Design Study

# European Grid Initiative Design Study (EGI_DS)

## EGI Consolidated Requirements and Use Cases

### Work Package 2

### EGI Requirements Consolidation

## EU DELIVERABLE: D2.1

| | |
|---|---|
| Document identifier: | EGI_DS_D2.1_v1.7.doc |
| Date: | **05/05/2008** |
| Work Package: | **WP2: EGI Requirements Consolidation** |
| Lead Partner: | **GRNET** |
| Document status: | **FINAL** |
| Document link: | http://web.eu-egi.eu/owl/browse.php?sess=0&parent=81&expand=1&fileid=231 |

Abstract: This document provides the essence of EGI_DS WP2 work, capturing the main EGI requirements as expressed by the involved communities. The WP2 work consists mainly of a call for contributions to a worked out list of stakeholders, namely the National Grid Initiatives, the application communities, the operations people, as well as other specific groups such as the networking, supercomputing and business communities. The processed findings of the above call for contributions, which was presented during the first EGI_DS workshop in Budapest, is also included, along with further useful feedback. Related WP2 material such as the EGI Knowledge Base, where all the EGI preparation material and input is stored, as well as the network of contact persons and links are also summarised in the deliverable. Note that a separate deliverable (D2.2) will focus on the EGI Knowledge Base.

## Delivery Slip

| | Name | Partner/ WP | Date | Signature |
|---|---|---|---|---|
| **From** | Fotis Karayannis | GRNET/WP2 | 29/2/2008 | |
| **Reviewed by** | Klaus Ullman<br>Juergen Knobloch | DFN<br>CERN | 25/3/2008<br>31/3/2008 | |
| **Approved by** | MB | | | |

## Document Log

| Issue | Date | Comment | Author/Partner |
|---|---|---|---|
| 0.1 | 15.10.07 | Draft Table of Contents (ToC) presented in WP2-WP3 meeting in Milano-Malpensa | Fotis Karayannis/GRNET |
| 0.2 | 28.11.07 | Updated ToC contents discussed during the WP2 conf call | Fotis Karayannis/GRNET |
| 0.3 | 7.12.07 | Updated ToC contents discussed in WP2-WP3 meeting in Munich | Fotis Karayannis/GRNET, Jan Kmunicek/CESNET |
| 0.4 | 17.12.07 | Final agreed ToC contents discussed in WP2 conf call | Fotis Karayannis/GRNET |
| 0.5 | 21.12.07 | First D2.1 contributions | Fotis Karayannis/GRNET |
| 0.6 | 7.1.08 | Fist D2.1 integrated draft | Fotis Karayannis/GRNET |

| | | | |
|---|---|---|---|
| 0.7 | 14.1.08 | Updated inputs from Jan Kmunicek, Diana Cresti, Panos Louridas, Per Oster, Katja Rauhansalo, Magdalena Wirtl | Fotis Karayannis, Panos Louridas /GRNET, Jan Kmunicek/CESNET, Diana Cresti/INFN, Per Oster, Katja Rauhansalo/CSC, Magdalena Wirtl, GUP |
| 0.8 | 28.1.08 | Applied agreed template and updated CESNET contribution | Fotis Karayannis/GRNET Jan Kmunicek/CESNET |
| 0.9 | 1.2.08 | Included executive summary and conclusions | Fotis Karayannis/GRNET |
| 1.0 | 15.2.08 | Included final version of CESNET contribution | Jan Kmunicek/CESNET |
| 1.1 | 19.2.08 | Editorial check | Fotis Karayannis/GRNET |
| 1.2 | 29.2.08 | Integration of final contribution, comments and reviews | Dieter Kranzlmueller/GUP Fotis Karayannis/GRNET |
| 1.3 | 25.3.08 | Document review | Klaus Ullmann/DFN |
| 1.4 | 31.3.08 | Second document review | Juergen Knobloch/CERN |
| 1.5 | 18.4.08 | Integration of review comments | Fotis Karayannis/GRNET |
| 1.6 | 24.4.08 | Integration of updated contributions from Diana Cresti (section 6-14/4/08) and Jan Kmunicek (reply to comments by Jesus Marco, 17/4/08) | Fotis Karayannis/GRNET |
| 1.7 | 29.4.08 | Uploaded document to the OWL and included the link in the document | Fotis Karayannis/GRNET |
| | | | |
| | | | |
| | | | |

**Document Change Record**

| Issue | Item | Reason for Change |
|---|---|---|
| | | |

# TABLE OF CONTENTS

## TABLE OF TABLES

# 1. INTRODUCTION

## 1.1. PURPOSE

This document is intended to provide an overview of the requirements gathering and consolidation of the main stakeholders of the European Grid Initiative (EGI) mainly expressed in the form of EGI Use Cases. This means how the EGI organisation should be used, offering services to its stakeholders (NGIs) and related communities (e.g. application communities). The document identifies the foreseen new EGI-NGI environment, the methodology to come up with the use cases, it summarises the use cases, and gives a first snapshot of the EGI Knowledge Base. In other words the document intends to cover the whole span of reported WP2 activities that all its partners have contributed.

## 1.2. DOCUMENT ORGANISATION

This document is organised in six main chapters. The Executive summary, the proposed new EGI-NGI environment, the results from the first EGI Workshop, the methodology and the collected EGI use cases, and a first introduction to the EGI Knowledge Base. In the executive summary, some main conclusions are drawn, along with a procedure for incorporating future feedback.

## 1.3. APPLICATION AREA

This document is a public deliverable intended to inform all the EGI related stakeholders, as well as readers not directly related with EGI (e.g. from other geographical areas) including the general public. Some knowledge and background is however required for readers to have a good understanding of the document.

## 1.4. REFERENCES

| | |
|---|---|
| R 1 | *Use Case in the EGI knowledge base http://knowledge.eu-egi.org/index.php/Use_Cases:Main* |

**Table 1: Table of references**

## 1.5. DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the main author or the project coordinator contact@eu-egi.org.

## 1.6. TERMINOLOGY

This subsection provides the definitions of terms, acronyms, and abbreviations required to properly interpret this document. A basic assumption that has been also agreed recently is that the term EGI (European Grid Initiative) refers to both the EGI organisation and NGIs, while the EGI organisation refers to the EGI headquarters, where the team with central responsibility is located, including the management structure. In addition, there is an Annex (Annex V) with more information on the use cases actors and objects. A useful glossary of terms can be also found under http://www.eu-egi.org/internal/glossary.

**Glossary**

| | |
|---|---|
| AB | EGI Advisory Board; This term is now superseded by the new term EGI Policy Board (EGI PB) |
| APEL | Accounting Processor for Event Logs; EGEE Accounting System |

| | |
|---|---|
| ARC | Advanced, Resource Connector; Middleware used in Nordugrid |
| BDII | Berkeley DataBase Information Index; The Information System service component |
| CA | Certification Authority; Main Grid authentication component |
| CE | Compute Element; A grid service |
| CREAM | Computing Resource Execution And Management; A Compute Element Implementation |
| DGAS | Distributed Grid Accounting System; An Accounting System used in Italy |
| D-Grid | Deutsches Grid; The German Grid Initiative |
| EFDA | European Fusion Development Agreement; is an agreement between European fusion research institutions and the European Commission to strengthen their coordination and collaboration www.efda.org |
| EGEE | Enabling Grid for E-sciencE; The major Grid infrastructure project for research |
| EGI | European Grid Initiative; The new EGI-NGI environment |
| EGI organisation | The legal body that will host the EGI headquarters; It includes the team with central responsibility, as well as the management structure |
| EMBL | European Molecular Biology Laboratory; major research center coordinating molecular biology research. |
| ESR | Earth Science Research; One of the disciplines deployed in the Grid environment |
| EUGridPMA | European Grid Policy Management Authority; The coordinating body of the national Certification Authorities in Europe |
| FTS | File Transfer Service; A grid service |
| GIIS | Grid Information Index Services; Element of the OGF information system |
| GridIce | A distributed Grid Monitoring tool |
| Gstat | Grid Status: An Information System Monitoring Tool |
| GSVG | Grid Security Vulnerability Group; A group aiming at eliminating any Grid security vulnerabilities in the Grid middleware and its deployment |
| HEP | High Energy Physics; One of the disciplines deployed in the Grid environment |
| IGTF | International Grid Trust Federation; The coordinating body of the Grid Authentication which sits on top of the EuGridPMA |
| ITER | Originally an acronym for International Thermonuclear Experimental Reactor; an international project to design and build an experimental fusion reactor |
| LB | Logging and Bookkeeping; A grid service |
| LFC | LCG File Catalogue ; A grid service |
| LHC | Large Hadron Collider; A major research infrastructure facility |
| LCG | LHC Computing Grid project |
| MonBox | Monitoring Box; A Grid monitoring service providing information on which resources are available on the Grid |
| MoU | Memorandum of Understanding |

| | |
|---|---|
| MyProxy | A Credential Management Service; A proxy for managing PKI security credentials |
| NGI | National Grid Initiative; The main building block of EGI |
| OAG | Operations Advisory Group; An EGEE group co-chaired by the Operations and Applications activities |
| OCC | Operations Coordination Centre; The coordinating structure of the EGEE operations activity |
| OGF | Open Grid Forum; The main Grid standardisation body |
| OMII-Europe | Open Middleware Infrastructure Institute – Europe; An EC-funded FP7 project working towards middleware standardisation |
| OSG | Open Science Grid;The US grid infrastructure analogous to EGEE |
| PPS | Pre-production service ; An EGEE service for deploying early releases of the Grid middleware |
| RA | Registration Authority; A building block of national Certification Authorities |
| RB | Resource Broker; A grid service |
| RC | Resource Centre (grid site) |
| R-GMA | Relational Grid Monitoring Architecture; A grid service |
| ROC | Regional Operations Centre- Structure for regional coordination of operations |
| SRM | Storage Resource Manager; An interface to Grid storage |
| Unicore | Grid Middleware used for supercomputing purposes |
| VDT | Virtual Data Toolkit; An ensemble of Grid software developed in the US |
| VO | Virtual Organisation; A main Grid building block gathering users of the same community |
| VOMS | VO Management (or membership) Service; An Authorisation service |
| WISDOM | Wide In Silico Docking On Malaria; Initiative for drug discovery for neglected and emerging diseases |
| WLCG | Worldwide LCG Collaboration; The name of the LCG project |
| WMS | Workload Management System; A grid service |
| WN | Worker Node; A grid service |
| WP | Work Package; A structure in the work plan of the project |

## 2. EXECUTIVE SUMMARY

Deliverable D2.1, entitled EGI Consolidated Requirements and Use cases, provides the essence of the EGI_DS WP2 work (EGI requirements consolidation) aiming at ensuring that the EGI organisation takes into account the requirements posed by the different actors involved with its development and usage. The main such actors are the National Grid Initiatives, application communities and Virtual Organisation leaders, operations people, as well as other specific groups representatives from the networking, supercomputing and business communities and coordinators of related projects. In particular, the EGEE project has been running a production infrastructure for almost 4 years having a very good understanding of the requirements. The EGI WP2 work has amongst other sources been based on the EGEE findings and experience of key personnel going one step further, updating, adapting and consolidating existing requirements into the new foreseen environment, i.e. of a central EGI organisation and of an NGI in each country.

The *organisation* of the D2.1 followed the WP2 structure, so the outcome of the three WP2 tasks and WP2 milestone is documented in this deliverable. The EGI-NGI's new environment and the list of NGI contact points (given as an annex) was the main outcome of Task 2.1 entitled "Establishment of official NGI and other related contact points". The requirements and use cases presentation is the outcome of the Task 2.3 entitled "prioritised requirements and use cases". It should be noted here that prioritisation will at the end take place as part of WP3 work in order to have a better picture, and the WP2 team will continue to work under WP3. The development of knowledge is obviously the outcome of corresponding Task 2.2 entitled "Development of an NGI Knowledge Base", which is summarised in this deliverable. In addition, a separate deliverable (D2.2) provides more details on the technologies and schema of the knowledge base. Finally the main WP2 milestone, entitled "Workshop WS1: EGI Requirements Consolidation and Use Case Definition (NGI evolution and the road towards EGI)" is also documented in D2.1.

The summary of the WP2 work and its findings documented here are the following:

- A list of actors related to the EGI-NGIs has been identified in an "actors model", i.e. a diagram capturing the actors and their main interactions. Lists of contact persons have been gathered, including above all the official NGI representatives list nominated by their countries.
- A call for contributions to the main identified actors/stakeholders, which was prepared by the WP2 team and sent to them in September 2007 in order to take into account their feedback. The call for contributions was mainly performed by collecting use cases for EGI (and NGIs) and included a specific template. In the letter some more questions on the evolution towards EGI-NGIs were included, leaving it to each contacted representative to express related views.
- The input received was processed and summarised by the EGI WP2 team and uploaded to the Knowledge Base in different forms. Use cases were categorised and clustered in various ways (e.g. according to the main EGI functionality or according to the EGEE activities). In addition, the original use cases were also included after getting the permission of the authors. It is felt that the sample is representative, although only around 1/3 of the NGIs did provide feedback. The total amount of use cases received until the end of February 2008 was 174, which is obviously a sufficient number. All the material is available in the web (http://*knowledge.eu-egi.org/index.php/Use_Cases:Main)* [R 1]
- The first findings of the call for contributions, were presented during the first EGI_DS workshop in Budapest in order to incorporate useful feedback. A summary of the first EGI workshop is included in section 4. Feedback and use cases were indeed received even as late as February 2008, but the majority of the use cases was received in the first wave. It has been agreed that further use cases will be accepted and their impact will be assessed in the WP3

work (EGI functionality definition) and if deemed appropriate it will be taken into account in the EGI specification.

- The major functions derived from the executive summary were the following:
    - Operation of a reliable Grid infrastructure
    - Application Support
    - Coordination of middleware development and standardization
    - Development and operation of build and test systems
- A section is devoted to the Knowledge Base, mainly dealing with its overview and sections, so as to have a complete view of the WP2 work. A separate deliverable on the Knowledge Base (D2.2) however is also available focusing on all the KB aspects.
- A rich set of annexes is finally included including also a snapshot of the NGIs' representatives at the beginning of 2008, as well as annexes related to the collection of use cases.

# 3. THE NEW EGI-NGIS ENVIRONMENT

## 3.1. ACTORS AND COMMUNITIES

The future EGI organization is intended to provide production grid services to the research communities in Europe. As a first step, it is important to identify relevant communities in order to obtain their requirements for the EGI-NGIs environment, hereafter referred to as the EGI system.

A natural distinction in respect to the EGI actors and communities relies on those providing the service and those using the service. On the provider side, it is possible to identify e.g. grid operators, middleware developers, and application support people. On the users' side, we see the range of different application communities, including those that are already using the grid and those, which may be grid users in the near future. Each of these groups has different requirements and expectations on the EGI organization, and it is essential that their views are incorporated into the preparations of EGI.

In order to get feedback for the EGI requirements consolidation work, a set of stakeholders and end user communities have been identified. These are called actors, which interact with the EGI system in one way or another. In addition, the scheme of these actors interacting with the EGI system is called the actors model – see Section 3.2. In a way, this resembles the usual software development process, when trying to design a system, and we tried to emulate the related system of actors, getting requirements, designing an interaction model and then designing the system functionality. The latter is going to take place inside EGI_DS Work Package 3.

The main stakeholders of the EGI organization are the National Grid Initiatives (NGIs), which provide the basis of the EGI production grid infrastructure. The future EGI organization can be seen as glue between the individual NGIs, improving the usage of the grid across Europe's internal boundaries. Each of the NGIs fulfils a certain set of criteria as follows.

Each NGI:

- should be a recognized national body with a single point-of-contact
- should mobilise national funding and resources
- should operate the national e-Infrastructure
- should support user communities (application independent, and open to new user communities and resource providers)
- should contribute and adhere to international standards and policies

It is expected that each country in Europe will sooner or later establish its own NGI entity that will take up these responsibilities. Within EGI_DS, it will be important to define the share of responsibilities between EGI and the NGIs, which will be done in WP3. The exact legal requirements for each NGI will be determined by WP4.

The NGIs are actively contributing to the EGI_DS not only by providing their input to WP2, but also as decision makers on the EGI Advisory Board (AB)[1], which is the board of the NGI representatives.

[1] The new agreed term for the EGI Advisory Board is EGI Policy Board

During the setup of the project, the EGI AB has been created by asking for nominations from the NGIs across Europe. A snapshot of the NGI representatives in the beginning of 2008 is included in Annex I.

Apart from the NGIs, international organizations such as CERN (European Organization for Nuclear Research), EFDA (European Fusion Development Agreement), EMBL (European Molecular Biology Laboratory), ESA (European Space Agency), ESO (European Organisation for Astronomical Research in the Southern Hemisphere), ESRF (European Synchrotron Radiation Facility), and ILL (Institut Laue Langevin) have been identified as possible stakeholders. The EGI_DS will identify a possible means of integrating international organizations in the structure of the EGI. In addition, EGI_DS should foresee mechanisms to include future stakeholders, which may become relevant in the course of time.

Additional actors in the EGI are the many existing and upcoming projects (national and international), which will either utilize the EGI structure or extend it for their particular domains. These include existing projects such as the major e-Infrastructure projects DEISA, EGEE, GN2 and PRACE and their communities (supercomputers, research networks, grids), but also smaller projects for a dedicated application domain, e.g. BioinfoGRID or Health-e-Child. Given that Grid standardisation is of vital importance for EGI, the OMII-Europe project should be added in the list of major e-Infrastructure projects. Major Research Infrastructure facilities listed in the ESFRI roadmap are also potential users of EGI. Finally, the EGI organization should establish links with policy makers and funding agencies, which need to understand the benefits of EGI for the research community and their own responsibility in shaping the future grid in Europe.

## 3.2. THE ACTORS MODEL

The actors' model has been developed, trying to show all the actors involved with the EGI system under study including their interactions.

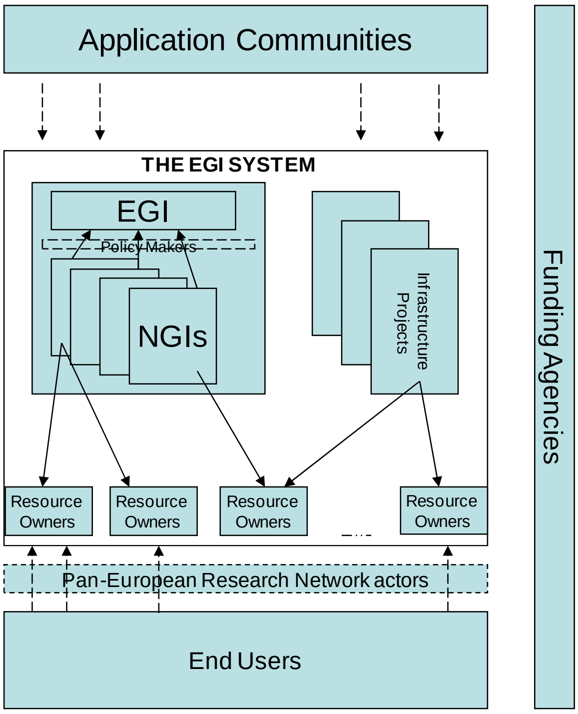



**Figure 1 – The EGI system**

Boxes (i.e. objects) represent organisations (or forms of collaborations), while human drawings represent interactions among related actors. A first simplified model of the envisaged system is presented below.

The EGI system is the central box encompassing the EGI and NGIs, as well as the other infrastructure projects, which establish some kind of relation to EGI. To avoid any misinterpretation, e.g. that EGI owns the other infrastructure projects, it should be stated that those infrastructure projects are independent from EGI, and the only reason that they appear is because they have some kind of relation to EGI. If e.g. an infrastructure project on Grid middleware standardisation is accepted at some point in FP7, this has a clear relation to EGI and should be kept in the overall picture. DEISA or PRACE are also infrastructure projects, which own resources, so this option is made possible through the arrows to the resource centres.

The main categories of actors are related to these two categories of objects, i.e. the EGI organisation and the NGI organisations. All the other shapes around this central box are objects interacting with the main system. Individual end users are shown in the bottom or the picture, while organised application communities expressed as Virtual Organisations, can be found on the top of it. On the right side we see the funding agencies objects and actors appearing, which are relevant to all the stakeholders (i.e. funding agencies of all the boxes). Other actors include policy makers, which can belong to another category of actors (shown in dotted lines). Policy makers can be also form separate actors (such as the e-IRG body), which are not shown in the first figure, rather in the more detailed one that follows.

In the latter, represented in figure 2, further stakeholders are shown, such as the supercomputing and industry or business community ones having some direct or indirect relation with the system. In addition, all the envisaged actors around the EGI system and its interacting systems appear. A detailed analysis of further actors and objects has also been produced based on the EGI use cases and can be found as an annex. This acts as a summary of the main categories of objects and users.

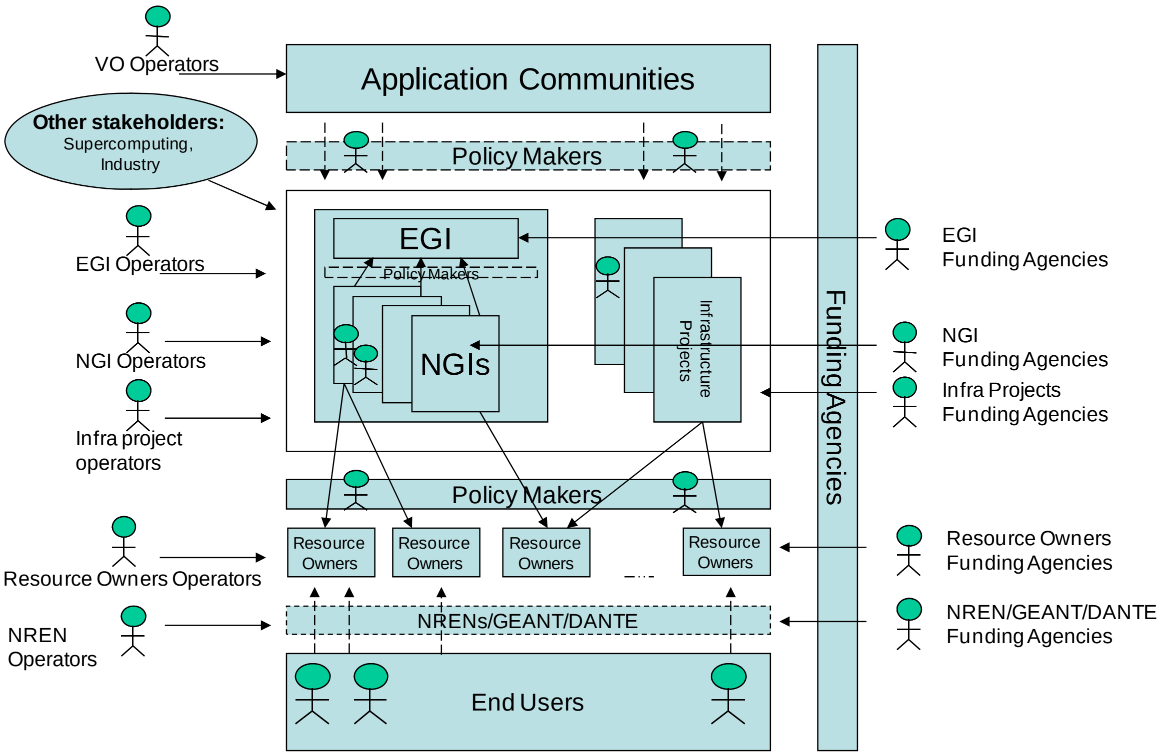



**Figure 2 – The envisaged EGI Actors model**

Note that the main objective of the above model is to get an idea of the overall system and the actors interacting with it, and not to have an exhaustive list of all possible objects and actors. Standardisation bodies are obviously important objects (with their related actors) acting as another vertical layer (similar to the funding agencies), but have not been included in the figure in order not to complicate the already loaded figure.

# 4. EGI FIRST WORKSHOP REPORT

The first EGI Workshop took place on October 2nd in Budapest, Hungary as part of the EGEE 2007 conference. The EGI Workshop brought together representatives of the involved stakeholders aiming at assessing the status of National Grid Initiatives (NGIs) and the progress towards the establishment of the European Grid Initiative (EGI) organisation. Moreover, following the call for requirements and use cases contributions sent out in September 2007, the initial set of input in the form of use case descriptions has been presented to the different stakeholders/actors and comments have been solicited for the next round of contributions and requirements consolidation. Detailed information including all the presentations can be found at http://web.eu-egi.org/events/workshops/oct07/

## 4.1. GOAL OF THE WORKSHOP

Besides a general presentation of the EGI-DS project, the first EGI workshop concentrated on the consolidation of EGI requirements and the definition of the use cases. The workshop aimed at bringing together representatives of the involved parties and to assess the status of NGIs and the progress towards the establishment of an EGI organisation. The initial set of use cases was presented and input and comments were solicited.

The workshop consisted of three separate sessions. Sessions one and two were open to all public and the third session was reserved only for the NGI representatives, i.e. members of the EGI Advisory Board (AB).

## 4.2. MAIN ISSUES AND RESULTS

The *first session* presented an overview on the EGI Design Study and its operations. Michael Wilson from STFC chaired this session.

- Dieter Kranzlmueller from GUP made a presentation on the EGI_DS present status and project plans. The goal of the European Grid Initiative was presented as a long-term sustainability of production grid infrastructures in Europe achieved by establishment of a new federated model bringing together the NGIs building the EGI organisation. The EGI organisation was defined as a coordinating and operationally common multi-national and multi-disciplinary grid infrastructure in order to enable and support pan-organisational grid-based collaboration, to provide support and added value to NGIs and to liaise with corresponding infrastructures outside Europe.

- Fotis Karayannis from GRNET spoke about the requirements of the EGI_DS project, and stated as an objective to ensure that the future EGI organisation takes into account the requirements posed by the NGIs and other related stakeholders in a prioritised way. He stated that in order to achieve the set of objectives, the update and consolidation of the existing preparatory work (EGEE, e-IRG) have to be completed. The major stakeholders (to start with NGIs and application communities) should be consulted and a knowledge base for NGIs and EGI has to be developed. Also the use-cases for EGI have to be collected and the conclusions of this feedback should be incorporated in all tasks.

- Diana Cresti from INFN did an overview on the definitions of EGI functions with respect to the NGIs at the start of EGI. She explained that to define the core of functions information would be collected on use cases, requirements of the NGIs and best practices presently implemented by different actors in grid computing. A model on how to implement these functions will be drafted including legal implications and identification of rules and guidelines for the NGIs. A model of how to implement these functions and to evaluate the legal form of EGI bodies will follow. The areas of activities were initially divided into the middleware, the operations and the networking

- Klaus Ullmann from DFN presented the federated research network operations in Europe through NRENs and DANTE by concentrating on their organizational and financial challenges. He highlighted the crucial question for EGI by asking how the EGI business cases (use cases) will be constructed. Will there be EGI budgetary driven grid operation or will the actual grid operations be federated more through the NGIs? He summarized also that the highest priority in EGI must be on the definition of the use-cases (business cases). He also asked how the NGIs would develop in the future.

The *second session* concentrated on the EGI requirements and the use-cases. Jürgen Knobloch from CERN chaired this session.

- Diana Cresti presented the new EGI Knowledge Base (http://knowledge.eu-egi.org), its functions and the internal pages. She emphasized the role of the Knowledge Base as a collaborative tool that uses MediaWiki and is edited by EGI_DS and the NGI representatives.

- Ludek Matyska and Jan Kmunicek from CESNET reported about the collection of the EGI use-cases. The primary contributors were the NGI representatives and secondary the EGEE application domain representatives, EGEE related project leaders and EGEE ROC managers. The managers of the EGEE Virtual Organizations were also addressed. The feedback will be collected and processed, the requests for specific comments and new use cases will be produced and the list of respondents will be modified.

The *third session* was closed and reserved for the NGI representatives and members of the EGI Advisory Board. The main discussion was about the EGI Advisory Board and its role in EGI_DS. The following issues were agreed:

- NGIs could contribute to EGI_DS, while EGI_DS should provide support to them.
- Selected experts from the NGIs should be invited to EGI_DS meetings in order to accelerate the consensus building process.
- User accounts for the knowledge base should be distributed on 3 October 2007
- The next NGI representatives meeting (i.e. AB meeting) will take place in March 2008 in Italy during the second EGI Workshop.

## 4.3. CONCLUSION

Both the workshop and the closed Advisory Board meeting was very well attended with 111 persons in the two workshop sessions and 37 NGI representatives present in the AB meeting. It is obvious that the EGI_DS is of highest interest to the NGIs. In the meetings there was a whole range of different interests represented; from newly established NGIs with small but very active and successful research groups that want to use grid resources, to the well-established NGIs with founders of LCG and other large grids as their stakeholders.

It was clear from the discussions that there are many different aspects EGI_DS has to take into account when proposing the EGI organisation. And the major conclusion is that openness towards the NGIs in the design study process is crucial to the success of EGI-DS.

# 5. REQUIREMENTS AND USE CASES PRESENTATION

## 5.1. BACKGROUND

The current Grid infrastructures like EGEE are usually built on a project basis. This approach does not guarantee long term sustainability, a necessary condition for users to commit themselves to use it and to rely on its existence in their own research. As the Grids can be seen as an extension of the global network infrastructure, we expect some similarities in the organization and sustainability guarantees. However, as there are also fundamental differences between Grids and networks, including the still wide spectrum of grid middleware and lack of a generally accepted layer resembling the IP in computer networks, we need to better understand the providers and users needs, requirements and expectations. Only with this input, using also the already gained experience we will be able to shape Grid environments into a sustainable model. To get such insight, users must be approached, so that their feedback is obtained. Although there are several approaches to get such feedback, the EGI team decided to exercise the use case collection, asking users, providers, and other actors to share their experience and expectations in a form of scenarios.

As the EGEE Grid currently represents a worldwide infrastructure utilised by many distinct application domains, the first apparent step towards collecting end user needs and requirements was the EGEE survey performed in 2006 by EGEE NA5 activity working on Policy and International Cooperation. The results of this survey were presented at EGEE conferences and currently they are stored and freely available[2] at EGI Knowledge Base (KB). The performed survey serves as a first step giving us a generic overview of trends within the end users communities. However, this collection is limited, as it was focused primary on the experience with and expectations of EGEE infrastructure, while the more general sustainable grid may in some aspects differ from the EGEE. The next logical step is therefore to approach existing and potential Grid users and researchers, but also existing and potential Grid resource providers, and collect from them use cases of Grid deployment and usage. The typical practice, their daily experience, usual common situations encountered when deploying and using Grids as well as their wishes and expectations of the role, functions and properties of the future Grid infrastructure will be thus identified, collected, and analyzed.

The term *use case* concerning grid computing should be understood as any information about any aspect of the Grid operation, management, monitoring and use by end users or other actors (including virtual organization setup and operation). Moreover, it can also involve related aspects like middleware, relationship to industry, governmental and other bodies and interaction with other infrastructures.

## 5.2. METHODOLOGY FOR USE CASES COLLECTION PROCESSING AND DISSEMINATION

As a first step towards the EGI mission, the potential actors in the new planned Grid environment had to be identified. Subsequently, their requirements in the form of use cases of Grid deployment and usage had to be collected.

Basically, there are two main groups of key Grid players. On one side, there are Grid infrastructure managers (owners) and administrators who build and take care of the Grid environment. On the other side, there are many distinct end user communities utilizing the Grid environment for their daily research work within miscellaneous application domains. In addition, there is the group of Grid tool developers; also the users are knowledgeable of Grids at very different levels of expertise and

[2] http://knowledge.eu-egi.org/index.php/2006_NGI_survey

expectations. Moreover, there is organizational (hierarchical) aspect, too. The separation is based on functional distinction and involves national Grid operators, computational resource owners, virtual organization administrators and members, funding agencies, policy makers, standardization bodies, specific institutes or even individual end users.

At the more explicit level, the conception of EGI DS project went parallel with the establishment of National Grid Initiatives (NGIs) in individual countries. The NGIs are usually composed from resource owners, but as the whole concept is new and evolving, also user representatives are present in many NGIs constituencies. This makes NGI representatives crucial actors in the future Grid infrastructure, their opinion, expertise, and expectations/requirements being very important. Notably without NGI endorsement there will be no sustainable Grid infrastructure.

As mentioned earlier, the Grid infrastructures rely on the underlying network infrastructure. This obviously causes networking projects and representatives to become other important actor. Similarly, the supercomputing facilities and corresponding projects have a lot of expertise that should not be left out.

While taking into account this heterogeneity, the EGEE project is considered to form the current worldwide etalon of Grid infrastructure. This means that EGEE personnel, especially leaders of individual activities, also belong to the important actors. EGEE already created many VOs, so it also offers a gateway to huge number of current Grid users.

The list of identified Grid actors has been divided into several groups that represent different categories of Grid communities. Table 1 summarizes Grid communities that have been later approached.

| | |
|---|---|
| NGI representatives | Leaders of National Grid Initiatives in Europe |
| EGEE application domains | Representatives of influential application communities |
| EGEE ROC managers | Operators of geographical regions of EGEE |
| EGEE related projects | EGEE Grid-related projects leaders |
| EGEE VOs | Virtual organizations representatives of the EGEE project |
| Supercomputing area | Representatives of supercomputing-related projects |
| Networking area | Representatives of European networking-activities |

**Table 2 – Identified Grid communities**

The next step consisted of contacts with the identified actors. Obviously, there are several ways how to collect the required information. One can oscillate between two extremes. The first one is represented by the situation when input in any form is accepted. While this approach minimizes a risk of pre-selection, it makes further analysis an eventual aggregation of results very difficult. Also, some actors may not understand the request and the answers may not be properly targeted, making the material of lower value. The other extreme is a questionnaire, with specific questions and prescribed form of answers—this has been used for the first round of use case collection within the EGEE project. The questionnaire has an associated risk of missing important areas that had been omitted or forgotten by the questionnaire authors. Also, actors with differing views (from the questionnaire authors) may be reluctant to contribute to a process they see as biased. Therefore, we decided to follow a less restricted approach, where actors received recommendations on input (but not a direct questionnaire), including examples of use cases we consider interesting. A **template** has been created, to help further with the preparation of use case description. On the other hand, we also offered the opportunity to use a completely free format in cases the individual use case contributor feels to restricted by our recommendations; also while proposing some major areas of interest we left on actors to pick up any area they deem as important or interesting. This way allowed freely accepting any ideas, requirements, comments and suggestions arising from Grid communities rather than forcing our views to the involved actors.

To achieve the defined goal a so-called **"use case collection letter"** was drafted[3] and subsequently sent to several distinct groups of key Grid actors identified so far. The prepared letter simultaneously served as a template containing guidelines helping the letter recipients in constructing their use cases (see Annex II and III for detailed content). Within the use case letter three specific use case areas were suggested—infrastructure, applications, and end user position. These areas are foreseen as the critical components of the sustainable Grid environment and the way how these areas are dealt with will have a strong impact on its shape.

The distribution of prepared use case letter to previously identified Grid actors/communities has been performed in several steps. At first, a personalized version of the use case letter has been sent to the primary audience, all the NGI representatives. Secondly, a generic use case letter has been sent to the EGEE application domains representatives, EGEE ROC managers and EGEE related projects leaders. To disseminate the generic use case letter within the EGEE virtual organizations we used a CIC portal broadcast tool[4] allowing us to spread the use cases collection request among all EGEE VO managers.

## 5.3. STATISTICS OF COLLECTED INPUT

The majority of use cases arrived in the form of the distributed template or in a template-like form retaining suggested distinction into infrastructure-oriented, application-oriented and end user-oriented use cases. Moreover, those in non-template form were easily transformable. Quantitatively speaking, **28 contributors** have provided us their opinions and concerns of establishment and management of the sustainable Grid environment. Those contributions contain **in total 174 individual EGI use cases** (as majority of the contributions contained multiple use cases). The detailed statistics can be found in Table 2.

| | no. of contributors | no. use cases |
|---|---|---|
| NGI representatives | 11 | 112 |
| EGEE application domains<br>EGEE ROC managers<br>EGEE related projects | 16 | 57 |
| EGEE VOs | 1 | 5 |
| **Total** | 28 | **174** |

**Table 3 - Use case input statistics**

[3] http://www.eu-egi.org/public/EGI_Use_Case_Letter.pdf

[4] https://cic.gridops.org/

The first inspection of the use cases collected revealed that some topics are mentioned more frequently. The following issues belonged to the most frequent ones:

- Interactions (access, monitoring, etc.) with the Grid infrastructure
- Definition of necessary procedures/policies/service level agreements
- Commercial application licensing issues
- Intensive support for parallel run of computational jobs
- Middleware development plus maintenance

On the other hand, the preliminary analysis of the use cases content also showed some content that was not directly expected (this confirmed the value of more free form approach to the use case collection). The most interesting not expected use cases deal with issues related to

- Closed community (private companies)
- Single user access to the infrastructure

### 5.4. USE CASES PROCESSING AND SUMMARISATION

The processing of obtained contributions was divided into several subsequent steps. First rough insight into the material was received through the identification of basic patterns. This allowed us to see what areas were mentioned and how many contributors considered a specific topic important. Based on this information a mechanism for detailed analysis has been implemented—we highlighted specific use cases patterns and distributed them into specific categories with subsequent assignment of respondents into these categories.

This allowed us to obtain a more compact view of specific issues and their relative importance from the respondent's point of view. The majority of these topics covers the same aspects of Grid infrastructure building and management that was already expected and mentioned in the use case letter. This can be illustrated on the extracted main topics covered in the raw use cases:

- Infrastructure access and operations
- Middleware development/deployment/enhancements
- Application scenarios

This list corresponds well with the anticipated significant areas in the use case letter that were focused in the infrastructure-related (infrastructure access and operations, resources monitoring/evaluation) and application-related issues (small communities/private sector, applications scenarios). All of them are logically complemented by middleware development/deployment/enhancements ensuring continual middleware progress and implementation of new features. Based on the primary information a geographical distribution of use cases origin dealing with the main topics could be evaluated and is indicated in the next figures.

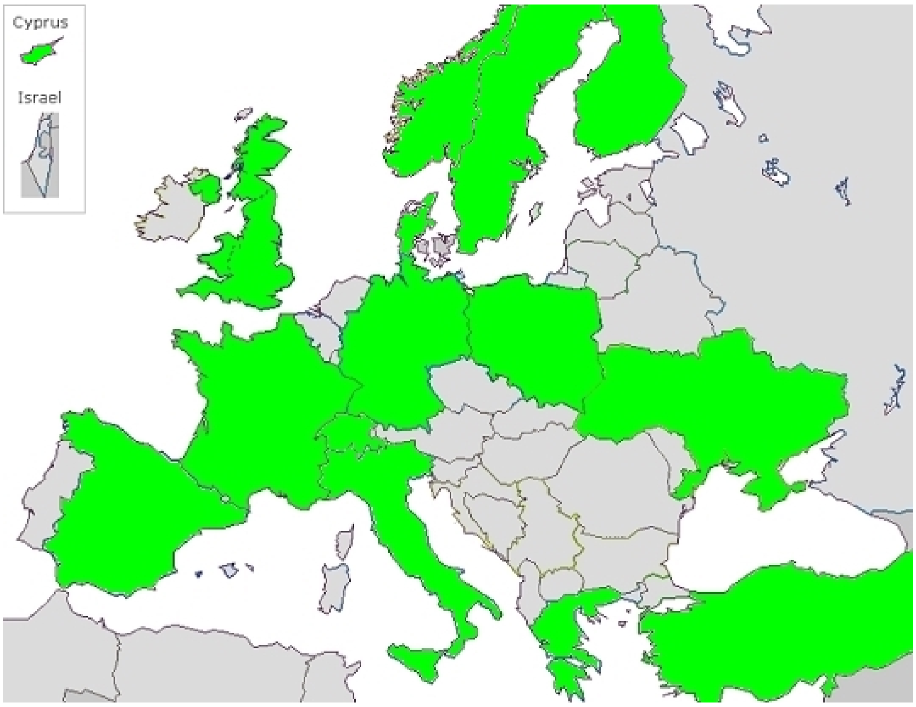



**Figure 3 – Infrastructure access and operations**

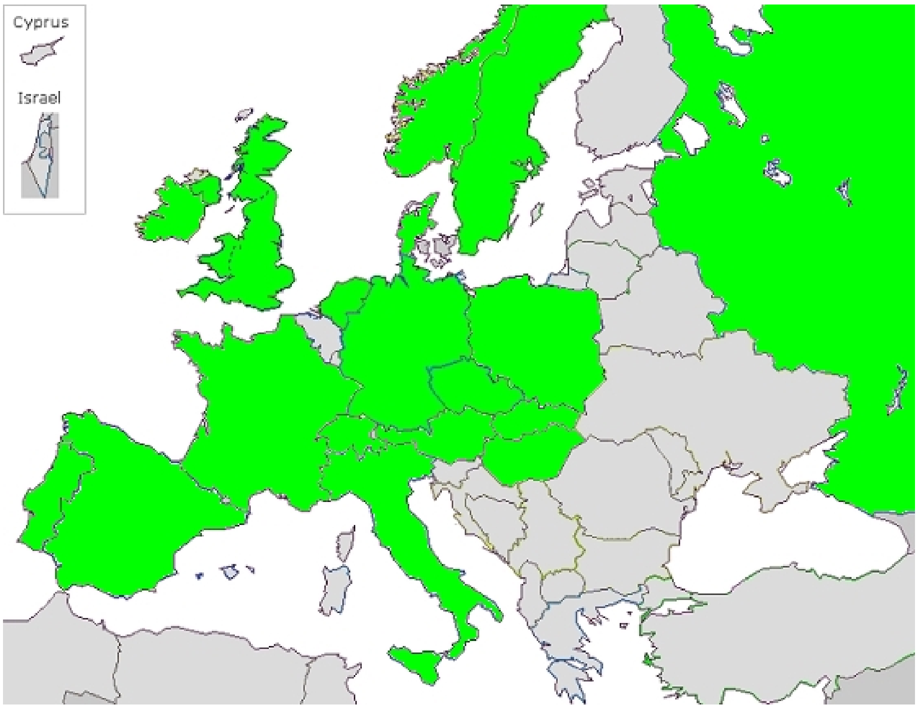



**Figure 4 – Middleware development/deployment/enhancements**

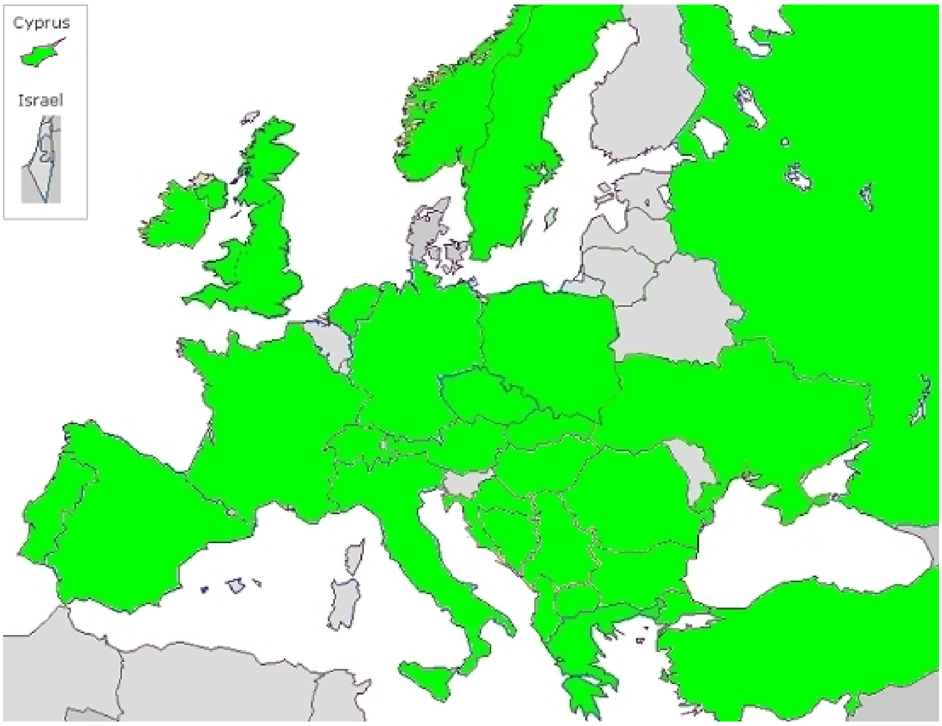



**Figure 5 – Application scenarios**

The use cases are collected with the primary goal of providing us with base requirements on the future Grid environment, its function, properties and shape. However, as the understanding of the principles of Grid operation evolves, new areas may become important not foreseen in the first set of use cases. It is therefore important not only to process the collected use cases, but also to make them available to all actors and potential contributors of new and modifications to existing use cases. To fulfill this function, the use cases have to be publicly, freely available for all Grid actors to allow for continuous contribution with further ideas and inputs. To ease this task, we reformatted the use cases into an easily readable and accessible way, giving each a unique identifier for further reference.

All individual use cases were reformatted into a textual form and were put into the EGI Knowledge Base (KB) (see Figure 6). The individual text use cases are exposed in the EGI KB through single pointers. Each use case is stored as a separate pre-processed file in a textual form that makes reference to an individual use case very easy. We used consistent naming of use cases in such a way that a use case name is created according to the use case origin (contributor). This approach makes navigation more intuitive and straightforward. The use cases are now freely available for all potential contributors willing to comment, modify or extend them[5]. At the same time, the collection of use cases used for the analysis presented in this deliverable is kept frozen and is a digital (virtual) appendix to this deliverable.

[5] http://knowledge.eu-egi.org/index.php/Individual_use_cases_gathered

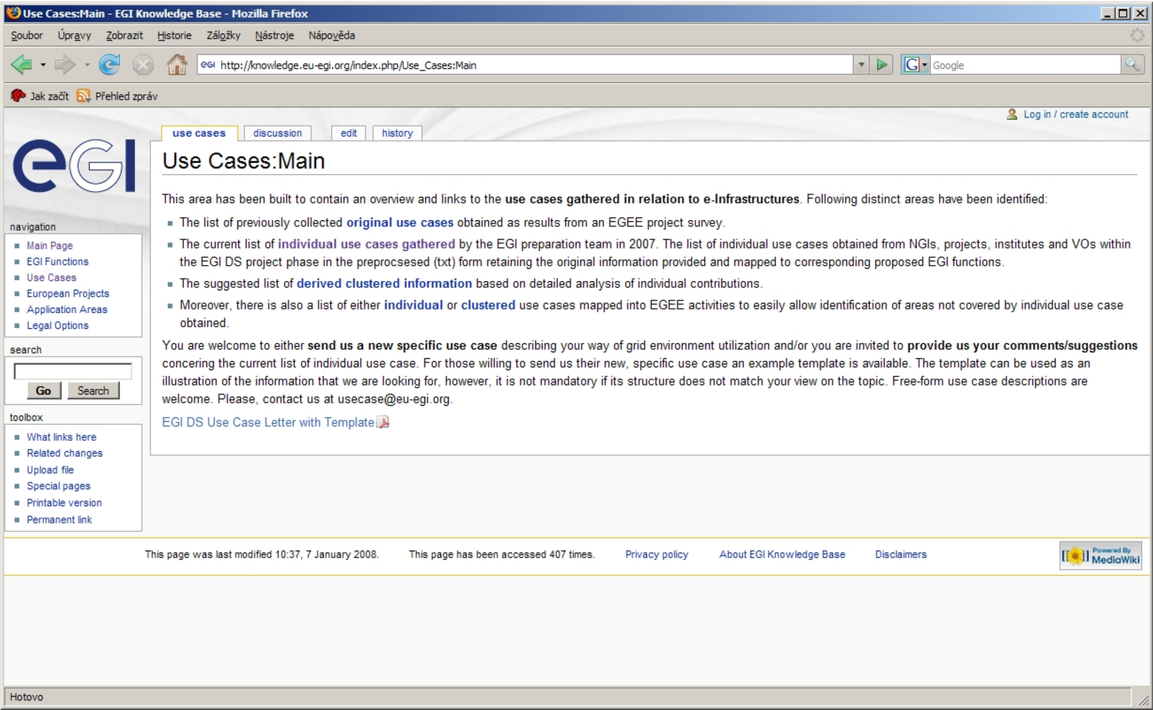



**Figure 6 - Use cases within EGI DS Knowledge Base**

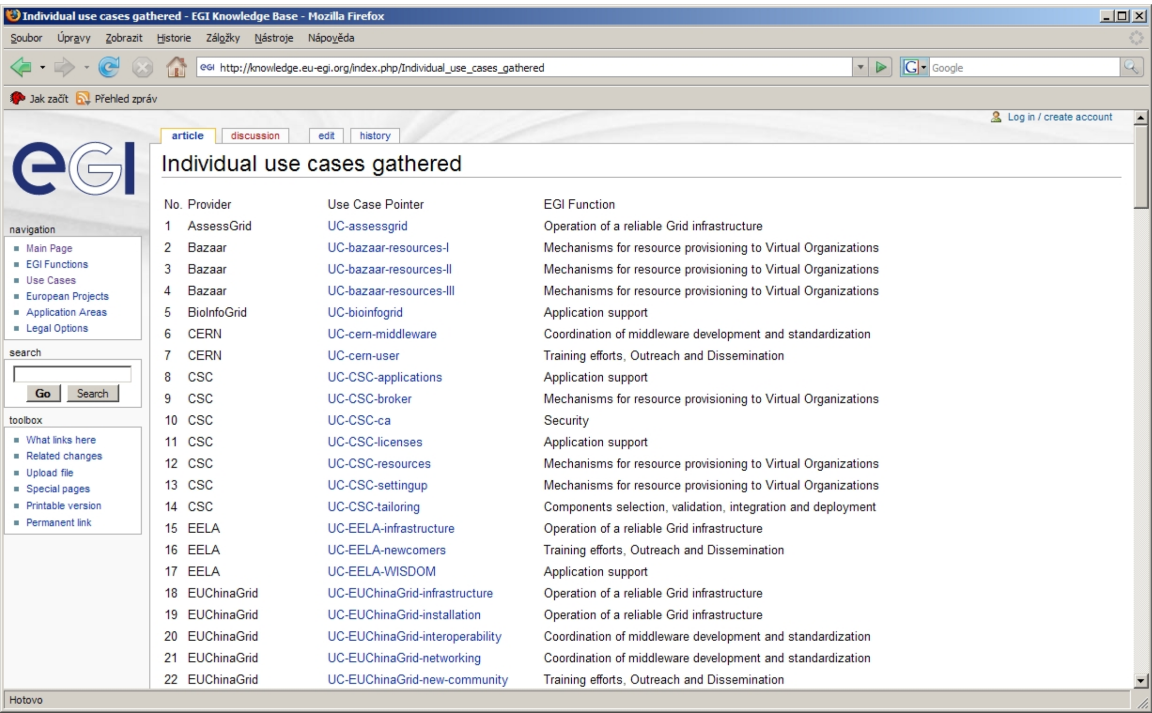
To further aggregate the information within use cases we derived a skeleton of issues to which all use cases have been matched. This hierarchical tree[6] represents critical areas that have been identified as cornerstones of any future Grid infrastructure attempting to achieve complete sustainability. Derived areas involve following key fields:

1. Middleware
    1.1. Smooth integration of new components
    1.2. Component lifecycle
    1.3. One or more MW stacks
    1.4. Support of interactive use
    1.5. Security issues
    1.6. Large redesign is required
    1.7. Other specific functionality requests

2. Infrastructure and Operation
    2.1. Joining the Grid
    2.2. Working with the Grid
    2.3. Organization of operations
    2.4. Policies
    2.5. Related commercial issues

[6] http://knowledge.eu-egi.org/index.php/Derived_clustered_information

The aggregated results (for details see the Knowledge Base) indicate that several issues have been considered repeatedly, but there might be some that were not even mentioned. Therefore, the next logical step involved the identification of missing areas. This challenge has been tackled from two different directions—first, the clustered use cases were mapped into EGEE project internal activities to easily allow identification of areas not covered by individual use cases collected. The matchmaking towards internal EGEE activities serves as methodological check preventing unwanted exclusion of specific areas or issues.



Individual use cases gathered

| No. | Provider | Use Case Pointer | EGI Function |
|---|---|---|---|
| 1 | AssessGrid | UC-assessgrid | Operation of a reliable Grid infrastructure |
| 2 | Bazaar | UC-bazaar-resources-I | Mechanisms for resource provisioning to Virtual Organizations |
| 3 | Bazaar | UC-bazaar-resources-II | Mechanisms for resource provisioning to Virtual Organizations |
| 4 | Bazaar | UC-bazaar-resources-III | Mechanisms for resource provisioning to Virtual Organizations |
| 5 | BioInfoGrid | UC-bioinfogrid | Application support |
| 6 | CERN | UC-cern-middleware | Coordination of middleware development and standardization |
| 7 | CERN | UC-cern-user | Training efforts, Outreach and Dissemination |
| 8 | CSC | UC-CSC-applications | Application support |
| 9 | CSC | UC-CSC-broker | Mechanisms for resource provisioning to Virtual Organizations |
| 10 | CSC | UC-CSC-ca | Security |
| 11 | CSC | UC-CSC-licenses | Application support |
| 12 | CSC | UC-CSC-resources | Mechanisms for resource provisioning to Virtual Organizations |
| 13 | CSC | UC-CSC-settingup | Mechanisms for resource provisioning to Virtual Organizations |
| 14 | CSC | UC-CSC-tailoring | Components selection, validation, integration and deployment |
| 15 | EELA | UC-EELA-infrastructure | Operation of a reliable Grid infrastructure |
| 16 | EELA | UC-EELA-newcomers | Training efforts, Outreach and Dissemination |
| 17 | EELA | UC-EELA-WISDOM | Application support |
| 18 | EUChinaGrid | UC-EUChinaGrid-infrastructure | Operation of a reliable Grid infrastructure |
| 19 | EUChinaGrid | UC-EUChinaGrid-installation | Operation of a reliable Grid infrastructure |
| 20 | EUChinaGrid | UC-EUChinaGrid-interoperability | Coordination of middleware development and standardization |
| 21 | EUChinaGrid | UC-EUChinaGrid-networking | Coordination of middleware development and standardization |
| 22 | EUChinaGrid | UC-EUChinaGrid-new-community | Training efforts, Outreach and Dissemination |

**Figure 7 – Individual use cases mapped into expected EGI functions**

Concurrently, the use cases were also mapped into expected EGI functions derived as the preliminary list[7] proposed by NGI representatives and further contributors. The expected functions listed there will be later examined and evaluated on the basis of the use cases. This will allow us to complete the definition of EGI functions. The comparison with EGEE activities showed that there are use cases

[7] http://knowledge.eu-egi.org/index.php/EGI_Functions

missing specifically from areas dealing with management, dissemination and outreach and networking support. This corresponds with the fact that we did not obtain any use case related to administrative or legal issues while majority of use cases was related to infrastructure itself and the applications running of the infrastructure utilizing underlying middleware stacks.

## 5.5. END USERS AND GRID OPERATORS INTERACTION RULES

There are at least two different ways how to describe the possible interactions between various grid actors, especially two main categories of involved entities - End Users and Grid Resources Operators. A natural process of how to present such a complex environment as the Grid to the researchers would require a very minimalistic layer of basic services which the users can either directly use or build their own customized solutions on top of them. However the requested layer should be complete and robust enough to prevent users from bypassing key Grid services and establishing their own solutions that will never converge into consistent integrated grid infrastructure (the risk of federated grids). The user support is expected to work in very tight cooperation with grid operation personnel to introduce fabric infrastructure as a uniform entity. On the other hand, the virtual organizations serving as mediators between end user requirements and grid resource centers' operators should define their policies describing the expected users' behavior. To achieve an equilibrium in this issue is one of most critical matters whose solution is subject of detailed and exhaustive definition of EGI infrastructure functions. The collected use cases clearly demonstrate that the use case providers are interested in the common layer and the coherent and cooperative behavior of grid actors, not a loosely coupled federation of incompatible grids.

## 5.6. CONCLUSIONS AND FUTURE PLANS

There are basically two main approaches how to analyze the content of use cases. First, the so-called "bottom-up" approach tries to cluster use cases and sort them into categories based on their content and relationship between individual use cases. This allows to identify primary areas of interest in use cases. Second, a complementary "top down" approach can be applied, in which use cases are mapped to a set of predefined areas of interest. Two different predefined areas have been used for this analysis—the set of proposed EGI functions and the set of EGEE project activities. Direct comparison of results from both approaches (see Table 4) clearly indicates that both types of analysis led to the same results, however top-down approaches yielded also areas that are currently not covered by obtained use cases.

| Bottom-up | Top-down |
|---|---|
| Infrastructure access and operations | Operation of a reliable Grid infrastructure |
| Middleware development/deployment/enhancement | Middleware development and standardization |
| Application scenarios | Application support |
| | Development and operation of build and test systems |
| | Legal. organizational, and administrative issues, management |
| | Policies and strategy |
| | Industrial take up |

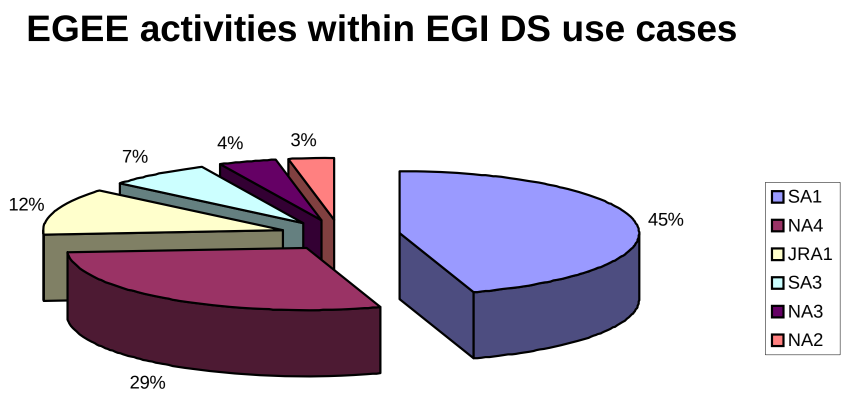
**Table 4 – Comparison of use cases analysis approaches**

To summarize: the obtained use cases results clearly indicate that Grid communities worldwide are willing to provide their day-to-day experience and/or requirements from the current Grid infrastructure utilization. During the collection of initial input from various Grid actors we collected and analyzed a substantial amount of provided material (see examples of use cases in Annex IV). The analyzed and processed content of all use cases were exposed in the EGI DS Knowledge Base and during December 2007 were made publicly available to the first community feedback. The fact that the use case collection is made publicly available has to do with the transparency in the EGI design process and with the openness in getting feedback on current use cases or acquiring new use cases[8].

| **Top expected EGI functions** |
|---|
| Operation of a reliable Grid infrastructure |
| Application Support |
| Coordination of middleware development and standardization |
| Development and operation of build and test systems |

**Table 5 – Proposed critical EGI functions**

To conclude, during the first phase of the EGI DS project we collected and analyzed material in the form of current Grid utilization use cases proposed by various Grid communities. This information resulted in a table of requirements (individual use cases) that has been mapped to the expected EGI functions. Mapping this to the proposed EGI functions indicates that some of the activities were more populated than others while some of them had been neglected. This can be easily seen when similar mapping towards various activities as recognized within EGEE project had been done (see Graph 1). This mapping shows that the following EGEE activities were recognized as the most important ones — SA1, NA4, JRA1, and SA3 (Grid operation, application support, middleware development and integration and testing respectively). This can be seen as a confirmation of the positive effect the EGEE project and the infrastructure built has on the understanding of Grids.




**Graph 1 – Statistics of use cases mapped to EGEE activities**

[8] Note that after the end of February 2008, new use cases have been received, e.g. from the SEE-GRID and RINGrid projects, that have been incorporated in the Knowledge Base and their impact will be evaluated in the WP3 processes.

Moreover, through the matching towards EGEE activities we identified several areas that have not been contributed in obtained use cases at all. The most prominent areas are the management (NA1 in terms of EGEE), legal matter/policy (NA5 in terms of EGEE), administrative matter and business models. A contribution to those topics will be subject of upcoming use cases "consolidation phase" during which missing areas will be covered as well as non-responsive recipients will be asked to provide comments to the current material. Simultaneously, the expected input from supercomputing and networking communities will be taken up directly to WP3 as a completely new source of contributions. From the long term point of view we would like to utilize the use cases collection as a so-called "validator" of the EGI model being prepared. Besides that purpose, obtained use cases are expected to have their own life—we expect extensions and modifications of the current use cases and incorporation of additional ones to be proposed by the project partners, collaborators and the Grid community at large. Overall the evolution of the use case collection is expected to broaden the knowledge of grid environment utilization and to summarize the feedback on the infrastructure, organization and support of sustainable grid environment.

# 6. THE EGI KNOWLEDGE BASE

## 6.1. INTRODUCTION

The purpose of the EGI Knowledge Base (EGI KB - http://knowledge.eu-egi.org) is to document the status and evolution of the European National Grid Initiatives (NGIs), to collect information about current e-Infrastructure activities and projects in Europe, to share the ongoing work done by the EGI Design Study team ( e.g. to share the use case collection – see section 5), and to provide a real time feedback loop between the team and the NGI representatives.

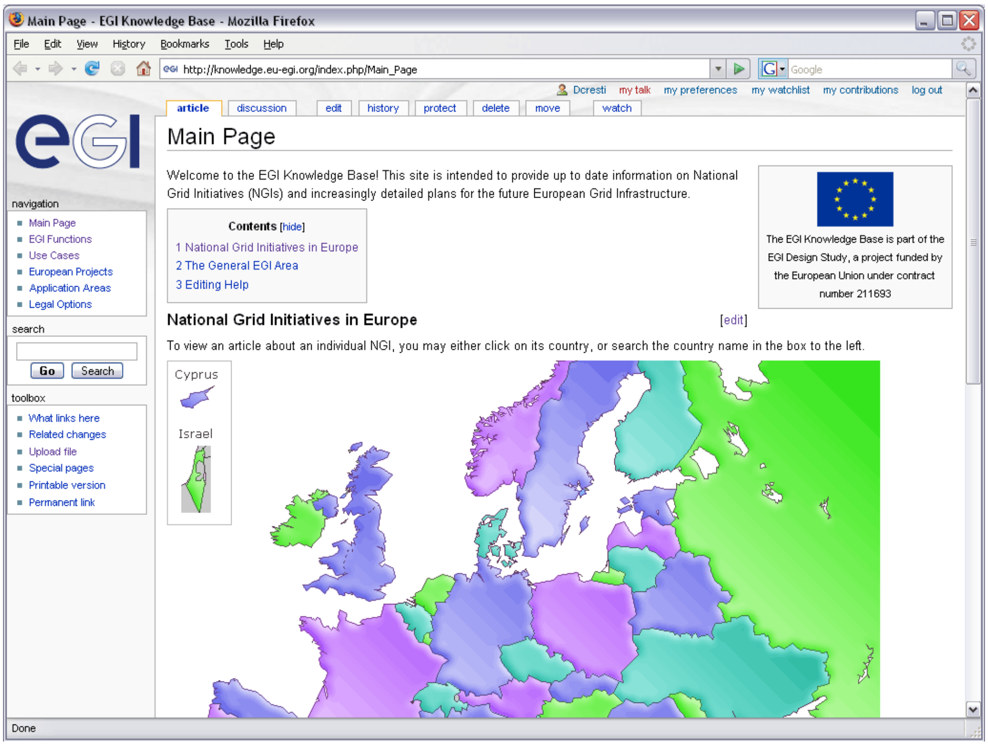



**Figure 8: The EGI KB Main Page**

The site is structured to contain (a) general information about the European Grid infrastructure, including relevant European projects and Application areas; (b) some results of the work done by the EGI_DS team – e.g. a collection and analysis of the Use Cases as discussed in Section **Error! Reference source not found.**, a description of the EGI Functions envisioned for the future EGI organisation, and the Legal Options for such organisation; and finally, (c) individual NGI pages. The individual NGIs each have a main page, reachable either by searching by country or by clicking on the image map in Figure 8. Each NGI can create further pages, such as the already existing Membership page for the Italian Grid Infrastructure (see http://knowledge.eu-egi.eu/index.php/Members_of_the_Italian_Grid_Infrastructure).

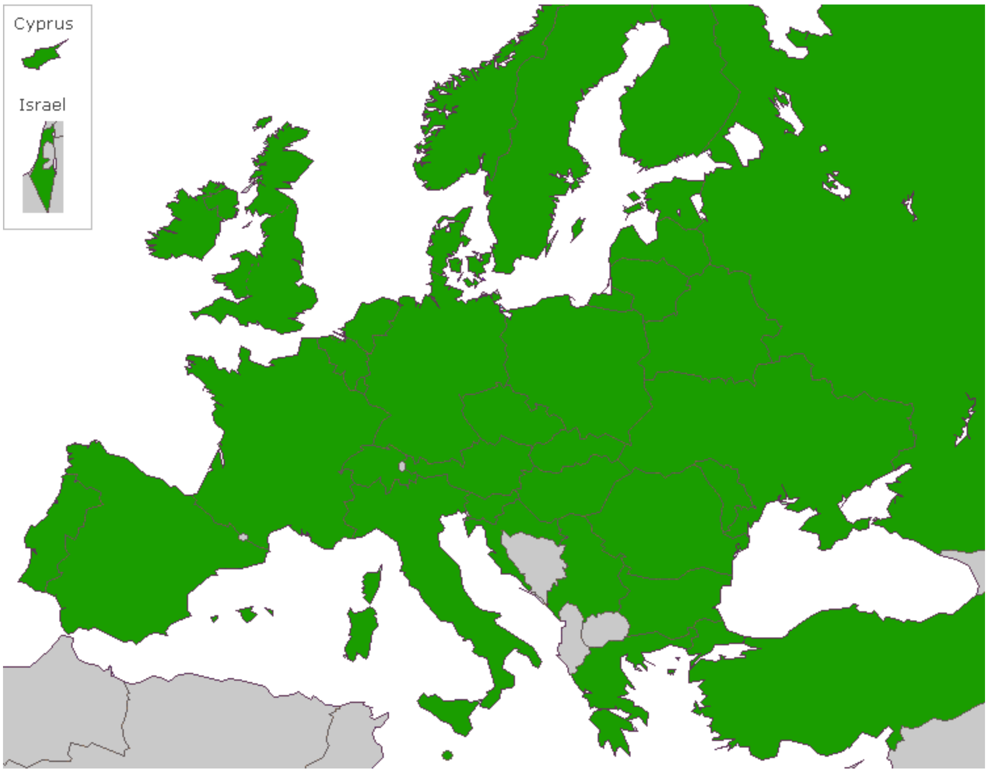
For a more complete report on the EGI KB, please see the WP2 deliverable **D2.2: The EGI Knowledbe Base: First Snapshot**.

## ANNEX I – LIST OF NGI MEMBERS (AS OF FEBRUARY 2008)

| No. | Country | Institution | Contact name |
|---|---|---|---|
| 1 | Austria | GUP, Joh. Kepler University<br>Federal Ministry of Science and Research | Jens Volkert<br>Stefan Hanslik |
| 2 | Belarus | Research Division of Belarusian National Technical University | Ihar A. Miklashevich |
| 3 | Belgium | BELNET | Rosette Vandenbroucke |
| 4 | Bulgaria | Institute for Parallel Processing,<br>Bulgarian Academy of Sciences | Kiril Boyanov |
| 5 | Croatia | SRCE, University computing centre, University of Zagreb | Ivan Maric |
| 6 | Cyprus | University of Cyprus,<br>Dept. of Computer Science | Marios Dikaiakos |
| 7 | Czech Republic | CESNET z.s.p.o. | Ludek Matyska |
| 8 | Denmark | DCSC - Danish Center for Scientific Computing | Rene Belso |
| 9 | Estonia | NICPB - National Institute for Chemical Physics and Biophysics | Martti Raidal |
| 10 | Finland | CSC - Scientific Computing Ltd. | Leif Laaksonen |
| 11 | France | CNRS - Centre National De La Recherche Scientifique | Guy Wormser |
| 12 | Germany | DFN-Verein - Deutsches Forschungsnetz (on behalf of D-Grid) | Klaus Ullmann |
| 13 | Greece | Greek Research & Technology Network – GRNET S.A. and HellasGrid | Panayiotis Tsanakas<br>Fotis Karagiannis |
| 14 | Hungary | NIIF - National Information Infrastructure Development Institute | Tamás Máray |
| 15 | Ireland | Grid-Ireland | Brian Coghlan |
| 16 | Israel | School of Physics, Tel Aviv University | David Horn |
| 17 | Italy | INFN - Istituto Nazionale di Fisica Nucleare | Mirco Mazzucato |

| | | | |
|---|---|---|---|
| 18 | Latvia | University of Latvia, Institute of Mathematics and Computer Science | Inara Opmane |
| 19 | Lithuania | Faculty of Mathematics & Informatics, Vilnius University | Algimantas Juozapavicius |
| 20 | Luxembourg | Fondation RESTENA - Réseau Téléinformatique de l'Education Nationale et de la Recherche | Antoine Barthel |
| 21 | Malta | University of Malta | Kevin Vella |
| 22 | Moldova | Research and Educational Networking Association of Moldova | Veaceslav Sidorenco |
| 23 | Montenegro | University of Montenegro | Bozo Krstajic |
| 24 | Norway | NorGrid, UNINETT Sigma AS | Jacko Koster |
| 25 | Poland | PL-Grid, ACK CYFRONET AGH | Michal Turala<br>Kazimierz Wiatr |
| 26 | Portugal | LIP - Laboratório de Instrumentacao e Fisica Experimental de Particulas | Gaspar Barreira |
| | | Knowledge Society Agency (UMIC) | Luis T. Magalhães |
| 27 | Republic of Serbia | AEGIS - Academic and Educational Grid Initiative of Serbia | Aleksandar Belic |
| 28 | Romania | Ministry of Education and Research, National Authority for Scientific Research | Ion-Sorin Zgura |
| 29 | Russia | RDIG - Russian Data Intensive Grid, Skobeltsyn Institute of Nuclear Physics | Viacheslav Ilyin |
| 30 | Slovakia | IISAS - Institute of Informatics, Slovak Academy of Sciences | Ladislav Hluchy |
| 31 | Slovenia | ARNES | Marko Bonac |
| 32 | Spain | Spanish e-Science network, Universitat Politècnica de València<br>IFCA - Instituto de Fisica de Cantabria | Vicente Hernández<br>Jesus Marco |
| 33 | Sweden | SNIC - Swedish National Infrastructure of Computing, Uppsala University | Sverker Holmgren |
| 34 | Switzerland | CSCS- Swiss National Supercomputing | Marie-Christine Sawley |

| | | | |
|---|---|---|---|
| | | Centre<br>SWITCH - The Swiss Education & Research Network | Fiorenzo Scaroni |
| 35 | The Netherlands | Netherlands National Computing Facilities Foundation | Patrick Aerts<br>Rossend Llurba |
| 36 | Turkey | TUBITAK-ULAKBIM | Cem Sarac |
| 37 | Ukraine | NAS - National Academy of Sciences | Gennady Zinovjev<br>Anatoly Shpak |
| 38 | United Kingdom | UK e-Science Institute<br>STFC - Science & Technology Facilities Council | Malcolm Atkinson<br>Neil Geddes |




**Countries supporting the EGI (February 2008)**

# 7. ANNEX II – USE CASES CONTRIBUTION INVITATION

**EGI - European Grid Initiative**

Towards a sustainable production grid infrastructure

www.eu-egi.org

23.08.2007

Dear Colleague,

Let me inform you, as one of key players in the area of Grid computing, supercomputing, and distributed computing, about the latest developments of the **European Grid Initiative (EGI)**.

The EGI Design Project (EGI_DS)[9] represents an effort to establish a sustainable Grid infrastructure in Europe, which requires a new model for the collaboration between the **National Grid Initiatives (NGIs)** and the future **EGI organisation**. Driven by the needs of the research community, EGI will enable the next leap in research infrastructures, thereby supporting collaborative scientific discoveries in the European Research Area (ERA).

As a first step towards this goal, we need to **identify the potential actors in this new Grid environment**, and **collect their requirements in the form of use cases of Grid deployment and usage**. This includes typical practices, daily experiences, common situations encountered when using Grids, as well as wishes and expectations of the role, functions and properties of the future Grid infrastructure.

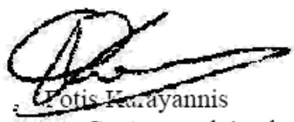
We ask you to send us one or more use cases that you believe are relevant to this goal. In the following pages, we provide a template and some guidelines which may help you in constructing your use cases, assuming you don't have any already prepared.

Please send us your input via email to usecase@eu-egi.org by **Friday, 21 September 2007**, so we can present a comprehensive set of use cases during the upcoming EGI Workshop on Tuesday, 02 October 2007.

If you have any other questions or remarks on the above, please do not hesitate to contact us at the same address, or for any other generic question at contact@eu-egi.org.

Yours sincerely,

Dieter Kranzlmüller
EGI_DS Coordinator
(on behalf of EGI_DS)



Fotis Karayannis
Requirements Capture and Analysis
Work Package Leader

[9] The EGI_DS project is expected to start on the 1st of September. The preparatory work before the official start of project is co-funded by the EGEE project

## ANNEX III – USE CASES TEMPLATE AND GUIDELINES

### I.Template

**Use Case title**: The title of the use case

*Example: Create Virtual Organisation (VO)*

**Short description**: A short description of the Use Case

*Example: A multi-national well organised application community wants to create a VO that will be able operate on the new EGI-NGIs environment.*

**Actors involved**: The envisaged actors involved.

VO coordinator, VO operator, NGI operator, (possible future scenario: EGI operator?)

**Related Requirement**: The related requirement posed by the corresponding actor.

A new community needs to quickly and easily join the Grid

**Pre-conditions (optional)**: Conditions that need to be in place before the use case takes place

The community needs to be already organised as a multi-national team

The community in some cases (depending on national policies) needs to bring resources in the natio infrastructure

**Steps**: The actual steps required, so that the use case can be materialised.

1.
2.
3.

..

**Post-conditions (optional)**: Conditions after the use case takes place

The VO is created and supported by X NGIs and the EGI central pool of resources.

**Project(s) involved (optional)**: The projects involved in the use case (if applicable)

**Middleware (optional)**: The middleware packages involved in the use case (if applicable)

**Application(s) (optional)**: The applications involved in the use case (if applicable)

## II. Guidelines

These guidelines are given to help you focus on the kind of questions that are relevant to a future NGI-EGI scenario, and to provide a context for the construction of use cases.

### Use Case Areas

In relation to the main actors involved in a Grid infrastructure, we identify three corresponding use case areas. Under each, we give some questions that can guide the construction of a use case. Some questions might be applicable to more than one type of use case.

I. **Infrastructure**-oriented use cases:

In what manner does an infrastructure provider/operator currently share its computational resources or data storage capacities? What are its current responsibilities and what kind of control does it have over its resources? What issues have arisen in this particular use case context, and how can the NGI-EGI environment address these issues?
If you are a resource provider, tell us about your view on how best to collaborate within the international community. How do you imagine the incorporation of grid research (grid middleware development) into the forming sustainable infrastructure? Would you expect grid research to be completely independent, interconnected within projects or rather subsumed (to some extent) under the responsibilities of EGI?

II. **Application-**oriented use cases:

What is the typical process a new application community goes through when interested in doing research on the Grid? How are VOs typically created or identified now, and how should they be in the new NGI-EGI environment? How to adapt a generic middleware service towards specific research group requirements? Is there an effort to develop tailored middleware services for specific research users groups? Are my applications already available on the grid?

III. **End user**-oriented use cases:

What are the required prerequisites to become a user? How to join the global infrastructure available? Which is the official registration procedure? Should I connect through the corresponding national Grid activity or go through a global registration pointer? Is the Grid certificate for authentication included in the registration procedure?
What are the main obstacles for you to use Grids comfortably on a daily basis - latency during job submission? Non-intuitive work with certificates? Non functional middleware services? Any other hassle?
Ultimately we are interested in your opinions on how you imagine the future of grid environments, such as:

- What are the main concepts required for seamless cooperation between already established worldwide grid infrastructures and local national grid activities?
- For which cornerstone services should the emerging EGI be responsible?
- What kind of interactions among different national/global grid initiatives should be drafted and foreseen?

For all use cases, a useful question to try to answer is “what kind of problems have you been forced to solve, did you solve it and how, and how could it be solved within the NGI-EGI environment?”

# ANNEX IV - SELECTED REPRESENTATIVE EXAMPLES OF USE CASES

Some representative examples of use cases are presented in this section taking into account the three main categories identified in section 5.5 (table 4), i.e. middleware, infrastructure &operations and applications. The use cases for each of the areas are obviously subjective, but try to cover a broad range of scenarios and sub-functions.

1. Middleware

    1.1. Smooth integration of new components

## Use Case title: Leading Middleware Development

Short description: Currently Grid computing is facing a “Grid interoperability contradiction” in general and on the European arena in particular. While Grid addresses interoperability over heterogeneous resources, there are major interoperability and portability problems between different Grids. As a result, Grid developers are constantly recreating the basics, leading to slow progress in the implementation of fundamental Grid services, which in turn makes the development of portable high-level Grid applications immensely difficult. Hence, the main motivation for seriously addressing Grid interoperability problems is actually not only to facilitate exchange of data between different Grids, or to facilitate cross-grid job management, although these things are also relevant and important. If not addressing these problems appropriately, we are wasting money, the time of software developers, and we are running out of patience of the end-users.

We expect multiple middleware to co-exist for a foreseeable future, and we cannot afford to develop every tool for each specific middleware or to make the high-level applications specific to a particular middleware. Instead, we need to strive for a development strategy that more easily allows the good tools for particular purposes to be used with different middleware. EGI has an important role in making this happen!

As of today, the best way we (the Grid community) knows on how to realize this follows from the well-established and basic idea of an eco-system of Grid components. Fundamental to this approach is a separation of concerns - a Grid component should ideally be designed to perform one specific and well-defined task, and its dependencies on other components should be well-defined and kept as small as possible. Following this strategy, one component (may it be a job submission tool, a resource broker, a logging system, a workflow system, a name service, a file replica manager, etc, etc) can more easily be adjusted for use in middleware other than its original target environment. Here, the analogue with the eco-system is that with replaceable components, a natural selection process of the best components for each specific task can take place based on competition, innovation, evolution, and diversity. On the contrary, with heavy components were too much functionality is built into the same tool (or a set of tools with too strong interdependencies) the reuse of individual components is virtually impossible.

Of course, standardization plays an important role here, but interface standardization alone cannot solve this problem if the general development strategies involve too monolithic systems and multifunction components.

We argue that one of the most important tasks for EGI could be to lead the European Grid development in the direction of a sound eco-system of Grid components, for the future we foresee with multiple co-existing middleware.

Actors involved: EGI, middleware development projects

1.2. Component lifecycle

### Use Case title: Grid services development, tuning and deployment

Short description: In production deployment, Grid services are exposed to load scale and non-deterministic behaviour of the infrastructure (delays, failures, and outages) virtually irreproducible in any testing environment. Therefore a standard software release process (prepare release candidate, test independently, fix problems found in testing, do the release, gather and react to users' bug reports) is rather inefficient. The most critical problems, e.g. race and congestion conditions, are not hit in testing because it is unaffordable to run independent tests in a scale comparable to the target production. On the other hand, due to the infrastructure complexity and non-determinism, it is extremely difficult for the users to provide problem reports that would be reproducible, or at least precise enough to spot the true problem.

A radical change in the approach to the software release is required to address the different conditions. Representatives of service developers, testers, infrastructure operators, and end-users must form task forces focused on a specific software release. One or few instances of the release candidate services are injected into the production environment, and the end-user representatives generate their load in a way that is controlled and reproducible, while still resembling real production, in both usage pattern and scale. The experimental services are continuously monitored by the infrastructure operators in a close cooperation with the developers. Then any behavioral anomalies and malfunctions are detected almost immediately, they can be repeated for further analysis with high probability and less effort than in real production, and they are fixed on the spot finally. Once the service operation is settled, its current codebase and configuration snapshot from the desired release which is spread in a conventional way then.

This use case is supported by the experience with ``Experimental services operation of new gLite Workload Management System on EGEE infrastructure in 2007. During this several months exercise the WMS services were successfully tuned to considerably outperform ``acceptance criteria specified by HEP applications (10k jobs/day sustained submission over 5 days, without manual intervention, and with less than 0.5% failure rate). The resulting code and configuration forms the first gLite 3.1 WMS release (aka patch #1251). The following CHEP'07 presentations provide further details (to appear in J.Phys.: Conference series):

* S. Campana et al.: Experience with the gLite Workload Management System in ATLAS Monte Carlo Production on LCG
* M. Checchi et al.: The gLite Workload Management System

Actors involved: local research community, VO operator, NGI operator

### 1.3. One or more MW stacks

Use Case title: Build European Grid infrastructure based on Grid middlewares from different providers

Short description: A user community wishes to share their existing resources via European Grid infrastructure. By the nature of the community specifics, these resources are already equipped with Grid middleware that suits this community most, and they do not intend to re-configure the resources, primarily because there is no manpower and the learning curve for any new middleware is very steep and takes time, all of which incurs unbearable costs, while the community hopes to reduce such, by joining EGI. The community then approaches EGI and requests to be accepted, much like a new domain joins Internet, or a new mobile operator is registered. EGI verifies that the community resources are compliant with international (e.g. OGF, IEEE) standards and satisfies EGI policies, and proceeds with including this set of resources into the infrastructure, no matter what middleware flavor or hardware is involved – just like it does not matter what firmware a mobile phone has, as long as its external interfaces are standards-compliant. As a result, the infrastructure consists of standard segments delivered by different manufacturers/providers, exactly like any other European infrastructure. In order to address needs of different customers, and to allow for healthy competition leading to technology progress, European Grid infrastructure must develop mechanisms that guarantee the possibility of including any Grid resource, independently of its internal architecture and set-up. A resource that provides a set of standard Grid interfaces and a number of functionalities that are needed by a user community.

Actors involved: User community (VO), possibly national Grid infrastructures, EGI operators, EGI security officers, possibly international standardization body, possibly funding bodies, possibly middleware providers.

Related Requirements:

* The procedure must be quick and drain no significant additional resources from the community
* EGI must not require installation of any specific software at community resource
* in case the community middleware is found to be non-compliant, community must have a freedom of selecting another middleware (a list with more than one suggestion from EGI is welcomed).

Pre-conditions:

* The community must have a working Grid structure already, not yet affiliated with EGI

* EGI must embrace international standards and develop a set of acceptance criteria based on them (not on de-facto standards of individual Grid players)

* EGI must have the corresponding administrative structure in place, that accepts and processes applications in a speedy manner

Steps:

1. User community (eventually organized as a VO) contacts EGI via an authorized representative (e.g. VO manager, if such exists, or a national Grid project coordinator etc.)
2. EGI officers provide a check-list of criteria that must be satisfied (including required OGF profiles, interfaces, policies)
3. User community provides evidence that the criteria are met
4. EGI and the User community in question exchange the necessary information to enable resource inclusion (service contact points, authorization details, signed agreements etc)

Post-conditions: The new resource enjoys operational benefits of being a part of EGI even when running a middleware of their own choice; EGI is obviously not required to offer any middleware support – this lies with middleware providers.

Projects involved: EGI, OGF, IGTF, national and international Grid development and deployment projects, any middleware provider

Middleware: Any middleware

Applications: All kinds of applications

### 1.4. Support of interactive use

#### Use Case title: Visualization of Plasma in Fusion Reactors

Short description: This application visualizes the behaviour of plasma inside a Fusion Reactor. Runs are foreseen as a part of a so called Fusion Virtual Session. The plasma is analyzed as a multi particle system consisting of N particles which are distributed among a number of processors, which calculate the individual trajectories. The inputs of the simulation are

Geometry of the vacuum chamber of the fusion device Magnetic field in the environment Initial number, position, direction, velocity of particles Possibility of collisions between particles Density of particles inside the device

The simulation takes place by solving a set of Stochastic Differential Equations with Runge-Kutta method. The dynamics includes already the interaction between the particles which is introduced as an effective field. Therefore the simulation can take place independently for every particle trajectory.

The outputs are the trajectories of the particles, which are used for visualization purposes, and the average of relevant magnitudes: densities, derives, temperature, etc…

A graphical interface with the application visualization pops up in the desktop of the user as soon as the application starts. The user can then choose the simulation parameters and start the simulation by clicking on the display with the mouse. The evolution of the plasma is then forwarded to this display in real time. The user can zoom in and out and view the system from different angles. The plug-in to the Migrating Desktop developed by the integration team of the project includes interactive and steering capabilities include control over the number of particles in the simulation, electric field, and several parameters affecting the Runge-Kutta algorithm for calculating the particle trajectories.

Actors involved:

1.5. Security issues

Use Case title: Medical Image Analysis on the Grid.

Short description: Sharing and organising medical imaging knowledge is a key issue in medical research and training. Evidence-based medicine is also demanding high-quality well-organised knowledge bases to check for second opinion and drive diagnosis. However, sharing and organising medical imaging data is not straightforward. Technological and legal problems on exchanging data make it difficult or even impossible with the current infrastructures. On the other side, the index criteria used in clinical practice are inefficient when searching for knowledge.

Actors involved:

* Data Providers: Medical centres (mainly hospitals) with Image Diagnosis Departments.

* Users: Radiologists.

* Application Developers: Integrators of the medical imaging processing tools, data interfaces and side-applications.

* Operators: In charge of the administration of the users, the update of templates and the maintenance of data gateways, central services and processing resources.

Related Requirements: The requirements are analyzed in four areas:

* Computing. Depends very much on the number of users and cases involved. A figure of a sustained availability of 30-40 CPUs with peak demands of 50-75 will be reasonable for a case with 5-7 users.

* Storage. The storage requirements are low, since data is kept distributed at the local centres (in order to be compliant with the EU regulations). Data is temporarily copied in the main storage and local disks of computing services. Central services only store catalogues and user profiles.

* Security. User authentication and authorisation is managed through VOMS credentials. Services are “VOMS-aware” and only provide access if the appropriate credentials are presented. Data is encrypted to avoid malicious access from unauthorised users (even local users with administrative privileges at the processing services). Encryption keys.

* Deployment: The services must deal with the particularities of hospital environments. Some nodes must be deployed within the hospital network, which implies that firewalls restrict heavily the ports and protocols that can be used. Adaptation of GridFTP, for example, needs to be done.

Pre-Conditions: A hospital infrastructure must be set-up to deal with it.

Steps:

1. Producer
    1. Authenticate and validate group credentials.
    2. Select case and copy and register it in the global indexing system.
    3. Fill-in the metadata related to the diagnostic of the case.
    4. Submit the metadata.
2. Consumer
    1. Authenticate and validate group credentials.
    2. Select the ontology and fill-in the searching criteria.
    3. Select the post-processing action (if any).
    4. Submit the job (data retrieval to the processing services, processing and copy of the results).
    5. Retrieval of the results.

Post-conditions: A group of operators update the structured report templates and ontologies, used in the indexing of the cases, as well as the groups and roles of the users.

Projects involved: CVIMO, “Valencian Cyberinfrastructure for Medical Imaging in Oncology” (http://www.grycap.upv.es/cvimo). CVIMO is a platform developed to share and organise medical studies and reports based on ontologies constructed upon the fields of structured reports. It is based on a Grid Software Architecture of WSRF services that organise coding, access rights and data location for different studies and reports.

Middleware: The middleware used is a proprietary system called "TRENCADIS" (Towards a Grid Environment for Processing and Sharing DICOM Objects), which is based on WSRF and uses components from GT4 and gLite.

Application: Application for searching cases by content, writing structured radiology reports and processing images.

### 1.6. Large redesign is required

Use Case title: Engage in future grid middleware redesign for European grid computing and beyond

Short description: Grid development faces serious de-fragmentation as well as the burden of early success: today's grids grow outside the scope of early grid visions. EGI should start an endeavour to start rethinking grids from scratch, taking into account all we know of today's and planned grid application fields and the experiences with all different flavours of grid middleware. This activity should be managed according the rules of the book (standards, version control, open source, planned targets, etc.)

Actors: the grid scientific community in the EU and US(Pete Beckman), funding agencies, NGI officials.

### 1.7. Other specific functionality requests

#### Use Case title: Guide novice users to the Grid

Short description: Today's procedure, from the request for a personal digital certificate up to data retrieval from a site is too complex. The chain Authentication – Authorisation – Job submission – access to results needs to be simplified. Integration of new communities to the Grid has to become faster and easier. Already within existing VOs, a lot of effort is spent to train new users, despite the experience available in their environment. Today, Grid usage is reserved to computing 'experts'. This is not a good forecast for funding.

Actors involved: Grid architects, Middleware designers and developers, User supporters.

Related Requirement: Users who don't belong to large communities but need to use the Grid cannot today unless they register a new VO. This situation doesn't scale and certainly will never make the Grid as popular as the Web.

Pre-conditions: (optional) Important middleware re-design effort needs to be invested. Security policies must be re-thought in a, still secure, yet more flexible way.

Steps: The actual steps required, so that the use case can be materialised.

* The Joint Security Policy Group to revise policies in the light of a large Grid community.

* Simplify or abandon certificate-based authentication for users and hosts.

* The middleware security group, which includes policy makers and developers, to review the Authentication and Authorisation procedures and tools.

* The VO Registration Procedures, VO Management and User responsibilities to be re-defined.

* Grid Storage and Data Management experts to re-design their code in the new 'plug-and-play' spirit.

Post-conditions (optional): Audit, monitoring and security update procedures should be well established and scrupulously applied for the Grid to be more easily accessible but remain secure.

Project(s) involved (optional): All special-purpose projects under the gLite software umbrella. Security experts, CAs and commercial partners from the computer industry, databases and, even, mobile telephony projects can be foreseen for the era, when Grid usage becomes really easy.

Middleware (optional): Integration to the Grid environment or replacement of certificate signing and renewal for hosts and humans. All middleware packages, especially VOMS/VOMRS, UIs, RBs, SEs.

Application(s) (optional): All information dissemination tools; documentation, software repositories, web pages, newsletters, registration forms, multiple purpose portals.

2. Infrastructure and Operation

2.1. Joining the Grid

### Use Case title: Adding a new NGI/EGI site

Short description: This use case presents a situation in which a new site is applying for participation in grid infrastructure. The site has some resources to be shared in grid environment. This procedure covers steps that need to be taken at NGI and EGI level. We foresee that NGI and EGI will have their own rules on which a new site is allowed to join the grid infrastructure. These rules may form a Service Level Agreement (SLA) which is signed as the site joins the infrastructure. We assume that all EGI resources are bound with NGI resources, but some NGI resources may not participate in EGI infrastructure.

Actors involved: RC operator - person responsible for the site, RC manager, NGI operator, EGI operator

Pre-conditions (optional): Hardware at the site is ready. NGI and EGI rules (SLAs) for sites are ready.

Steps:

1. The RC manager asks the NGI operator for guidelines and requirements (SLA) to join the NGI infrastructure as well as EGI infrastructure
2. The RC operator sets up the necessary services
3. The RC manager signs requirements (SLA).
4. The NGI operator checks if the site complies with the EGI requirements (SLA) and if so, accepts (certifies) the site for the EGI infrastructure
5. (optional) The EGI operator confirms the site in the EGI infrastructure

Post-conditions (optional): A new site with its resources contributes to the NGI infrastructure and to the EGI infrastructure.

## 2.2. Working with the Grid

### Use Case title: Collaboration within international research communities

Short description: A local research community will be granted a new international research project. This community has several international research partners and all together would like to interconnect their individual computational/data resources (some of them may already be on the EGI grid, most are not). What will be the role of EGI in this process? What help/support/advices can the research community joined in a project expect from EGI/NGI? How the situation will change is some of the collaborators are from outside of Europe (e.g. USA, Japan, …)?

Actors involved: widely distributed research community, VO operator, NGI operator

## 2.3. Organization of operations

### Use Case title: Expand Virtual Organization outside of country

Short description: NGI has created a VO for people within its country's borders; these people would now like to make use of European resources and collaborate with similar researchers in other countries. A recent example is the Auger VO (AUGER experiment); we were contacted by colleagues in Prague who already had a VO set up, they knew we were involved in Auger and wanted us to support their VO.

Actors involved: NGI coordinators, VO representatives (and by VO here I mean the computing people), experiment officials (these people are the bosses of the VO reps).

## 2.4. Policies

Use Case title: Interconnect the existing European and Chinese Grid Infrastructures, EGEE and CNGrid, through interoperability.

Short description:

Actors involved:

Pre-conditions: WP2 carry out studies on the available and foreseen network connectivity promoting new high-bandwidth links between Europe and China, WP3 carry out studies on CNGrid and EGEE interoperability proposing deployment solutions, WP5 create synergies and encouraging the creation of new Grid Initiatives.

Steps:

* Analysis;
* Design and development of a gateway service allowing the interoperability between the two middlewares;
* Test the gateway service;
* Deploy the solution;

Post-conditions: Different instances of a gateway service, connecting the two the infrastructures, has been deployed. The research and education communities of China and Europe have a transparent access to a larger world widely distributed amount of storage and computing resources.

Middleware: Installation of gateway nodes (interoperability between gLite and GOS middlewares) allows the interconnection between the two infrastructures, CNGRID and EGEE.

Applications: A set of existing Euro-Chinese collaborations in research, marked by strong requirements in terms of analysis of large quantities of data and needs for wide amounts of computing power, can exploit the resource of the two infrastructures.

## 2.5. Commercial related issues

**Use Case title: Privately funded hardware**

Short description: A group of users needs to join the Grid with privately funded and dedicated CPU and storage

Actors involved: RC manager, RC operator, VO administrators, local technical staff groups

Related requirement: CPU/storage dedicated exclusively to the project

Pre-conditions: Technical feasibility (electric power, space, cooling etc.), enough manpower for administration and service provision. Some technical and manpower redundancy required.

Steps:

1. Evaluation of technical conditions
2. Evaluation of service availability
3. Hardware installation
4. Configuration of queuing systems
5. VO setup
6. Service configuration

## 2.6. Training & User Support

### Use Case title: Application group "grid guru" training and support

Short description: A group of scientists uses local computational resources regularly, and they want to join grid to leverage its potential. However, a realistic learning curve is still too steep, most of these people cannot afford spending their time learning how to use grid tools. On the other hand, there are few people in the group, more experienced in computer science, who are still familiar with the needs of these users, while being capable, willing, and ideally paid for becoming "grid gurus" of the group.

However, even for these people it is rather difficult to keep pace with development of the grid technology, and to be aware of all available functionality of grid services and map it to the needs of their group. Therefore they easily fall into the trap of reinventing the wheel, providing their custom solution to a problem which has been already addressed by some advanced grid services.

These "gurus" form a special category from the point of support and training. As they hold the potential to bring their communities to the grid, and they also affect how the grid is perceived by the true end-users, the national and international grid infrastructure should establish and maintain collaboration with these people actively, offer them specific training opportunities etc.

Actors involved: local research community, VO operator, NGI operator

2.7. Monitoring and Accounting

**Use Case title: Resource allocation**

Short description: A ROC manager wants to know how the resources of a VO are distributed over the regions and which services are assumed where. This might help in capacity forecasts or for failover planning of critical services.

Actors involved: ROC managers.

3. Applications

3.1. Commercial software licenses

Use Case title: **Licensing of commercial software to be used within Grid environment**

Short description: There is a research community with its own computational resources willing to offer those resources to the global grid infrastructure. The community is daily dealing with a set of commercial applications (coming from distinct application areas as molecular modelling, material simulations etc.). What will be the EGI model for a situation when this community would like to join EGI infrastructure but have to fulfill specific restrictions related to individual licenses of commercial software packages?

Actors involved: small to medium research group of end-users, local NGI representatives, VO coordinator

3.2. Parallel job run

**Use Case title: Parallel software application execution on specific resources**

Short description: This use case shows a target model for submitting and executing parallel software applications on the Grid. The current model (EGEE) allows for such jobs, however their execution is very complicated. It is not possible to specify how to make given resources accessible to the job (for example, it is not possible to run a multithreading job).

Actors involved: user, application, RC manager, RC operator, NGI manager

Related requirement:

1. A job language description, allowing users to specify desired resources (for example, the requirement that the job has to be run on a multicore node or SMP machine, or request for a given number of cores/processors on several nodes)
2. Middleware capable of reserving resources requested by user jobs
3. Site(s) with specific hardware resources, allowing parallel execution (SMP machine, large cluster, etc.)
4. Software library (libraries) allowing parallel execution

Steps: Run a parallel job with as much ease as present sequential jobs.

Application(s) (optional): chemical or biological software (ab initio and molecular dynamics code)

### 3.3. Interactive jobs

#### Use Case title: UltraSound Computing Tomography

Short description: Ultrasound Computer Tomography is a new method of medical imaging based on the reconstruction by numerical techniques of an image, using as input the data measured by a scanner of Ultra Sounds which surrounds the object of interest.

In the basic setup every point (transducer) of a cylindrical scanner emits a pulse with a circular wave front, while all others receive the scattered signal. When the process is repeated for all transducers, it is possible to reconstruct the object that has produced the scattering patterns recorded by all transducers along the process.

The researchers from the Forschungszentrum Karlsruhe are investigating the application of this technique to early detection of breast cancer. This means detection of tumours as small as possible, and with the most precise spatial location possible. For this we need sub millimetre resolution in three dimensions. Conventional reconstructions techniques are slow and in practice suppress its usage on a large scale in Hospitals.

Actors involved:

Requirements

The computational cost of image reconstruction is very high. In a typical situation we face an amount of data of about 20GB. The reconstruction of the full data set would take in a single workstation about 180 years. This would imply sub millimetre resolution, for the whole volume.

However such resolution is rarely needed, and in Hospital cases one restricts de volume of study and also the areas of high resolution to be more limited. In a normal use case we would be talking about reconstruction times of about 30 days in a single workstation. Evidently this is too much time, and our goal is to reduce this time to something acceptable profiting from Grid technologies.

The bottleneck in this application is data transmission. If we split the calculation in a number N of jobs, we have to submit the 20GB of data along with every job. This strategy has the advantage that the calculation of every job can be restricted to a subset of the image, and the output of the individual jobs is smaller. It has however the drawback of having to send at submission time N times 20GB of data.

Data are submitted as a tar ball containing 3.5 Million input files, which is locally unpacked into many files and many directories. This will put stress in the filesystems on the local clusters. Accessing a lot of small files will probably put stress on the local network because NFS or even parallel distributed filesystems are very sensitive to accessing many small files.

Besides filesystem issues, we should consider to use always SMP machines for this application with at least 2 physical processors per motherboard. The situation is that the linux kernel, by construction, uses one of the CPU for data access and the other for floating point calculation.

Besides computing requirements, there are other user requirements. One is trying to hide as much as possible the complexity of using a Grid infrastructure, in terms of Grid certificates, submission of large amount of jobs, etc… Efforts have to be put into a reliable strategy to upgrade software versions on the repositories, and of course, into fault tolerance issues.

### 3.4. User support for individual application group

Use Case title: A new community needs to port its application on the grid.

Short description:

Actors involved:

Pre-conditions: A stable and efficient e-infrastructure is available. Proper actions are planned to identify new communities and new applications to be ported on the grid infrastructure.

Steps:

* Identify and attract a new community and its applications through dissemination events (tutorials, grid schools, etc.);

* Evaluate the application and its infrastructural requirements;

* Grid Design;

* Functionality and needs audit: a set of proper tests is performed in a testing environment. If there is no failure in tests, the grid experts can certify the application as a “supported application” to be deployed in the production grid;

* Show-case in Educational and Business Briefing: the application ported on the grid is presented as cases study/example during training events and workshop where EUChinaGrid project is involved.

Post-conditions: The new application is running on the grid infrastructure. The new community can benefit of the huge amount of storage and computing resources available on the infrastructure.

## ANNEX V-USE CASES TERMINOLOGY

## Actors

The actors below were considered as self-evident or the definition has been given in the corresponding use cases text, thus no explanation has been provided.

- Local NGI representatives
- Small to medium research groups
- VO coordinator
- NGI operator
- Local community
- Global community
- Service developers
- Testers
- Infrastructure operators
- End-users
- Grid gurus
- Central operations management
- Security officer
- Operational Security Coordination team
- ROC managers
- Operations tools developers
- Site manager/administrator
- NREN
- Security officers
- Virtual organization
- National data centers
- European data centers
- National libraries
- Certification Authority
- Bioinformatics scientists
- Deployers (users)
- Grid Trainers
- OGF
- IGTF

**Objects**

- *Middleware* The software that is used on top of the operating system for providing the necessary services for the operation of the Grid infrastructure and the interaction with the users.
- *MPI* Message Passing Interface (MPI) is a language-independent protocol to allow the communication between processes distributed over different processors.
- *Parallel computation* A form of computation in which instructions are carried out in parallel.

- *Commercial software* Software that is developed for commercial purposes. In the context of Grid infrastructures the term usually refers to proprietary software that requires licenses to run.
- *Software licensing* The provision of licenses to allow lawful execution of software on some equipment.
- *Computational resources* The term is used to distinguish resources that perform computation only, in contract to other computer-related resources such as bandwidth, storage, etc.
- *Data resources* The term is used to distinguish resources that store data only, in contrast to other computer-related resources such as computation, bandwidth, etc.
- *Infrastructure/grid gateway/portal* An application that allows the user to interact in an integrated way with the Grid resources. Such applications typically allow users to submit their jobs, to see the status of submitted jobs, to retrieve their results, etc.
- *Pilot job* A pilot job is an empty shell, which is submitted onto the grid through standard submission channel. Once it gets executed, it calls back its controlling service, which is run under a complete control of the user or the users group, and the service provides the pilot with its actual payload.
- *Dynamic scheduling* A scheduling mechanism for jobs submitted to the Grid that takes account of changing workloads to adjust job allocation to resources on the fly.
- *Workflow execution* The execution of a set of jobs that form part of an integrated workflow.
- *Bulk submission* The submission of many related jobs in a single batch by the user.
- *IdP* An Identity Provider (IdP) is an entity responsible for supplying information about users at a domain to relying parties.
- *User certificate* An electronic certificate that incorporates a digital signature from a Certificate Authority that binds together a user's public key with user credential information.
- *Experimental service* A Grid service that is not ready for production usage yet and is installed in certain nodes in the Grid infrastructure only.
- *Security incident* A breach of security in a computer system.
- *Monitoring* The task of watching a system's operation and collecting operational data in order to assess its performance, help in resolving problems and security incidents, etc. In a Grid infrastructure monitoring may be hierarchical, i.e., performed at different levels comprising the infrastructure.
- *Trouble tickets* Problem reports submitted by users of the infrastructure and entered to a suitable issue resolution system.
- *Resource allocation* The allocation of computation, network, data storage, or other kind of resources, to research groups, users, and jobs submitted to the infrastructures.
- *Resource usage* The use of the infrastructure resources. The term is used both in the context of the actual observed usage, and the desired resource usage; the two may be different due to over- or under-utilisation of the infrastructure resources.
- *Site configuration* The arrangement of an infrastructure site's functional components with respect to its participation in the Grid infrastructure.
- *Data challenges* A large scale experiment using Grid resources intensively for a specified period.
- *Resource contention* The situation where the resources required by submitted jobs extends the resources available at the infrastructure.
- *VOMS server* (Virtual Organisation Membership Service server) A server that authorises Grid users based on their certificate that has been authenticated by a Certificate Authority.
- *File catalog* A hierarchical directory structure, usually like a Unix file system, that provides access to files residing on the Grid infrastructure.

- *Fragmentation of grid infrastructure* The phenomenon of evolving towards different, incompatible Grid infrastructures.
- *Data archive* Large-scale archive of scientific data that are available for processing on the infrastructure.
- *Sustainability* The creation of a permanent Grid infrastructure that will not depend on short-term funding models, but will be able to maintain itself financially and materially in the long term.
- *Scientific data curation* The task of archiving and cataloguing scientific data (such as those used in experiments, research work, etc.) in order to make it available to the research community at large beyond the time and resource constraints of individual projects.
- *Digital libraries* Large scale repositories of digital data.
- *Visualisation tools* Tools that help researchers and scientists visualise, i.e., get a visual perspective on, their data.
- *Throughput* The amount of work that a computer system can perform in a given time period.
- *Numerical computation* The use of computers to solve problems involving computation of real numbers with increased precision requirements. Numerical computations are required in the application of computing to scientific and engineering disciplines.
- *Real time tools* Software that gets input and produces input in real time; this is often used in contrast to standard Grid software that runs at scheduled times and produces output in batches.
- *Shared workspace (virtual work arena)* A workspace that allows shared access to objects (data, programs, communication facilities) to geographically distributed users.
- *Bandwidth QoS* The provision of guaranteed bandwidth to selected services or applications for specific amounts of time.
- *Testbed* A collection of resources (computing, data, etc.) that is used to test specific functionality of services.
- *SLA* Service Level Agreement (SLA) An agreement between parties in an infrastructure whereby one party guarantees the provision of specific service quality (based on appropriate metrics) to the others.
- *License server* A server that dispenses licenses to be used with commercial software. Commercial software copies usually acquire licenses from a license server.
- *Resource Allocation Portal* A web site that allows the allocation of infrastructure resources in an integrated way. A resource allocation portal allows the resources dispersed in different sites to be viewed together and allocated from a single point.
- Web Service According to the W3C, a web service is a software system designed to support interoperable Machine to Machine interaction over a network.
- *WSDL* Web Services Description Language (WSDL) is an XML language that describes the interface of a web service to the outside world.
- *OWL* Web Ontology Language (OWL) is an XML language that describes web services ontologies, including descriptions of classes and their properties and instances.
- *OWL-S* Semantic OWL is an OWL language that describes web services semantics.
- *Gridification* The process of transferring an application to the Grid. This may require extensive restructuring and refactoring in order to take advantage of Grid resources (for instance, it may require breaking the application in parts to be run in parallel, dividing the application data among the nodes in the Grid infrastructure, etc.).
- *Job management* The set of tasks related to handling jobs in the Grid, including jobs submission, scheduling, job monitoring, job termination, provision of input, retrieval of output, etc.
- *Authentication* In computing, the confirmation of an entity's identity.

- *Authorisation* In computing, the permission of the use of a certain resource by an entity based on its identity.
- *Audit* In computing, a systematic assessment of a system or an application. In computer security, an assessment of the events that led to a security breach, usually based on collected log data.
- *Resource markets* An approach whereby computing resources are not allocated by decree, but through a market where users place their bids for the resources they need.
- *Dissemination* In Grid computing, the set of activities that aim at spreading Grid technologies and uses.
- *Training* In Grid computing, the set of activities that aim at educating on Grid technologies. Grid training may concern training end users, training application programmers, or training Grid technology trainers.
- *Command Line Interface* A mode of human-computer interaction where the user provides input by typing in a computer terminal and the computer responses by displaying messages on the terminal.
- Job Description Language A language that describes the tasks that constitute a job to be run on a Grid infrastructure.
- *P2P* Peer to Peer (P2P) is a network where participant are largely equivalent, and few, if any, centralised servers exist.
- *Interactive jobs* Jobs that interact directly with the user who submitted them, instead of simply running in the background and reporting their results. The term is usually employed for jobs requiring continual user input.