

A very popular solution was to use NIS (Network Information System), which allows different machines to access the same database with user accounts. More secure solutions use Kerberos authentication to access remote hosts. The Distributed Common Environment uses Kerberos to authenticate users across the DCE cell. Also, some distributed file systems like DFS or AFS employ Kerberos authentication to access files at a remote location. All these solutions were designed for local networks and do not solve the problem of Grids with hundreds of thousands of users.
The most common system to implement a Grid environment nowadays is Globus. In the Globus environment users are authenticated with X.509 certificates. Logging into the Globus environment only involves creation of an X.509 proxy, and this proxy works on behalf of the user without any need for further authorization. However, to use any Globus resources the users must be included in the Grid-mapfile on remote hosts, where the resources are located. The Grid-mapfile is checked by the Globus-gatekeeper to find the local Unix account, which should be used to run processes on behalf of the Globus user ID. Users can have individual accounts or many users can be mapped to the same Unix account. Some enhancements for Globus authorization were made by the DataGrid project. DataGrid is based on the idea of virtual organizations. Each virtual organization (VO) runs a dedicated server that maintains certificates of all people that are working on the same experiments. Each computing machine downloads a list of authorized users from all approved virtual organization servers. Then these users are added to the Grid-mapfile and are mapped to the pool of accounts dedicated to this VO. The accounts from this pool are assigned to consecutive users when needed, but are not recycled or reused. Hence, all users have their own accounts on all the machines they use, but they do not need to apply for them. All the users need to do is to enter any approved VO.
Several Grid environments employ the idea of virtual accounts that are temporarily mapped to a pool of physical accounts when needed. Such accounts are called virtual, scratch, generic, template or shadow accounts. In PUNCH users have their own logical user-accounts and the system manages its own physical accounts on remote resources and dynamically recycles them among users as necessary. Generic accounts can also be used in Condor and Legion, but in both systems it is recommended that users have their own accounts on every machine. Condor uses a nobody UID to run jobs for users that do not have an account in a Condor flock. Legion also manages the pool of generic accounts that are assigned for Legion use. To the best of our knowledge, none of the systems using a pool of virtual accounts allows for full accounting of resources used. Some work is underway to define the mechanisms of distributed accounting on the Grid.

Accounting

There are four research groups within GGF currently working on accounting-related standards. The Grid Resource Allocation Agreement Protocol (GRAAP) working group is currently specifying a Service Level Agreement (SLA)-based protocol called Agreement-based Service Management (WS-Agreement). The Resource Usage Service (RUS) is a service that can be used to publish and query resource usage data. It heavily relies on the Usage Record (UR) format, which is a standard XML document composed of various usage properties. The Grid Economic Services Architecture (GESA) essentially extends the OGSI Grid service model into an economic service model, where you can charge consumers for service usage [GESA]. In order to achieve this, GESA defines an architecture based on OGSI-Agreement contracts, and Resource Usage services. This combines all accounting-related efforts within GGF into a common model.

The DataGrid project implemented the DGAS component for scheduling jobs so that resources are fully utilized to the lowest possible price [DGAS1, DGAS2]. Each user has an account in a local bank called the Home Location Registry (HLR). When a job is submitted by the user, the resource broker receiving the request contacts a pricing authority at various resources and the local bank to check whether there are sufficient funds to run the job. If the bank grants the transaction, the request is passed to the job controller which sends it to an available resource that matched the user requirements. A resource monitoring service then tracks the job status and the resource usage and sends periodic reports to the HLR. When the job completes, the total cost is calculated and possible holds on amounts in the HLR are unlocked and the credits spent are withdrawn from the user HLR and deposited into an account in the resource HLR.

SweGrid (Swedish Grid) is preparing the system (SGAS) for allocating resources to project and account used resources [SGAS]. The system bases on and extends GGF standards.
Another approach is SNUPI, the System, Network, Usage and Performance Interface which provides an interface for resource utilization reporting for heterogeneous computer systems, including Linux clusters. SNUPI provides data collection tools, recommended RDBMS schema design, and Perl-DBI scripts suitable for portal services to deliver reports at the system, user, and job for heterogeneous systems across the enterprise, including Linux clusters [SNUPI].

Extended context

Network Monitoring Services

The monitoring services have been addressed by the research group including INFN and FORTH.
Grid.IT-WP3 (http://www.grid.it) is a national project for the development of an advanced network infrastructure.
DATATAG-WP4 (http://datatag.web.cern.ch) is an European project to promote interoperability between Grids, supported the development of the GlueDomains prototype, a network monitoring architecture that incorporates active sensors and a configuration database that is used to control monitoring sessions.

The network monitoring architecture is deeply concerned with scalability issues, and we have considered (INFN-FORTH research group) several facets of such issue.

The INFN-FORTH research group has been addressing the scalability problem at several levels [CP,2006]:

- we propose an extensive use of passive monitoring techniques, which eliminates the problem of traffic induced by monitoring sessions;
- we introduce an overlay topology which should simplify the topology of the system, thus reducing the number of paths under test;
- we provide an interface for on demand monitoring sessions. Such sessions in principle do not suffer of scalability problems, since their number is linearly bound to running applications.

While the first two alternatives are currently available, the third one depends on the existence of Grid-aware applications. When such technology will become widely applied, the scalability problems bound to the monitoring footprint will be eliminated.

Another kind of scalability problem is bound to the existence of a shared knowledge, here represented by the overlay topology. The management of such shared knowledge is a potential scalability limit, if based on any kind of centralized control. Therefore we have defined a distributed algorithm that keeps reasonably updated an instance of the whole database, describing the composition of the overlay domain partitions, on each Network Monitoring Element. The algorithm, as well as the size of the database, has a complexity that linearly increases with the size of the system, and is based on a random gossip scheme. The latency of an update, which amounts to minutes in a system of 1000 domains, according to simulation experiments, is adequate for the application. But the deployment of optical networks might drasically cut these figures [AC,2005].

Workflow Services

There is no complete environment covering all the above-mentioned topics. Instead, there are a lot of projects related to the specific Grid infrastructure or resulting from specific local needs. Task 5.3 partners are involved in a number of projects dealing with workflows.

K-Wf Grid is a STREP project funded by EC that started in 09/2004 under the leadership of Fraunhofer FIRST with the objective to develop a "Knowledge-based Workflow System for Grid Applications". The K-Wf Grid system will assist its users in composing powerful Grid workflows by means of an expert system. All interactions with the Grid environment are monitored and evaluated. The knowledge about the Grid itself is mined and reused in the process of workflow construction, service selection and Grid behaviour prediction.

The OPENMOLGRID project will provide a unified and extensible information-rich environment for solving molecular design and engineering tasks. The project contains the high-level workflow description & processing in UNICORE [UNICORE], [D-GRID].

The goal of the NextGRID project is to develop next-generation architectural components enabling a more economically viable use of the current Grid infrastructure for business and research as well as the general public. This project will aim to work out a universal model to integrate resources in workflow processing.

The GGF Grid Scheduling Architecture Research Group was set up to standardize workflow efforts. The goal of this research group is to define a scheduling architecture that supports cooperation between different scheduling instances for arbitrary Grid resources. The considered resources include network, software, data, storage and processing units. The research group will particularly address the interaction between resource management and data management. Co-allocation and the reservation of resources are key aspects of the new scheduling architecture which will also include the integration of user- or provider-defined scheduling policies.

Another workflow-related project is the Ganga Grid user interface which will be deployed on the LHC Computing Grid to manage the analysis jobs for high energy physics. It is a front-end for the configuration, submission, monitoring, bookkeeping, output collection, and reporting of computing jobs run on a local batch system or on the Grid. In particular, Ganga handles jobs that use applications written for the gaudi software framework shared by the Atlas and LHCb experiments. Ganga exploits the commonality of gaudi-based computing jobs, while insulating against Grid-, batch- and framework-specific technicalities, to maximize end-user productivity in defining, configuring, and executing jobs.

UOW implemented an environment where users can construct, manage and execute visual workflows from a Grid portal in order to support the execution of Grid-enabled Legacy Codes. The middleware called GEMLCA (www.cpc.wmin.ac.uk/gemlca) has been developed by the University of Westminster during the UK EPSRC funded OGSA testbed project. GEMLCA uses Globus Toolkit 3 and is being migrated to GT4 (WSRF). GEMLCA has been demonstrated at IST 2004 in Hagues and there is an article about it in the 'Next Generation Grids' publication document.

User Account Management and Accounting Services

PSNC has been performing research on user accounts management and distributed accounting information system for several years. The project name is the Virtual Management System (VUS).

The idea of VUS is universal and can be implemented with different job management systems. VUS is just an extension of the system that runs users' jobs (e.g. queuing system, Globus Gatekeeper, etc.) and allows running jobs without having a user account on a node. This allows minimizing overhead related to creating and maintaining additional user accounts. On the contrary to other solutions, VUS assures an accurate security level achieved by user authorization and possibility of charging the user with costs of resource usage. Additionally, it respects the local policy of sites and makes it possible for the local administrator to differentiate between local and remote users. The system is transparent for the users and to some extent also for administrators. Users just run their jobs and do not care about account assignment. Administrators just configure access to a machine for VO and do not need to create accounts and manage grid-mapfile. The system does not interfere with local access policy either.

The history of VUS started in 1998 [DMW,1999],[KLM,2001]. The first implementation of VUS was an extension to queuing systems (e.g. LSF) and it was successfully exploited 3 years ago in the Polish national cluster which connected several HPC centres in Poland [KLW,2001],[MW,2003],[JWM,2004],[DJK,2005]. Later we focused on Globus. The first version of VUS for Globus 2.4 was using a modified Gatekeeper [DJM,2005]. The current implementation is GRAM "callout", a mechanism introduced in GT 3.2. The major advantage of callouts is that there is no need to modify Globus Toolkit (GT) codes to install VUS (the older versions of GT needed slight modifications in Gatekeeper and GridFTP).

The research on account management and distributed accounting processing is performed in several projects.

- National Computing Grid – set up by PSNC as an attempt to connect supercomputers from different Polish supercomputing centers with the LSF queuing system.

- SGIgrid - "High Performance Computing and Visualization with the SGI Grid for Virtual Laboratory Applications". The goal is to design and implement an environment for remote access to unique and therefore expensive laboratory equipments. The environment will support the end user by delivering enough computational power for pre- and post-processing computations and data intensive visualization. One of the project goals is to provide a backup data center for the national Institute of Meteorology and Water Management.

- Clusterix – "National CLUSTER of LInuX Systems". The goal of the project is to build a new generation distributed PC cluster by connecting 64bit PC from 12 Polish computing centers. Middleware implemented in this project will allow a test application to run in a dynamic environment with a changing number and configuration of computing and network infrastructure.

- BalticGrid - The goal of the BalticGrid project is to extend the European Grid by integrating new partners from the Baltic States (Lithuania, Latvia and Estonia) in the European Grid research community and to foster the development of the Grid infrastructure in these countries. To this end, the BalticGrid consortium has enlisted the help of experienced EU Grid computing centers whose aim will be to guide our new partners through the process of deploying Grid resources and applications at their respective institutions.

Checkpointing Services

In the recent years, PSNC has been very active in the area of R&D issues concerning the checkpointing technology. Interest in that area resulted in the three checkpointing packages:

- PsncLibCkpt - the user-level library that provides the checkpointing functionality for Solaris 8 OS. The core part of the psncLibCkpt is based on the libCkpt library. The package has been developed as part of the national PROGRESS project co-funded by Sun Microsystems and the Ministry of Education and Science.
- psncC/R - kernel-level checkpointing package. The product is on Solaris 8 and 9, developed for UltraSparc CPU.
- AltixC/R – a kernel-level checkpointing package designed for Altix systems equipped with Intel IA64 processors and running under the ProPack environment (a Linux-based environment prepared by SGI). Currently we have versions of our package that works with ProPack based on Linux kernel 2.4 as well as with a more recent ProPack that is based on Linux kernel 2.6. The development was done in a national project called SGIgrid, funded by SGI and Ministry of Education and Science.

MTA SZTAKI aimed at making parallel – GRAPNEL, PVM, MPI – applications capable of being checkpointed and restarted on clusters. GRAPNEL is a graphical language to define parallel applications in the PGRADE development framework. Until now MTA SZTAKI has been working on the following checkpoint issues:

- User-level, automatic, binary checkpoint for GRAPNEL applications
- Automatic checkpoint/restart mechanism for GRAPNEL applications under Condor scheduler (without any modifications in the scheduler)
- User-level, automatic, binary checkpoint for native PVM applications
- User-level, automatic, binary checkpoint for MPI applications

MTA SZTAKI had several local projects in the last 3-4 years. They were about to sponsor the research and development regarding the above mentioned areas. These are the “Hungarian Supercomputing GRID” sponsored by the Ministry of Education (IKTA4-075), “Cluster Programming Technology and Its Application in Meteorology” sponsored by the Ministry of Education (IKTA3-029), “Chemistry GRID and its Application for Air Pollution Forecast” sponsored by the Ministry of Education (IKTA5-137) and the currently active project is the “Hungarian SuperCluster project” sponsored by the National Office for Research and Technology (IKTA-00064/2003) which aims at giving parallel checkpoint support for the nationwide Hungarian Clustergrid infrastructure which is the largest cluster-based Grid in Hungary.

4. Vision, Strategy and Roadmap

Vision and Scenarios (end-users, technologies, computer science)

The IRWM Institute is one of the 6 Institutes which will focus on working out the knowledge and excellence on various disciplines of Grids and P2P technologies.

> ”The Grid concept demands novel approaches to system design and management – and thus to the operating system behaviour, middleware requirements and services offered to applications. *A three-level conceptual architecture is emerging. 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.*
>
> *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 the existing operating systems with Grids Foundations Middleware to provide the required functionality.”*
> – Next Generation Grid report [NGG1, NGG2]

The IST group of experts identified in the NGG and NGG2 reports some crucial features which will allow migration from the 'scientific' Grid to the production Grid, allowing to introduce computing on demand and make the Grid suitable for industry challenges.
Some of the features are mentioned below:

"In addition to the properties mentioned in the NGG1 Report, the 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 the 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."*

The vision of the IRWM Institute is coherent to the main outcome of the NGG document, i.e. provides the Grid middleware layer with services allowing stable migration to industry-based Grids with a reliable behaviour, secure, pervasive and scalable, being able to make self-reconfiguration.

All services must be designed to establish a fault-tolerant and flexible behaviour in a large-scale heterogeneous environment. It is impossible to achieve that without relevant information about the state of services, properly collected, merged and filtered if necessary. Current models do not scale to the Grid level or are focused on specific aspects. The primary objective of the Institute is to study and provide general information and monitoring service for the underlying Grid management required by the Next Generation Grid. The Grid management services considered here include Grid middleware (core) services and components as well as higher-level services and components on top of the Grid middleware.

Therefore, IRWM scope of tasks will focus on the following objectives:
- Providing multi-grain and dynamic monitoring for Grid resources and services
- Developing scalable Grid monitoring architecture with enhanced robustness and QoS guaranties
- Enabling reliable online monitoring of status and performance for a large range of resources
- Providing monitoring of the progress of complex job workflows
- Support for extraction and representation of job workflows from programming models
- Realizing middleware support for complex job workflow execution
- Framework for user management and user and job separation
- Supporting accounting services
- Providing checkpoint restart functionality in a heterogeneous environment supporting dynamic job migration
- Supporting kernel and application level checkpointing.

The new objectives defined in the JPA2 programme are the drivers of the future development vision, which at this stage is focusing on release prototypes in co-operation with other tasks and Institutes. These are the crucial issues for the second phase of IRWM research.
The availability of connectivity measurements is an issue that is a precondition for many features that are of interest in a Grid environment. Disregarding applications that come immediately to mind, like optimization of replica utilization, there are other aspects that are bound to the availability of such information, for instance job migration, which can be required by a recovery activity, and workflow analysis. Such applications are the end users of a **Network Monitoring Architecture (task 5.1)**. Unless this functionality is implemented in a scalable way, the applications that should take advantage of it will (and in fact are) simply ignore the presence of network limits, and therefore incur in performance degradation.

Introduction of the **checkpointing service (task 5.2)** to the Grid environment will affect the Grid as a whole. Both end user and the Grid itself will benefit from this service. The user can expect that the job that is running for a really long period of time will not have to repeat the whole computation once again because of some sort of failure. From the scheduler point of view, checkpointing and migration allows to use dynamic scheduling algorithms because application is no longer required to run from start to end on the same node. Implementation of checkpointing and migration is justified from the economical point of view as it allows better utilisation of the Grid resources.

The problem of **user management (task 5.4)** is a non-trivial one in an environment, that includes bulk number of computing resources, data, and hundreds or even thousands of users participating in lots of virtual organizations. The complexity rises from the point of view of time required for administration tasks and automation of these tasks. The existing solutions are still not satisfactory. Also the **accounting** services become more and more important for scientific applications and will be crucial for commercial Grids, but they are in their early stage.

Strategy

The strategy on development agreed at the first stage (roadmap version 1) includes the following actions:
- Definition of short-term research groups within each task.
  The research groups focused on several goals, defining the subjects of research and short term milestones with clearly defined outcomes.
- Clearly defined milestones and deliverables of the IRWM Institute:
  - The milestones and deliverables take into account the integration work of research groups
  - After providing common solutions within tasks, co-operation between tasks has been foreseen to integrate a joint architecture of this Institute
  - WP5 will finally provide the architecture which is planned to be used by other WPs. The IRWM Institute will use the results of other CoreGRID Institutes.

The research groups (RG) may describe short- or long-term cooperation between partners. It means we can have several RG during the project lifetime. The long-term cooperation should result in prototypes.
The outcome of the strategic view of WP5 is a list of deliverables and milestones put in JPA2 programme.

Several general WP5 meetings have been organized since the beginning of the project:
- The CoreGRID Kick-off Meeting, 13–14 September, 2004, Charleroi (BE)
  During the first CoreGRID meeting we organised the first WP5 meeting in order to present all WP5 tasks and establish first steps towards the creation of research groups. All partners presented their possible contribution to WP5. This was also an opportunity to meet other partners and define possible collaboration
- The First official CoreGRID WP5 Meeting, January 19, 2005, Crete (Greece)
  At this meeting the WP5 partners presented their research activities carried out. This led to the first approach towards the creation of research groups defined after the kick-off meeting. A plan of short visits between partners was also prepared during this meeting.
- The second CoreGRID WP5 meeting, July 15, 2005, Barcelona (Spain)
  Presentation of research work done in each task. A new call for making new RG in terms of workflow services was announced.
- VRVS teleconference of WP5 partners, December 2005
  The state of integration work done in each research group was discussed.

The strategy of the second project time period includes the following general activities:
- Presentation of the results and work done by each partner in R&D projects in the scope of WP5 activities
- Integrating the R&D results
- Working out a common architecture of the services
- Making prototypes
- Integrating services and interface into a consistent WP5 architecture
- Defining requirements to other tasks
- Starting the process of inter tasks cooperation
- Integrating results between Institutes.

The mentioned activities will be implemented by WP5 meetings, short visits between partners, longer visits (fellow programs), joint CoreGRID reports, publications and thematic workshops.
Common demonstrators have been planned as a proof of the concept.

Roadmap

The roadmap is defined by 4 deliverables in the JPA2:

- D.IRWM.03: Roadmap version 2 on Grid Information, Resource and Workflow Monitoring Services (M18)
- D.IRWM.04: Integrated framework architecture the Grid Information, Resource and Workflow Monitoring Services (M20)
- D.IRWM.05: Report about the results on the joint research topics of all research groups, including implementation of prototypes (M25)
- D.IRWM.06: Report on the integration of WP5 services with other architectures developed by CoreGRID Institutes (M30).

Also, the following milestones of the research and development work up to the project month 30 have been planned:

- M.IRWM.02: Integration of workflow service with the coordinated multi-level Grid scheduling strategies (M16)
- M.IRWM.03: Evaluation of Implementation and Test of Grid Information, Resource and Workflow Monitoring Services (M20)
- M.IRWM.04: The IRWM Institute workshop organized in conjunction with an international European conference (M24)
- M.IRWM.05: Joint publication about integrated IRWM services (M30).

The effort done during the first project year was focusing on information exchange between partners about the knowledge, experience, research interests and development made in R&D projects correlated with CoreGRID. Then we identified research groups within each task, worked out common understanding of the problems being the subject of research and integration. Finally the research groups worked out a definition of common terms and naming used for future models.
The Integration Workshop in Pisa (November 2005) was an opportunity for all research groups to present the results of the 18 months of cooperation.

The roadmap version 2 defines the forthcoming new phases of the research and integration work:

Phase one:
Consolidation of common perception of the features to be included in the model.
Definition of models and architecture

Phase two:
Integration work
Joint reports and publications

Phase three:
Proof of concepts
Providing prototypes

Phase four:
Integration between tasks
Cooperation between CoreGRID Institutes.

The following sections include detailed information about contributions, common interests and integration work, which defines the roadmap of work to be done in each task within the established research groups.

Grid Information and Monitoring Services

Task 5.1 identified one research group between INFN and FORTH in terms of network monitoring systems. Once the Architecture of the Network Monitoring system is defined, and we are currently at this stage, we need to establish the hardware support for its components. After this point, we will proceed with the integration of a

prototype Network Monitoring Element within the GlueDomains network monitoring system. The next step will consist in the modification of the GlueDomains database to support the distributed caching mechanism.

The specific roadmap for the coming months are the following:
1. Define the architecture of the Network Monitoring System, intended as a set of functionalities needed to carry out such task, and their location in a Grid layout; this also entails identifying security issues related to the Network Monitoring System, and the proposal of scalable solutions (March 2006)
2. Define the hardware characteristics of Grid components that are devoted to Network Monitoring. This requires assessment of the network characteristics that can be effectively observed, and the way they are processed and published. Survey file transfer Grid services in order to assess their potential as a source of network performance measurements (June 2006)
3. Implement a prototype focussing on modularization and implementing basic passive monitoring features. Study architectural features inherent to the collection and publication of measurements related to the amount of data exchanged in a Grid infrastructure over time (October 2006)
4. Use GlueDomains as a testbed for early assessment of the prototype. Study of the publication of performance measurements through standard Web service interfaces (GGF NM-WG schemata) – December 2006.

The network monitoring architecture which is under study within the FORTH-INFN research group is deeply concerned with scalability issues, and we have considered several facets of such issue. The amount of bandwidth and computing power consumed by the monitoring activity is a first source of scalability problems: if every end-to-end path were permanently under test, the footprint of the monitoring activity would grow with the square of the number of resources in the system, which raises a scalability problem. A detailed study of the scalability can be found in the subsection 'Extended Context'of this document.

Checkpointing Services

Within task 5.2 the following research groups have been identified:
PSNC, SZTAKI: The group is working on the implementation of integration of TCKPT and PSNC's checkpointer.
Goal: Implementation of TCKPT and selected PSNC's low-level checkpointer.
Results: Framework allowing to checkpoint PVM applications. Paper describing the solution.

UCO, PSNC: The group working on a definition of requirements in relation to the storage services and on recognizing joint research areas with UCO.
Goal: Definition of required storage functionality for distributed checkpointing purposes.
Results: Contribution to paper.

We have also identified the main goals for a new research group working in connection with WP6:

PSNC, NN (will be defined after the joint meeting WP5 and WP6 in April):
The group is working on integrating the GCA with the Broker and other external services.
Goal: Work out the changes (if any) and interfaces defining the way the GCA interacts with the Broker and other external services.
Result: Paper describing the designed changes and interfaces.

The outcome of the PSNC-SZTAKI research group will be the first prototype of checkpointing functionality working on PVM and kernel level. This is an approach which will prove the general proposition of delivering a complex Grid Checkpointing Architecture.
UCO-PSNC is a new group established last year, which is starting the research in terms of adding checkpointing to Grid Desktops (cross activity between WP5 and WP4).

During the next eighteen months we plan to participate in writing the common papers on the following topics:
- A paper describing the revised version of GCA. The changes were introduced as a result of feedback of the first publication. Planned release time is the 2nd quarter of 2006.
- A paper describing the TCKPT, how it allows checkpointing the PVM applications and how it can utilize the third party checkpointers. Additionally the idea and rationale of integratingon the TCKPT and one of PSNC's checkpointers will be presented. Anticipated release is the 3rd quarter of 2006.

- When the integration of TCKPT and PSNC's low-level checkpointer finishes, a paper describing the results and the new product will be written. Expected publishing time is the 4th quarter of 2006 (it also depends on when the integration itself is finished).
- Task 5.2 plans to contribute to one of UCO's, storage for checkpointing in desktop-Grids related, papers. In that paper we want to express the GCA's requirements in relation to the storage system.
- After the GCA becomes even more integrated with external Grid Services, a paper describing the changes and interfaces to the external services will be created.

Within the nearest months we plan to publish technical papers on the following topics:

- Analysis of the possibility of adapting the AltixC/R checkpointer to the TCKPT requirements. A Technical Report prepared by the Research Group which strives for working out the plan of TCKPT and one of PSNC's checkpointers integration.
- A Technical Report describing the revised version of the Grid Checkpointing Architecture. As a result of feedback received after the first version of GCA and as a result of internal task 5.2 discussion, the GCA was a little refined. The changes will be presented in that report.

Partners	Contributions/Interests and R&D background projects	Task no.
PSNC	In the recent years, PSNC has been very active in the area of R&D issues concerning the checkpointing technology. Interest in that area resulted in the three checkpointing packages that are briefly described below. **psncLibCkpt** The user-level library that provides the checkpointing functionality for Solaris 8 OS. The core part of the psncLibCkpt is based on the libCkpt library. The most important novelty of the psncLibCkpt is the ability to checkpoint and restart multi-process programs that utilize System V IPC objects to mutual communication and synchronization. It is our first product in which we have introduced the virtualization of identifiers and keys related System V IPC. Thanks to that, when the program is recovered, it is cheated that the identifiers have not changed (even though due to technological reasons, it is very likely that they have). To utilize the psncLibCkpt library, the program has to be recompiled against the library and the only modification of source codes encompasses replacing the name of the main() function with the ckpt_target() name. The library is designed to be used with programs written in the C language. No special installation or deployment activities are required so the library can be used even by any regular, not privileged user. The package has been developed as part of the **PROGRESS** project. **psncC/R** Our first kernel-level checkpointing package. The product is aimed at Solaris 8 and 9 OS running on UltraSparc CPU. The main advantage of the kernel-level approach is full transparency for programs that are to be checkpointable and independent of the programming language that was used to write these programs. From the end-user's point of view, the utilizing of kernel-level checkpointing package is really convenient and simple but requires some deployment work that has to be done by the system administrator. Similarly to **psncLibCkpt**, that package has been developed within the **PROGRESS** project. **AltixC/R** The kernel-level checkpointing package designed for Altix systems equipped with IA64 processors and running under the ProPack environment (a Linux-based environment prepared by SGI). Currently we have versions of our package that works with ProPack based on Linux kernel 2.4 as well as with a more recent ProPack that is based on Linux kernel 2.6. The package is characterized by all features typical for kernel-level approach. It is easy to use; there is no assumption on the availability of source codes or the programming language that was used to write the programs that are to be checkpointed. The	5.2

	package has to be deployed by the system administrator. The package allows doing checkpoints of multi-process programs that communicate through System V IPC objects. Additionally, the idea of virtualization of some system global keys and identifiers has been employed in that product as well (advantages of such virtualization are the same as in case of the psncLibCkpt library). The package has been developed as part of the **SGIGrid project** but we intend to further extend the functionality of that package also beyond the SGIGrid project. Currently we are making every effort to add support for programs that use threads and "local" sockets. Such features would allow us to prepare the package with the capability of doing checkpoints of some MPI programs. **CoreGRID activities** The aim in CoreGRID project is to define the high-level checkpoint-restart Grid service and to alocate it among other Grid services. We aim to define the Grid Checkpointing Architecture that will encompass both, the abstract model of that service and the lower layer interface that will allow the service to co-operate with the diverse existing and future checkpointin-restart tools. PSNC is involved in the following related activities: • a group working on integrating the GCA with the Broker and other external services (prototype) • a group working on integration of TCKPT and PSNC's checkpointer (prototype) • A group working on a definition of requirements related to the storage services (initial co-operation with UCO).	
MTA SZTAKI	In the recent years MTA SZTAKI aimed at making parallel – GRAPNEL, PVM, MPI – applications capable of being checkpointed and restarted on clusters. GRAPNEL is a graphical language to define parallel applications in the PGRADE development framework. Until now MTA SZTAKI has been working on the following checkpoint issues: • User-level, automatic, binary checkpoint for GRAPNEL applications • Automatic checkpoint/restart mechanism for GRAPNEL applications under Condor scheduler (without any modifications in the scheduler) • User-level, automatic, binary checkpoint for native PVM applications • User-level, automatic, binary checkpoint for MPI applications All the issues mentioned above are mainly targeted in a way that the underlying single process checkpointer is taken off-the-self, which means we do not deal with research on single process checkpointer, we are focusing on handling parallelism and the software architecture that services the mechanism. MTA SZTAKI had several local projects in the last 3-4 years. They were about to sponsor the research and development regarding the above mentioned areas. These are the "Hungarian Supercomputing GRID" sponsored by the Ministry of Education (IKTA4-075), "Cluster Programming Technology and Its Application in Meteorology" sponsored by the Ministry of Education (IKTA3-029), "Chemistry GRID and its Application for Air Pollution Forecast" sponsored by the Ministry of Education (IKTA5-137) and the currently active project is the "Hungarian SuperCluster project" sponsored by the National Office for Research and Technology (IKTA-00064/2003) which aims at giving parallel checkpoint support for the nationwide Hungarian Clustergrid infrastructure which is the largest cluster-based Grid in Hungary. **CoreGRID activities** In CoreGRID project SZTAKI closely cooperates with PSNC in order to define and refine the Grid Checkpointing Architecture. The CoreGRID Research Groups that SZTAKI is involved in include:	5.2

	• The group working on integrating the GCA with the Broker and other external services. • The group working on implementation of integration of TCKPT and PSNC's checkpointer.	
UCO	The fault-tolerance and checkpointing issues are addressed by the Dependable Systems Group. The background projects that the group is involved in are: **DBench - Dependability Benchmarking** The Dbench's goal is to define a conceptual framework and an experimental environment for benchmarking the dependability of COTS and COTS-based systems. The final product will allow system developers and end-users to evaluate the dependability of a component or a system. It will also allow to identify malfunctioning or less weak parts, requiring more attention, tuning a particular component to enhance its dependability, and comparing the dependability of alternative or competing solutions. **mCrash** In this project UCO intends to evaluate the robustness of a standard kernel for mobile devices. The project will focus on the Applications Programmer's Interface (API) since this is the most common cause of crashes and malfunctions. **MATER** The project will research into models that can reproduce the demographic development processes that produce such type of distributions. Computer models and realistic representation of territories will be used to test how empirical population distributions can be approached by simulation. **RAIL** The objective of this project is to develop a library that allows .NET assemblies to be manipulated and instrumented before they are loaded and executed in the CLR virtual machine. **TACID** The goal of this project is to investigate ways to add timeliness properties to the typical ACID transactions. **VAL-COST-RT** This project aims at researching a methodology to assist software engineers in lowering the risk of using COTS components in avionics and space subsystems. The methodology will be used in case-studies defined with the NASA IV&V facility team. **DWS** In this project UCO proposes to develop an innovative technology to implement a data warehouse over an arbitrary number of computers (typically cheap workstations) and, at the same time, integrating this approach in the data warehousing technology available in the market. **CoreGRID activities** UCO group is interested in increasing the reliability and flexibility of Desktop Grids. In this context the recently established research group between UCO and PSNC will deal with checkpointing and storing of images and management issues, which allow to increase reliability of Desktop Grids computing of . The group is working on requirements storage services.	Cross-activity 5.2, 4.3

Workflow Services

Within Task 5.3 the following research groups have been identified:

UMUE, FHG: Workflow description languages using high-level Petri nets for Grid workflows

This research group continues its research about the workflow management of Grid applications using high-level Petri nets. Several joint papers have been published and further articles are currently in preparation:

Published:

- Martin Alt, Andreas Hoheisel, Hans-Werner Pohl, Sergei Gorlatch: Using High Level Petri-Nets for Describing and Analysing Hierarchical Grid Workflows. In: Proceedings of the CoreGRID Integration Workshop 2005, Pisa, 2005
- Martin Alt, Andreas Hoheisel, Hans-Werner Pohl, Sergei Gorlatch: A Grid Workflow Language Using High-Level Petri Nets. In: Proceedings of the PPAM05, Poznan, 2005
- Martin Alt, Sergei Gorlatch, Andreas Hoheisel, Hans-Werner Pohl: A GridWorkflow Language Using High-Level Petri Nets. CoreGRID Technical Report Number TR-0032, Institute on Grid Information and Monitoring Services, CoreGRID, 2006

In preparation:

- Andreas Hoheisel and Martin Alt: Petri Nets. In: Workflows for eScience, Springer, 2006 (in preparation)

Research tasks

After the specification of the common workflow description language "GWorkflowDL" has been established by the partners, the next steps are to implement tools and Grid services on top of this specification. This research group now focuses on the coordination of the research and development within other projects which use or aim to use this approach. Currently the work of this research group has strong impact on the projects K-Wf Grid (EC), MediGrid (Germany), and Instant-Grid (Germany), and the Fraunhofer Resource Grid. Other projects already announced their interest in this approach and we plan to further promote it at conferences and within the Global Grid Forum.

FHG, UMUE, UNICAL, INFN: Compatibility and conversion of different Grid workflow description languages

The objective of this research group is to publish a joint survey about compatibility and conversion (mapping) issues between commonly available workflow description languages, such as BPEL, GridAnt or DAGman. This group is currently in the formation process, so the final number of participating persons and institutes is not fixed yet. The call for participation has been announced at various CoreGRID events and at the Grid Workflow Forum (see http://www.gridworkflow.org/snips/gridworkflow/space/Workflow+Description+Languages/Compatibility+and+Conversion).

Research tasks

We will start with theoretical research about possible mappings between different workflow description formalisms that are either very wide-spread (such as BPEL), or used in one of the partner's projects (e.g. DAGMan) and continue with the analysis of available conversion tools, such as the BPEL to Petri Net converter developed by the Humboldt University in Berlin.

SZTAKI, UoW: Fault tolerance in Grid workflow execution

The objective of the reserch group is to elaborate on the theory of fault tolerant workflow execution on top of widely used Grid middleware systems, namely Globus, LCG-2 and gLite. The group members are currently using the Condor DAGMan workflow manager as a workflow enactor tool in their P-GRADE Portal and GEMLCA systems. Within Task 5.3 the group would like to implement an intelligent fault tolerant layer on top of Condor DAGMan. The first results of the research group are described in the following publication [BKT,2006].

Fault tolerant workflow execution needs the intelligent integration of an information system, resource brokers and workflow scheduler/enactor tools. While other Tasks of this WP are working on information and monitoring systems, scheduling is the topic of WP6. Consequenlty, this research group must be a bridge between WP5 and

WP6 researchers in order to build high-level workflow manager systems onto low level information sources and schedulers.

Research tasks:
- collecting information from different sources for workflow management purposes (building onto the results of Tasks 5.1 and 7.2)
- generating high-level availability and performance predictions for workflow scheduling algorithms
- fault tolerance at the workflow level: job resubmission, resource re-allocation
- fault tolerance at the job level: job checkpointing and migration (building onto the results of the task 5.2).

Members will have short visits on a two-month basis and will use VoIP (Skype) and document exchange (BSCW) tools.

MU, SZTAKI: Extending the SEAGRIN semantic overlay Grid infrastructure with the collaborative workflow management support of the P-GRADE portal

SZTAKI and MU are working on an ontology layer to be inserted into the P-GRADE Portal. The purpose of this layer is to make the portal capable of operating in a medical environment in which different medical assistants would like to execute several diagnoses on a patient. In April SZTAKI is going to have a short visit to MU in order to present the latest release of the P-GRADE portal and to work on the specification of the integrated SEAGRID – P-GRADE portal system.

Research tasks:
- Defining ontology for medical computations
- Defining a workflow manager which is able to understand and apply an ontologically described component to compose goal-oriented workflows.
- Management of workflows specified by ontological definitions.

Further contributions, interests and research & development background of the partners:

Partner	Contributions/Interests and R&D background projects	Task No.
FHG	**K-Wf Grid:** K-Wf Grid is a STREP project funded by EC that started in 09/2004 under the leadership of Fraunhofer FIRST with the objective to develop a "Knowledge-based Workflow System for Grid Applications". The K-Wf Grid system will assist its users in composing powerful Grid workflows by means of a rule-based expert system. All interactions with the Grid environment are monitored and evaluated. The knowledge about the Grid itself is mined and reused in the process of workflow construction, service selection and Grid behaviour prediction. Workflows are dynamic and fault-tolerant beyond the current state of the art. At the beginning of the project we analyzed the existing technologies in the domain of workflow composition and execution. After that we specified a graph-based workflow model that overcomes the disadvantages of the commonly used Directed Acyclic Graphs. The workflow description itself includes semantically-rich metadata on an abstract level that is independent of the Grid infrastructure. Furthermore, we will develop a "Grid Workflow Execution Service", acting as a software layer between the user-interactive Grid Application Building layer and the service-oriented lower-level Grid middleware. Other developments will include user interfaces for execution control and the monitoring of Grid jobs as well as administrative tools. **Fraunhofer Resource Grid, D-Grid:** Within the operation of the Fraunhofer Resource Grid and current contributions to D-Grid (German Grid Initiative), further research and development on our Petri-Net-based workflow approach will be done. The scientific goals related to Task 5.3 are: • Improvement of the Petri-Net-based Grid job description language (GWorkflowDL) • Service for the execution and monitoring of Grid jobs • Research on interoperability • Bringing further applications to the Fraunhofer Resource Grid • Higher level checkpointing of Grid workflows **CoreGRID Task 5.3 Research Groups** Within CoreGRID Task 5.3, FhG participates in the research groups “Workflow description languages using high-level Petri nets for Grid workflows” and “Compatibility and conversion of different Grid workflow description languages”.	5.3 UMUE, UNICAL, INFN
INFN	**Matchmaking of a set of activities involved in a workflow schedule:** Matchmaking sets of activities has been addressed with the Gangamatching approach that does not provide a high degree of orthogonality between coordination and computation. We are based on some standard workflow patterns that have been formalized with Petri nets but miss the description of important patterns as long running transactions. Negotiation protocols have been investigated in the area of Distributed Artificial Intelligence. A classic protocol is the Contract Net Protocol (CNP), proposed in the scenario of distributed problem solving. CNP has then been adapted to different types of multi-agent systems. INFN studies the formal definition of negotiation protocols, describing similarities and differences with respect to well known distributed transaction protocols as BTP that have already been formalized with process algebras. Some efforts are addressing the definition of standards for negotiation in the Grid/Web Service scenario (e.g. WS-Agreement by GGF), mostly in an agnostic way with respect to the underlying protocol. WS-Negotiation	5.3, WP2 FHG, UMUE, UNICAL

	considers negotiation in all its facets from a Web Service perspective but it lacks a formal description of the negotiation protocol. **Semantic resource discovery in Grid Environments (in cooperation with WP2)**: With the aim of enabling automation of services matchmaking, in the context of WP2 we are developing an extension of a service description language (possibly OWL-S) that takes into account non-functional parameters and negotiable parameters. The goal of this project is to investigate the usage of the service descriptions expressed in the proposed OWL-S extensions. Within this project we aim to investigate the problems of workflow matchmaking and the negotiation issues. We aim to clarify which workflow patterns to consider in our formal framework based on the existing work on Petri nets for standard workflow patterns and mainly on pi calculus concerning transactional behaviour. We intend to define a set of service features for each pattern that should be considered in the matchmaking of a composition. Our goal is to consider the use of an inferential Grid information system, a definition of the requirements for a negotiation protocol in a Grid scenario, investigation on the relationship between negotiation and Web-distributed transactions, and the formal definition of a suitable negotiation protocol. **CoreGRID Task 5.3 Research Groups** Within CoreGRID Task 5.3, INFN participates in the research group "Compatibility and conversion of different Grid workflow description languages".	
MU	**SEAGRIN –** Semantic Service-oriented Workflow Infrastructure Development of concepts and prototype implementation of service-oriented workflow systems combined with the collaborative environment. Finding ways to create, manage and inspect workflows within the collaborative environment, where different partners have different knowledge of and interest in the complex workflow parts (e.g., the processing of laboratory data and some medical images), while sharing a common goal expressed by the workflow (e.g., to converge on a diagnosis). The secondary goal is to study the use of ontologies to glue different workflow components (services) into a consistent whole. Scientific Goals: • To provide workflow within the collaborative environment • To provide examination possibilities of workflow semantic consistency checking. **CoreGRID Task 5.3 Research Groups** Within CoreGRID Task 5.3, MU participates in the research group "Extending the SEAGRIN semantic overlay Grid infrastructure with the collaborative workflow management support of the P-GRADE portal".	5.3 SZTAKI, UoW
SZTAKI	**Workflow Management in Grid Environments:** Creation of a prototype implementation of a workflow management infrastructure which enables the automated execution and monitoring of workflows consisting of parallel or sequential applications. Building onto the transparent job-level checkpointing and migration support provided by the components SZTAKI will develop as part of its 5.2 contribution, the manager will be able to execute job-workflows in a fault-tolerant way in Grids. The workflow manager will be built onto the core scheduler functionality of Condor DAGMan. **CoreGRID Task 5.3 Research Groups** Within CoreGRID Task 5.3, SZTAKI participates in the research groups "Extending the SEAGRIN semantic overlay Grid infrastructure with the	5.3 MU, UoW

	collaborative workflow management support of the P-GRADE portal" and "**Fault tolerance in Grid workflow execution**".	
UNICAL	The project is concerned with the adoption of WF as a model for defining complex service-oriented Grid applications involving a number of jobs interacting with each other in various ways. Scientific goals: Defining a specification mechanism simple enough to be used in several application domains and powerful enough to exploit the Grid infrastructure. **CoreGRID Task 5.3 Research Groups** Within CoreGRID Task 5.3, UNICAL participates in the research group "Compatibility and conversion of different Grid workflow description languages".	FHG, UMUE, INFN
UOW	**GEMLCA** We have an environment where users can construct, manage and execute visual workflows from a Grid portal in order to support the execution of Grid-enabled Legacy Codes. The middleware called GEMLCA (www.cpc.wmin.ac.uk/gemlca) was developed by the University of Westminster during the UK EPSRC funded OGSA testbed project. GEMLCA uses Globus Toolkit 3 and is being migrated to GT4 (WSRF). GEMLCA was demonstrated at IST 2004 in Hagues and there is an article about it in the 'Next Generation Grids' publication document. The current workflow system includes support for GT2 and service-oriented GT3 Grid using Gemlca jobs. Gemlca supports the execution of Grid-enabled legacy codes. The current workflow execution manager is based on Condor DAGman. Scientific goals: Research tasks include the following: - Workflows o Specifying workflow with QoS requirements o Scheduling, brokering, and support for different production Grids (interoperability issues – Globus, EGEE) - Dynamic scheduling of workflow jobs onto Grids in a fault-tolerant way o Dynamic testing of Grid resources o Assigning tasks of a workflow application onto most reliable and best performing resources with a view to minimize workflow execution time. o Workflow-level fault tolerance, on-demand re-scheduling and re-submission of jobs - Grid portals **CoreGRID Task 5.3 Research Groups** Within CoreGRID Task 5.3, UoW participates in the research group "**Fault tolerance in Grid workflow execution**"	SZTAKI
UMUE	**Using Coloured Petri Nets for Grid Workflows:** The usage of Coloured Petri Nets for describing complex Grid Workflows is investigated. A prototype system will be implemented, which takes local scheduling policies of resources into account. In a theoretical part, the task of modelling different aspects (e.g. cost information) of Grid services or components using Coloured Petri Nets are studied in order to analyse complex Grid workflows and allow for efficient execution. The system should make use of available Grid middleware but not be bound to a certain one. Therefore, a common interface to different underlying middleware systems has to be developed. The scientific goals are to provide a system environment for users to describe complex Grid workflows intuitively. Upon execution, the scheduling policy of used resources should be taken into account when distributing subtasks of the workflow over the Grid.	FhG, UNICAL, INFN

	Deliverables: The project has already started (still in its early stage), and the first concepts have been sketched to Nov/Dec 2005: prototype system information. Until mid 2006: Enhancement of the system, modelling procedures, extensive tests **CoreGRID Task 5.3 Research Groups** Within CoreGRID Task 5.3, UMUE participates in the research group "Workflow description languages using high-level Petri nets for Grid workflows" and "Compatibility and conversion of different Grid workflow description languages".	

Accounting and User Management Services

One research group between PSNC and MU is identified as a continuation of work in terms of user account management and accounting in Grid environments.

The specific roadmap for the coming months is the following:
- Preparing a detailed architecture of the proposed user management framework.
- Integration and implementation of the system.
- Analysis of requirements for accounting a gathering system.
- Definition of a minimum set of resources that the accounting system in a production Grid must gather and process.
- Definition of a database structure that will allow storing other and nonstandard accounting data.
- Design of a model architecture for the accounting system that will comply with different resources from different Grids and will take into account local and VOs policies.
- Implementation of a model system.

The specific roadmap for the coming months is the following:
- Preparing a detailed architecture of an integrated framework for accounting and account management services – March 2006
- Implementation of an integrated framework for accounting and account management services – September 2006
- Analysis of requirements for the accounting system, design of the model architecture – March 2007
- Implementation of accounting services – September 2007

Partner	**Contributions/Interests and R&D background projects**	***Task No***
PSNC	Poznan Supercomputing and Networking Center has performed research on user accounts management and the distributed accounting information system for several years. The project name is the Virtual Management System (VUS). The idea of VUS is universal and can be implemented with a different job management system. VUS is just an extension of the system that runs users' jobs (e.g. queuing system, Globus Gatekeeper, etc.) and allows to run jobs without having a user account on a node. This allows to minimize overhead related to creating and maintaining additional user accounts. On the contrary to other solutions, VUS assures an accurate security level achieved by user authorization and possibility of charging the user with costs of resource usage. Additionally, it respects the local policy of sites and makes it possible for the local administrator to differentiate between local and remote users. The system is transparent for the users and to some extent also for administrators. Users just run their jobs and do not care about account assignment. Administrators just configure access to the machine for VO and do not need to create accounts and manage grid-mapfile. The system does not interfere with local access	5.4

MU	policy either. The history of VUS started in 1998 [DMS,1999], [KLM,2001]. The first implementation of VUS was an extension to queuing systems (e.g. LSF) and it was successfully exploited 3 years ago in the Polish national cluster which connected several HPC centers in Poland [SGAS],[SNUPI],[MW,2003]. Later we focused on Globus. The first version of VUS for Globus 2.4 was using a modified Gatekeeper. The current implementation is GRAM "callout", a mechanism introduced in GT 3.2. The major advantage of callouts is that there is no need to modify Globus Toolkit (GT) codes to install VUS (the older versions of GT needed slight modifications in the Gatekeeper and GridFTP). The research on account management and distributed accounting processing is performed in several projects. • National Computing Grid – set up by PSNC as an attempt to connect supercomputers from different Polish supercomputing centers with the LSF queuing system. [KLM2,2001], [KLM,2001] • SGIgrid – "High Performance Computing and Visualization with the SGI Grid for Virtual Laboratory Applications". The goal is to design and implement an environment for remote access to unique and therefore expensive laboratory equipment. The environment will support the end user by delivering enough computational power for pre- and post-processing computations and data intensive visualization. One of the project goals is to provide a backup data center for the national Institute of Meteorology and Water Management. The project started in December 2002 and finished in November 2005 [MBM,2001], [NBM,2001]. • Clusterix – "National CLUSTER of LInuX Systems". The goal of the project is to build a new generation distributed PC cluster by connecting 64bit PC from 12 Polish computing centers. Middleware implemented in this project will allow the test application to run in a dynamic environment with a changing number and configuration of computing and network infrastructure [WMS, 2005]. • BalticGrid - The goal of the BalticGrid project is to extend the European Grid by integrating new partners from the Baltic States (Lithuania, Latvia and Estonia) in the European Grid research community and to foster the development of Grid infrastructure in these countries. To this end, the BalticGrid consortium has enlisted the help of experienced EU Grid computing centers whose aim will be to guide our new partners through the process of deploying Grid resources and applications at their respective institutions. Some cooperation has also been done within the 5FP EU GridLab Project.	

Mechanisms

Partner Meetings/Workshops
The all hands gatherings are considered to be the most important tool for collaboration and integration promotion. They will be held every 4 to 6 months, either as independent meetings (probably related to similar meetings of other Institutes) or associated to some public workshop.
The purpose of these meetings is manifold:

- To serve as a place for the presentation of results, with emphasis on work done in collaboration between partners
- To serve as a discussion venue for future work and direction of research (both operational and strategic decisions will be prepared)
- To provide a forum for the preparation of deliverables and other CoreGRID-related documents
- To serve as a research/development environment (extended meetings for several days, focused on a particular research topic to be approached from different perspectives by several partners in the collaboration).

The first meeting of this Institute was held in Heraklion (at the FORTH premises, Cyprus, Greece) on January 19th, 2005. This meeting followed a one-day WP4 organized workshop and a one-day WP4 meeting, thus strengthening also the inter Institute collaboration. The major purpose of this meeting was the identification of individual partners' contribution to the Roadmap and therefore to the actual scientific work of this Institute.
The second WP5 meeting was held in Barcelona on July 15th, 2005. It was an open forum of research group presentation – the main results of the integration work.

The outcomes of the integration work were presented at the first CoreGRID Integration Workshop (November 2005, Pisa).

The IRWM Institute is planning the following meetings:

- April 2006 – a general WP5 meeting, a joint meeting with WP6
  Presenting the current state of co-operation

- August 2006 – a joint workshop of IRWM, KDM and RMS Instituties [GRM, 2006]
  A presentation of joint research activities, cross activities between the KDM, IRMW and RMS Institutes. It will be also a possibility to see the research done by institutions, which are not CoreGRID partners.

- 4Q2006 – the Second CoreGRID Integration Workshop

E-meetings and tele-conference meetings
Most of the information exchange between partners will occur through e-mail exchange (either individual or through the Institute e-mail list) and through the document sharing on the CoreGRID BSCW. The IRWM Institute and task leaders will have semi-regular tele-conference meetings (usually with bi-weekly periodicity).
Partners will try to setup an e-collaborative environment, using either publicly available systems like VRVS, through H.323 videoconferencing tools where available or using simpler tools like MBone tools with some reflector. These potential e-meetings will be focused mostly on specific research topics, they will not make a complete substitute for the face to face partners meetings.

Researcher/Student Exchanges
On top of short visits funded directly by the CoreGRID, partners plan to promote joint PhD programs in relevant areas. The activity will start with short visits, students' exchange, and is expected to culminate in shared PhD supervision.
If knowledge/expertise gaps are discovered, partners plan to invite also researchers external to the CoreGRID or to visit institutes with relevant scientific knowledge.

Dissemination of results
Due to the primary research focus of the CoreGRID, the primary dissemination platforms are scientific journals, conferences, workshops, and similar venues. Partners expect to submit joint publications, which will also be available on the CoreGRID BSCW (unless the publishing policy prohibits such form of dissemination).

Apart from joint publications, partners will extensively use the results achieved within the CoreGRID in their own work, thus extending strength and knowledge of the CoreGRID-related research.
The CoreGRID public portal pages are used to promote the knowledge gained through joint work within this Institute.

CoreGRID BSCW data base
All partners have access to the CoreGRID BSCW and it is used to exchange documents and other digital material related to the CoreGRID activities. The portal is also extensively used for cross Institutes collaboration. Both internal and external web pages are maintained by the partners.

Exchange of documents
All documents are stored on the CoreGRID BSCW server, which provides a uniform www-based document and file sharing platform. As all the CoreGRID participants have secure access to the BSCW server, it is used by the Institute for easy upload and sharing of all textual information (documents, source code, meeting minutes, etc.).

Short visits
Short visits are considered to be one of the most important tools to initiate and later strengthen the collaboration between partners. The individual short visits will always have a clear purpose, usually a discussion, writing a joint scientific paper or preparing a deliverable contribution.
Most short visits are around 3 days and will not exceed a week. The actual length of stay may change even during the stay if an interesting research subject is discovered.
The list is not complete, as short visits are expected to be a flexible tool used as necessary (Table 1). In addition the role of short visits is not as important now as it was in the first stage. The first phase of the roadmap version 1 was focusing on exchanging experience, knowledge and finding common research interests. After the work was done, further integration is done remotely, using teleconferences, e-mail exchanges and meetings at workshops, CoreGRID or other related conferences.

From	To	TO DO
FORTH	INFN	A CoreGRID report describing the results of the research done so far. Preparation of a conference/journal paper
INFN	FORTH	Workshop at the end of April, about Network Monitoring and related security issues (in conjunction with a meeting at INFN
MU	PSNC	Technical Report with detailed architecture – April-May 2006 Paper prepared in 2Q 2006 (Conference or Journal) WP5 meeting – April 2006 MU to PSNC – June 2006 Common meeting during the Euro-Par 2006 conference – Dresden, August - September 2006 PSNC to MU - March 2007
SZTAKI	PSNC	The goal is the presentation of and discussion on the results of the performed analysis concerning the possibility of integrating TCKPT and PSNC's checkpointer. The actual platform and integrated products have to be chosen. Additionally a discussion on adapting the GCA to the external Grid Services and on planned joint papers is required
PSNC	UCO	PSNC and UCO established the Research Group that aims to define required storage functionality for distributed checkpointing purposes. Then we intend to meet in order to present our experiences and positions in the fields of interest of each institute.
PSNC	WP6 partner	As part of work on integrating GCA with the Broker and other external Grid Services we plan to meet with at least one member of WP6. The actual partner is not known at the moment; however, in order to place

		the GCA within the Grid infrastructure and to integrate it (as proof of concept implementation) with selected schedulers we require cooperation with people from WP6.

Table 1: A list of short visits between partners planned in the first stage of second project year

Depending on the requirements and according to the circumstances the teleconferences are planned. At the moment we are taking into account at least two mandatory teleconferences that will take place:

- A preliminary teleconference between PSNC and UCO to establish the preliminary arrangements concerning the work on a definition of required storage functionality for distributed checkpointing purposes. The agreement on a Short Visit is required.

- A teleconference between PSNC and SZTAKI to perform a preliminary debate about GCA and Broker integration and about the need to find a new partner from WP6. The agreement on a Short Visit is also required.

What we basically intend to do next on the basis of the roadmap is to:

- promote the participation of partners into a new research proposal, assuming the results of this WP as the starting point
- ensure stable and durable cooperation between WP partners

5. Trust & Security Issues

We recognize security as an issue for **network monitoring**, and we propose a solution for a distributed certificate caching aimed at network monitoring management. We have planned a meeting in Bologna which is finalized to discuss this and other issues in the domain of security. Both FORTH and INFN have recognized competence in the field, and there is a possible confluence of interests in the area of security, which could extend to other WPs. This will be a matter of discussion during a meeting organized in Bologna by INFN, scheduled around the end of April.

We have taken an assumption that the implementation of **Grid Checkpointing Architecture** will be performed with help of technologies that allow to employ the build-in security mechanisms. The most recent Web Services based technologies (i.e. those available in GT4) ensure privacy, integrity and proper authorization and authentication based on widely accepted world wide standards (X.509 certificates, proxy certificates, digital signs, TSL transport protocol). However, the security issues as such are not within the area of interest of task 5.2. Therefore the services designed within task 5.2 to provide an adequate trust and security level will take advantage of mechanisms offered by other services and the proof-of-concept implementation environment.

Trust and Security within **workflow services** is a very challenging issue, because distributed Grid workflows are mostly composed of activities distributed over a set of organizations that maintain diverse security policies, connected with each other over the insecure Internet. In order to reflect this importance, we plan to evaluate the possibility in Task 5.3 of writing a joint proposal for the CoreGRID fellowship programme, in order to involve additional researchers in this topic.
Consequence of the nature of workflows that usually a large number of resources are required for the optimal execution of component, for the optimal selection of resources. Advanced workflow enactor systems and scheduler algorithms must interface with production Grids that can provide large resource pools. Tools developed by Task 5.3 participants should be able to work on the most important European production Grids such as EGEE, UK NGS or Nordugrid. In order to achieve this goal, workflow managers must implement the security requirements of these infrastructures.

The main aim of **user account management systems** is to provide controlled and secure access to grid resources. Security requires authentication of the user and authorization based on combined security policy from the resource provider and the virtual organization of the user. Users must be allowed to use the resources to the extent allowed by the user roles and the policy, while resources must be secured against unintentional as well as

malicious policy breaks. The issue of authentication is well addressed by existing standards and solutions. There are some solutions in authorization area, but the subject is still being investigated also within this task.
The second important thing is the possibility of logging user activities for accounting and auditing (security reasons) and then gathering these data or their aggregates both by the resource provider and virtual organization of the user. The logging features are closely related to one of the main problems of user management: mapping the global user id to the local one, because they require identification of the user who performed some action. So that, we try to propose a solution that simplify the user management and will be scalable, but still allow for user identification.
The access to the **accounting and logging** data must be properly limited depending on the users' roles: consumers of the resources, administrators o f resources and managers of the virtual organizations. Thus, the accounting services must be secured and the access to the data must be controlled.

6. Link with other CoreGRID Institutes

This Institute will not be isolated. This synergy is to be expected especially between WP5 and the rest of WPs, particularly WP2, WP4, WP6 and WP7, as all the tools and environments developed must become part of the monitored infrastructure and will provide data to be evaluated and fed into the scheduling systems (Fig. 1).

Apart from the integration of research already done by individual partners within these Institutes, these services (information and monitoring, checkpointing, workflow, accounting and user management) are essential to other Institutes 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. Strong collaboration is thus essential and we foresee mutual benefits resulting from the common work within the CoreGRID NoE. This synergy is to be expected especially between the WP5 and WP2, WP4, WP6 and WP7, as all the tools and environments developed must become part of the monitored infrastructure and will provide data to be evaluated and fed into the scheduling systems. The first 12 months of collaboration also emphasized a real need of horizontal activities between CoreGRID Institutes, e.g. WP5 and WP6, WP5 and WP4, WP5 and WP7.
The work will be mostly organized into Research Groups, units of two or more CoreGRID partners collaborating closely together on common goals. Joint technical reports, publications and prototype implementation will be the major measurable outcome, while better personal networking and mutual trust will be the intangible, but even more important results

The distributed caching mechanism that we propose to cope with certificates management may share points of interest for the KDM Institute, in the domain of self-stabilizing algorithms. This is a possible link that will be explored in the future.

Thanks to CoreGRID NoE up to now we have defined the concept of **Grid Checkpointing Architecture** (GCA). Our next task is to make it even more integrated with the Broker and other external services from the point of view GCA. To accomplish that task the links with WP6 are required. We plan to set up the Research Group with participants from WP6. Simultaneously, as the need for defining required storage functionality has emerged, the cooperation with UCO from WP4 has been established. To define this required functionality, PSNC and UCO have set up the Research Group.

Workflows provide a high-level approach to compose compound Grid jobs and programs. The execution of such a composed Grid program is managed by workflow services which map the static, compile-time job flow description onto the dynamic Grid system at runtime. Therefore, these services have to access static information provided by the program itself and the workflow description.
Therefore, this task is strongly linked with the work on Grid programming models done in the Institute on Programming Models (PM). This linkage will be enforced in the future by means of nominating a specific person, in charge of building an information bridge between PM and IRWM Institutes.
Additionally, the runtime mapping has to acquire dynamic information from monitoring services and take resource management information and scheduling policies into account. Therefore, this task working on workflow services is linked as a technology user with RMS Institute, which provides a technical basis for the retrieval of resource management and scheduling information. Additionally, there is a strong connection to the task working on Grid architecture adaptability in the Institute on System Architecture. The workflows provided by IRWM provide input data for the analysis of overall demand for Grid resources and therefore provide a decision base for automatic Grid reconfiguration.

The **user account management services** should be transparent for other Grid layers as much as possible. However, there ate some situations, that the upper layers must be aware of the user management issues. e.g. in case of workflows, it must be assured that the subtasks of the workflow scheduled to the same node must be mapped to the same local identity (account, virtual machine etc.). Thus, there is need for cooperation with task 5.3 and scheduling related tasks from Institute on Resource Management and Scheduling (tasks 6.1-6.5).
There is group of researchers working on security in lightweight grid architecture in Institute on Grid Systems, Tools and Environments, task 7.1. The common interest would be authorization and user management in the lightweight Grids.

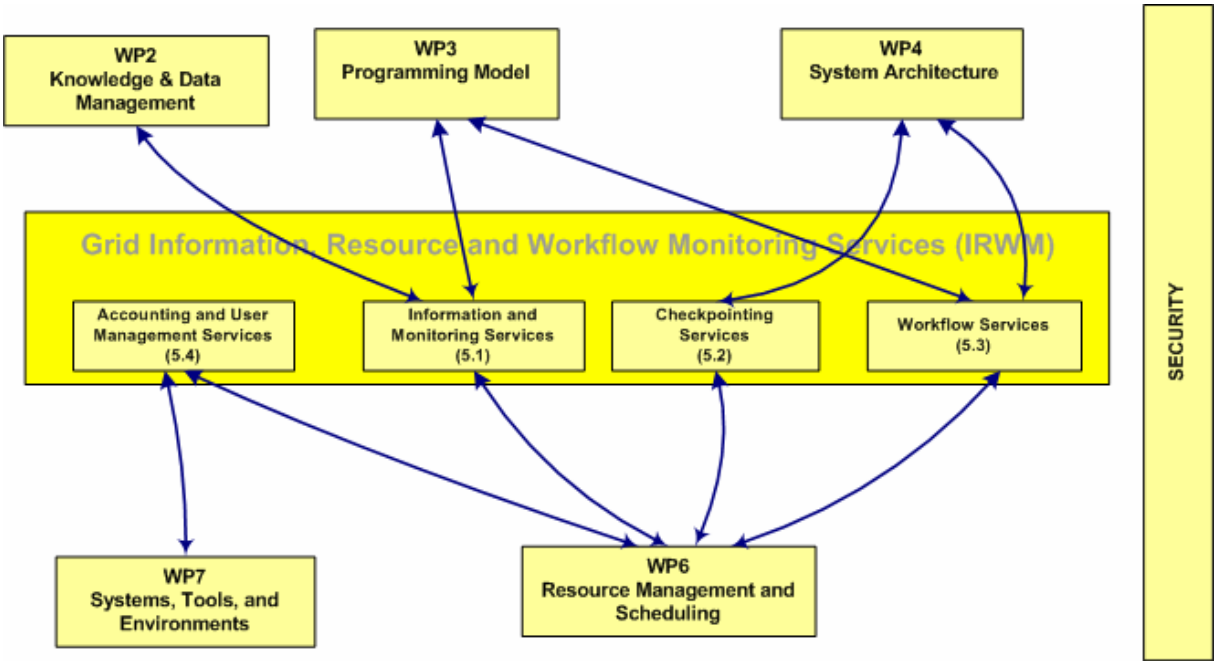

Fig. 1 Links with other Institutes

7. References

[AAC, 2005] S. Andreozzi, D. Antoniades, A. Ciuffoletti, A.Ghiselli E.P.Markatos and M.Polychronakis and P.Trimintzios, Issues about the Integration of Passive and Active Monitoring for Grid Networks, CoreGRID Integration Workshop, November 2005;

[AC,2005] Augusto Ciuffoletti, "Scalable accessibility of a recoverable database using a wandering token", Tech. Report TR-06-02, Dept. of Computer Science, Univ. of Pisa (also submitted to EUROPar);

[ACE,2005] G. Aloisio, M. Cafaro, I. Epicoco, S. Fiore., D. Lezzi, M. Mirto, and S. Mocavero : *iGrid, a Novel Grid Information Service*. Proc. EGC 2005, Amsterdam, The Netherlands, 2005 ; revised in LNCS 3470, Springer-Verlag, pp.1-9, 2005.

[BG,2003] (Mercury) Balaton, Z., Gombás, G.: *Resource and Job Monitoring.* In the Grid. Proc. of the Euro-Par 2003 International Conference, Klagenfurt, 2003. 404–411

[BKK,2005] L. Bocchi, O. Krajíček, M. Kuba. Infrastructure for Adaptive Workflows in Semantic Grids. In Proceedings of the first CoreGRID Integration Workshop. Pisa : University of Pisa, 2005. s. 327-336.

[BKT,2006] L. Bitonti, T. Kiss, G. Terstyanszky, T. Delaitre, S. Winter, P. Kacsuk, Dynamic Testing of Legacy Code Resources on the Grid, Conf. proc. of the ACM International Conference on Computing Frontiers, Ischia, Italy,May 2-5, 2006.

[BPEL] http://www-128.ibm.com/developerworks/library/specification/ws-bpel/

[CFF,2001] (MDS 2) Czajkowski, K., Fitzgerald, S., Foster, I., Kesselman, C.: *Grid Information Services for Distributed Resource Sharing.* In Proceedings of the 10th IEEE International Symposium on High-Performance Distributed Computing (HPDC-10), IEEE Press, August 2001.

[CFK,2006] Coviello, T., Ferrari, T., Kavoussanakis, K. et al."Bridging Network Monitoring and the Grid", in Proc of the CESNET Conference, Prague (CZ), Mar 2006.

[CKPT] http://psnc.checkpointing.pl

[CP,2004] Ceccanti A. Panzieri F.: *Content-based Monitoring in Grid Environments*. In Proc. 13th IEEE International Workshop in Enabling Technologies, WETICE 2004.

[CP, 2006] Augusto Ciuffoletti, Michalis Polychronakis "Architecture of a Netwrok Monitoring Element", CoreGRID Report n. 0033 (submitted on February 2006).

[DGAS1] http://www.to.infn.it/grid/accounting/main.html

[DGAS2] www.eu-egee.org

[D-GRID] http://www.d-grid.de/index.php?id=1&L=0

[DJK,2005] J.Denemark, M.Jankowski, A.Krenek, L.Matyska, N.Meyer, M.Ruda, P.Wolniewicz, *Best Practices of User Account Management with Virtual Organization Based Access to Grid*, PPAM 2005 Conference Proceedings,

[DJM,2005] J.Denemark, M.Jankowski, L.Matyska, N.Meyer, M.Ruda, P.Wolniewicz, User Management for Virtual Organizations, CoreGRID Integration Workshop Proceedings, Pisa 2005.

[DJMM,2005] J.Denemark, M.Jankowski, L.Matyska, N.Meyer, M.Ruda, P.Wolniewicz *CoreGRID Technical Report TR-0012: User Management for Virtual Organizations.*

[DMS,1999] W. Dymaczewski, N. Meyer, M. Stroiński, P. Wolniewicz, *Virtual Users Account System for Distributed Batch Processing*, in: P. Sloot, M. Bubak, A. Hoekstra, B. Hertzberger (Eds.): HPCN 1999, LNCS 1593, Springer-Verlag, 1999, 1231-1234.

[DMW,1999] W. Dymaczewski, N. Meyer, M. Stroiński, P. Wolniewicz, *Virtual Users Account System for Distributed Batch Processing*, in: P. Sloot, M. Bubak, A. Hoekstra, B. Hertzberger (Eds.): HPCN 1999, LNCS 1593, Springer-Verlag, 1999, 1231-1234.

[GESA] http://www.doc.ic.ac.uk/~sjn5/GGF/gesa-wg.html

[GGF] http://www.gridforum.org/

[GLOBUS] http://www.globus.org/

[GMA-WG] GMA-WG: Grid Monitoring Architecture Working Group, http://www-didc.lbl.gov/GGF-PERF/GMA-WG/

[GMW, 2006] http://www.coregrid.net/GMW2006

[GN2] http://www.geant2.net/

[GRIDCPR] http://gridcpr.psc.edu/GGF/

[HKM,2004] Holub, P., Kuba, M., Matyska, L., Ruda, M.: *Grid Infrastructure Monitoring as Reliable Information Service.* Proc. 2nd European AcrossGrids Conference, LNCS 3165, Springer-Verlag, pp.220-229, 2004.

[JJK,2005] Gracjan Jankowski, Radoslaw Januszewski, Jozsef Kovacs and Norbert Meyer and Rafal Mikolajczak, Grid Checkpointing Architecture - a revised proposal, presented at CoreGRID Integration Workshop 2005.

[JJM,2005] Gracjan Jankowski, Radoslaw Januszewski, Rafal Mikolajczak and Jozsef Kovacs, On integration possibility of TCKPT and psncLibCkpt, CoreGRID Technical Report 0019.

[JKM,2005] Gracjan Jankowski, Jozsef Kovacs, Norbert Meyer, Radoslaw Januszewski and Rafal Mikolajczak, Towards Checkpointing Grid Architecture, to be published in PPAM2005 proceedings.

[JWM,2004] M.Jankowski, P.Wolniewicz, N.Meyer, *Virtual User System for Globus based grids*, Cracow Grid Workshop '04 Proceedings, Cracow 2004.

[KLM,2001] M.Kupczyk, M.Lawenda, N.Meyer, M.Stroinski, P.Wolniewicz: *Experiences with Virtual Users' Accounts System in Polish National Cluster*, CUG, Indian Wells, May 2001

[KLM2,2001] M.Kupczyk, M.Lawenda, N.Meyer, P.Wolniewicz: *Using Virtual User Account System for Managing Users Account in Polish National Cluster*, HPCN,Amsterdam, June 2001, 587-590.

[KLW,2001] M.Kupczyk, M.Lawenda, N.Meyer, P.Wolniewicz: *Using Virtual User Account System for Managing Users Account in Polish National Cluster*, HPCN,Amsterdam, June 2001, 587-590.

[MBM,2001] N.Meyer, M. Bubak, J. Madajczyk, Virtual Laboratory using *HPC/HPV infrastructure*, PIONIER 2001 conference, April 2001, Poznań, p. 155-163, (in Polish)

[MUPBED] http://www.ist-mupbed.org/

[MW,2003] N.Meyer, P.Wolniewicz, *Virtual User Account System*, Conference – 10th Anniversary of Poznań Supercomputing and Networking Center – New Network Technologies, Grids and Portals, Poznań, October 2003,

[NBM,2001] J. Nabrzyski, A. Binczewski, N. Meyer, S.Starzak, M. Stroiński and J.*Węglarz,First Experiences with Polish Optical Internet*, Terena 2001 conference, May 14-17, 2001, Antalya

[NGG2] Expert Group. *Next Generation Grids 2 (NGG2)—Requirements and Options for European Grids Research 2005–2010 and Beyond*, EC Report, ftp://ftp.cordis.lu/pub/ist/docs/ngg2_eg_final.pdf, 2004

[NGG1] Expert Group. *Next Generation Grids—Requirements and Options for European Grids Research 2005–2010 and Beyond*, EC Report, http://www.cordis.lu/ist/grids/pub-report.htm, 2004

[SGAS] http://www.sgas.se/

[SNUPI] SNUPI http://snupi.sdsc.edu/snupidoc.html

[SC,2005] Rafal Mikolajczak, Radoslaw Januszewski and Gracjan Jankowski, Checkpoint - Restart funcionality for fault tolerance and migration purposes, Stand and left on Supercomputing 2005.

[SF,2001] S. Fisher : *Relational Model for Information and Monitoring.* Technical Report, GWD-Perf-7-1, GGF, 2001.

[SMK,2004] Sitera L., Matyska L., Krenek A., Ruda M., Vocu M., Salvet Z., Mulac M.: *Capability and Attribute Based GRID Monitoring Architecture.* Proc. CGW04, 2004.

[SNUPI] http://snupi.sdsc.edu/

[TNC,2005] Rafal Mikolajczak, Radoslaw Januszewski and Gracjan Jankowski, Checkpoint Restart Packages Originated in PSNC, Presentation on TNC 2005 (published on TNC's web page).

[TPP,2006] Panos Trimintzios, Michalis Polychronakis, Antonis Papadogiannakis, Michalis Foukarakis, Evangelos~P. Markatos and Arne Oslebo, “DiMAPI: An Application Programming Interface for Distributed Network Monitoring, Proceedings of the 10th IEEE/IFIP Network Operations and Management Symposium (NOMS), April 2006;

[UNICORE] http://www.unicore.org/

[WMS, 2005] Roman Wyrzykowski, Norbert Meyer, Maciej Stroiński, Concept and Implementation of CLUSTERIX: National Cluster of Linux Systems, The 6th LCI International Conference on Clusters, April 25-28, 2005, Chapel Hill, North Carolina (USA),

http://www.linuxclustersinstitute.org/Linux-HPC-Revolution/Archive/PDF05/25-Wyrzykowski_R.pdf

[ZS,2005] Zanikolas S., Sakellariou R.: *A Taxonomy of Grid Monitoring Systems*. FGCS, 21(1), 163–188, 2005

8. Participants

			Task			
Partner	**No.**	**Researchers and Ph.D. students**	**T1**	**T2**	**T3**	**T4**
FHG	8	Andreas Hoheisel Thilo Ernst			X	
FORTH	11	Spyros Antonatos Michalis Polychronakis Paraskevi Fragopoulou Evangelos Markatos Panos Trimintzios	X			
INFN	13	Antonia Ghiselli Tiziana Ferrari Augusto Ciuffoletti	X		X	
MU	16	Daniel Kouril Ondrej Krajicek Ales Krenek Ludek Matyska Miroslav Ruda Lukas Hejtmanek Jiri Denemark	X		X	X
PSNC	17	Gracjan Jankowski Michal Jankowski Norbert Meyer Dominik Stoklosa Radek Januszewski Rafal Mikolajczak Pawel Wolniewicz		X		X
SZTAKI	20	Gabor Gombas Gergely Sipos Jozsef Kovacs	X	X	X	
UCAM	24	Mark Hayes Andy Parker Mark Calleja Rizos Sakellari		X	X	
UNICAL	23	Antonio Congiusta Domenico Talia Paolo Trunfio Domenico Talia			X	
UOW	37	Vladimir Getov Ariel Goyeneche Tamas Kiss Gabor Terstyanszky Steve C. Winter	X		X	
UMUE	22	Sergei Gorlatch J. Mueller Martin Alt			X	
UNI DO	29	Ramin.Yahyapour			X	
UCO		Luis Silva		X		

Table 1: Participants of the IRWM Institute