eInfrastructure Reflection Group

White Paper

Version 5.51

13th April, 2004

Extended Work in Progress

Subject to Further Discussion and Improvement

Table of Contents

0 Executive Summary

The eInfrastructures Reflection Group (eIRG) intends to support on the political, advisory and monitoring level, the creation of a policy framework for the easy and cost-effective shared use of electronic resources in Europe (focusing on Grid-computing, data storage, and networking resources) across technological and national domains.

In this way, the eIRG will cross-fertilize the major European grid activities in order to get the highest return on EU investments and remain at the forefront of the corresponding activities world-wide.

The eIRG White Paper support-team proposes the endorsement of the following two items related to authentication policies, while identifying areas, such as authorization, that need further elaboration and eIRG actions during the next presidencies.

Decisions for immediate action have to be taken to:

- Promote interoperable authentication and authorisation infrastructures enabling seamless sharing of eInfrastructure resources from network access to Grid interactions.

- Endorse the EU Grid Policy Management Authority www.eugridpma.org, as a group of mutually trusted Certification Authorities, and the TERENA TACAR www.terena.nl/task-forces/tf-aace as the common repository for storing and validating the CA root certificates and policies; these constitute concrete first steps towards common EU authentication policies, essential for resource access and sharing in the major FP6 e-Science projects.

Subsequent actions should be foreseen in order to

- Enable the use of federated solutions, decoupling local authentication procedures at a user's origin organization from local authorisation at the target resource. Origins and targets are connected by the trust links built by the federation.

- Apply techniques for privacy preservation, oriented towards avoiding unnecessary data leakage when performing AA interactions, providing users with the ultimate control over what information about them is exchanged for what transactions.

- Start working out policies towards resource sharing and accounting:
    - Intra-grid policies
    - Inter-grid policies

The latter actions are part of a policy roadmap proposed by the white paper support team at the end of the document.

1 Background – The New Environment of eInfrastructures

The explosive growth of technologies associated with computing and electronic communication is providing unprecedented opportunity for growth and change in society and has the potential to drive economies based on information and knowledge. Applications currently being developed in this dynamic environment require ubiquitous distributed electronic infrastructures, dubbed eInfrastructures, created from the integration of existing and developing research networks, large-scale computing fabrics, and nascent grid middleware environments. The exploitation of the many electronic resources (computing, storage and others) within the various user communities which are linked by broadband optical networks, is of key importance for the European Union and international scientific and economic communities. Initiatives such as the EU eInfrastructures initiative will help to make this vision a reality, by harmonizing policies governing the resource usage and in this way facilitating the collaboration of the user communities.

To date, the World Wide Web has provided transparent access to information for millions of Internet users. The new electronic infrastructures are intended to extend this to provide rapid, secure, and transparent access to distributed computing resources and services. This "World Wide Grid" of resources will form the basis of the Information and Knowledge Society, and will be built upon the software and hardware necessary to establish virtual collaborative environments, tools for education and research, planning and simulation tools for complex problem solving, economic modelling analysis tools, virtual environments for medical treatment, storage and analysis of high-resolution digital data, pictures, and video and for providing access to massive scientific databases for disciplines from bio-informatics and bio-chemistry to meteorology, physics, and astronomy.

Pioneering work in these areas has been done by the academic research and scientific communities. These electronic science (e-Science) applications are building the frameworks and creating the necessary impetus for the growth of the required architectures and standards. At the same time, these Grid technologies are being adopted by the wider community of the Information Society, with applications such as e-Government: civilian transactions with administrations and governments, e-Business: providing tools and services for business, and areas such as financial modelling, data storage and analysis for medical and pharmaceutical sciences, entertainment and advertising, and the simulation of complex technological systems.

Today, electronic research infrastructures are implemented through grids of computing and storage resources connected through electronic networks of local, national, and international scales. The field is now learning how to transform these research environments into production-quality infrastructures capable of supporting these communities.

These new technologies provide unprecedented opportunities for novel means of education, economics, collaboration, and scientific endeavour amongst others. However, they also bring new issues regarding policies that must be understood in this new environment in order to exploit their full potential. Such issues as models for acceptable resource sharing and accounting of the associated cost, entitlement of communities or individuals to access and use resources, responsibility, privacy, to name but a few, cross traditional national, economic, and political boundaries. This paper is intended as a first look at such issues, providing an overview of the current state of the art, bringing together experiences and knowledge gained by the current generations of grid and networking projects.

1.1 The EU eInfrastructures initiative

For the development and support of the eInfrastructures environment a series of workshops has been launched by the European Union under the aegis of the European Union Presidencies, see

http://www.einfrastructures.org, in cooperation with the European Commission. The "eInfrastructures" paradigm will reach its broadest scope and cross-border relevance, with policy decision mechanisms that will satisfy the diverse end-user communities' requirements of performance, service transparency and security, while achieving scale economies in providing ever-growing resources at attractive cost.

1.1.1 Athens, June 2003 - Launching the eIRG

On the 12th of June 2003 under the auspices of the Greek presidency of the EU, the 1st workshop was held in Athens, organised by the General Secretariat for Research & Technology (GSRT), the European Commission and the Greek Research and Technology Network (GRNET) in collaboration with the Greek National Documentation Centre (EKT). The workshop, entitled Towards integrated Networking and Grids infrastructures for eScience and beyond – The EU eInfrastructures Initiative, aimed at discussing the creation of the necessary policy decision mechanisms for the successful deployment of "eInfrastructures" within the extended European Research Area. Among the key recommendations of the workshop was the establishment of an **eInfrastructure Reflection Group (eIRG)** with a membership "built from national representatives". The eIRG "should consider and communicate clear messages on policy issues to both European Commission and existing infrastructure projects".

1.1.2 Rome, December 2003 - Consolidating the eIRG and defining the scope

On the 9th December 2003, the 2nd eInfrastructure Open Workshop entitled: "eInfrastructures (Internet and Grids) – The New Foundation for Knowledge-Based Societies" took place organized by the Italian Ministry for Education, University and Research (MIUR) under the aegis of the Italian Presidency of the European Union with the High Patronage of the President of the Italian Republic, Carlo Azeglio Ciampi, and in co-operation with the European Commission.

One of the primary goals of the meeting, at which key players in the construction of the EU eInfrastructure were present, was to review the perspectives and the technical and political issues related to the usage of the eInfrastructure for Science and Society at the national, European and international level.

1.1.3 Dublin, April 2004 - The first concrete steps

On the 15th of April 2004, the 3rd eInfrastructure Workshop, entitled: "eInfrastructures (Internet and Grids) – The New Foundation for Knowledge-Based Societies", will take place, organized by the Irish Office of Science and Techology (OST), Science Foundation Ireland (SFI), the CosmoGrid project and Grd-Ireland, under the aegis of the Irish Presidency of the European Union and in co-operation with the European Commission. The workshop aims at planning further initiatives to progress the discussions on Grid-empowered infrastructure and create new enhanced facilities for performing research and for fostering innovation in Europe. The key objectives of this workshop include:

- Create more awareness on the eInfrastructure initiative and involve other important actors currently missing from earlier initiatives;
- Decide on concrete next steps and actions in the context of the afore-mentioned policy framework in Europe across technological and administrative domains, for example;
    - *A concrete policy proposal relating to grid authentication;*
    - *The way forward for grid authorization policies;*
- Facilitate and support the next meeting of the eInfrastructure Reflection Group;

- Discuss and promote the establishment of cooperations between the EU-eInfrastructure programme with similar programmes of other regions of the world (most notably, the US Cyberinfrastructure and the Japanese GRID projects). Participants in the meeting will therefore also include international representatives.

The items that the Irish presidency choe to tackle were based on the recommendations that arose during the Rome eIRG meeting. These can be summarized as:

- create a registry of EU resources;
- create a repository of relevant policy documents;
- create a draft eIRG Questionnaire for acquiring resource and policy information;
- consider authentication and authorization (AA) policies;
- extend the eIRG White Paper.

The first two of these are at the elaborated prototype phase, supported by a XML database that is targeted for extension for semantic web access. The third, the Questionnaire, is thereby enabled to be supported by webforms for direct data entry to the XML database; the creation of the draft Questionnaire is deferred to the Dutch Presidency, who have experience of this within the Arcade community. The fourth item, AA, is being pursued very actively. Authentication is the special focus for the Irish Presidency. Authorization is also deferred to the Dutch Presidency, as they have particular interest and expertise in authorization, and this maximizes the possibility of a successful outcome and continuity for both Presidencies. The last item has given rise to this document, the Dublin eIRG White Paper.

1.1.4 eInfrastructure Projects – Dual role of generating and adopting policies

In this respect it is important to note that three significantly sized initiatives will be launched in the context of FP6 eInfrastructures, which are expected to help structuring the grid infrastructures in Europe and building upon the already established GEANT infrastructure. These projects – EGEE, SEEGRID and DEISA – will be major actors in this context as they both generate policies and are major players in adopting and generalising them.

- **EGEE- Enabling Grids for E-Science in Europe,** which is led by the European Organization for Nuclear Research (CERN), aims to build the largest international grid infrastructure to date, operating in more than 70 institutions throughout Europe, providing 24-hour grid service and a computing capacity comparable to 20,000 of today's most powerful personal computers. In EGEE a specific networking activity has been proposed to assist the work of the eIRG and a selected number of partners have committed to promote it.
- **DEISA**, which will build and operate a distributed terascale super-computing facility, whose integrated power will be close to 30 teraflops in 2004. The principle objective of this project is to advance computational science in leading scientific and industrial disciplines, by deploying an innovative GRID-empowered infrastructure to enhance and reinforce High Performance Computing in Europe. The proposed infrastructure is based on the tight coupling – using dedicated network interconnects – of six homogenous national supercomputers, to provide a distributed platform, and is based on an innovative operational model, capable of providing substantial European added value to the existing national infrastructures. The distributed multi-cluster platform is in turn integrated into a larger heterogeneous Grid.
- **SEE-GRID – South East European GRid enabled eInfrastructure Development,** intends to provide specific support actions to pave the way towards the participation of the SE European countries to the Pan-European and worldwide Grid initiatives. This will be accomplished through dissemination and training material including cookbooks, pilot and demonstration test-beds for hands-on experience, adaptation of applications to be able to use the Grid, operational

and support centre schemes and organisation, and finally feasibility studies and roadmaps for the integration of the SEE to the European Research and Innovation Area via an extended Pan-European eInfrastructure.

2 The eInfrastructure Reflection Group (eIRG)

As agreed during the 1st eIRG meeting in Rome-10 Dec. 2003

2.1 The Mission

The main objective of the eIRG is to support on the political, advisory and monitoring level, the creation of a policy and administrative framework for the easy and cost-effective shared use of electronic resources in Europe (focusing on Grid-computing, data storage, and networking resources) across technological, administrative and national domains.

2.2 eIRG Objectives

- To identify the fundamental fabric, services and resources needed to enable pan-European e-Science
- Recommend resource sharing policy guidelines to:
    - National Grid initiatives
    - Regional and European eInfrastructure projects
- Contribute to International policy fora
- Give input to other policy drafting bodies e.g. ESFRI, NREN PC, etc.
- Focus first on eScience application user groups (as enablers of novel architectures) but also address wider application domains (e.g. eLearning, eGovernment, eHealth, eCulture, eBusiness, etc.) within the European Research and Innovation Area.
- Identify, inform and promote GRID awareness among communities who can benefit from sharing resources
- Address Governance issues of Grid deployment
- Draw upon the experience of the NREN community (Structure, Operations, AUPs)

2.3 eIRG Structure

- Consists of appointed Member, Accession and Associated States Representatives plus EC officials
- Initially co-ordinated by rotating EU Presidencies with the scheme of a troika of past, current and future presidencies
- Supported by a Technical Support Group, via resources from EU flagship projects (EGEE, DEISA, GN2) and possibly other initiatives.

3 The White Paper

3.1 Introduction

An initial White Paper was drafted prior to the Rome meeting, as a first input to articulate the discussion, and this has become a live document to continuously support and reflect on the work of the eIRG. The editors form also a live entity, with active participation by FP6 projects such as EGEE, DIESA, SEE-GRID and GEANT. The editors of the Rome eIRG White Paper comprised the following members:

- Victor Alessandrini (CNRS-DEISA)
- Kyriakos Baxevanidis (EU)
- Ian Bird (CERN-LCG)
- Alan Blatecky (SDSC)
- Brian Coghlan (TCD-Ireland)
- Fabrizio Gagliardi (CERN)
- Francois Grey (CERN)
- Fotis Karayannis (GRNET) – editor
- Dave Kelsey (UK GridPP)
- Mirco Mazzucato (INFN)
- Jesus Marco, (UNICAN-CSIC)
- Federico Ruggieri (INFN)
- Matti Veikko Johan Heikkurinen (CERN)

The editors for the Dublin eIRG White paper comprised the following members:

- Patrick Aerts, aerts@NWO.NL,
- Kyriakos Baxevanidis, Kyriakos.Baxevanidis@CEC.EU.INT,
- Kors Bos, k.bos@NIKHEF.NL,
- David O'Callaghan, david.ocallaghan@CS.TCD.IE,
- Brian Coghlan, coghlan@CS.TCD.IE,
- Licia Florio, licia@TERENA.NL,
- Antonia Ghiselli, antonia.ghiselli@CNAF.INFN.IT,
- David Groep, davidg@NIKHEF.NL,
- Peter Hanak, peter.hanak@NKTH.GOV.HU,
- Matti Heikkurinen, Matti.Heikkurinen@CERN.CH,
- Eduardo Jacob, jtpjatae@BI.EHU.ES,
- Jens Jensen, J.Jensen@RL.AC.UK,
- Christos Kanellopoulos, skanct@PHYSICS.AUTH.GR,
- Fotis Karayannis, fkara@GRNET.GR,
- Dave Kelsey, d.p.kelsey@RL.AC.UK,
- Diego Lopez, diego.lopez@REDIRIS.ES,
- Paolo Malfetti, p.malfetti@CINECA.IT,
- Mirco Mazzucato, Mirco.Mazzucato@PD.INFN.IT,
- Tom Sheedy, tom.sheedy@ENTERPRISE-IRELAND.COM,
- Ferenc Vajda, vajda@SZTAKI.HU,
- Dany Vandromme, vandrome@RENATER.FR,
- Jules Wolfrat, wolfrat@SARA.NL,
- Anders Ynnerman, andyn@ITN.LIU.SE

The White Paper series should summarize trends concerning general policies used to address resource access and sharing at the pan-European and International level across different technological and administrative domains. This version focuses on the first major issues that need to be addressed, such as authentication and authorisation, and identifies a roadmap for future development of a political framework in Europe and beyond, in order to allow a real effective exploitation of the eInfrastructures.

Initial input has been taken from the major Grid projects, like the EU Datagrid project, major Particle Physics Grid project such as the Large Hadron Collider Computing Grid project (LCG), other national Grid Initiatives such as the UK e-Science programme, the Italian INFN grid, NorduGrid, and what has been achieved in major Super-computing centres and test-beds, e.g. the Teragrid in the US. In addition the white paper has taken into account the experience of the National Research Networking Community (NRENs) for both technical and policy issues.

The white paper identifies a roadmap for the future development of a political framework in Europe and internationally, to allow a rapid deployment and effective exploitation of the eInfrastructures.

3.2 Methodology

The following methodology constitutes the basis of the White Paper. The policy framework will be the outcome of two main streams. The first stream will be the usual "use-model, requirements-capture, use-case-specification, architecture" chain, where the entities and their roles are first identified, their requirements are captured and analysed along with use case specifications resulting in the policy framework layout. The second stream is the capture and analysis of current practises and achievements in resource access and sharing. The combined analysis of the two streams will provide the policy framework and will include a series of policies. Note that at this stage a formal requirements capture process has not been started, but this is now enabled to be prototyped through corresponding questionnaire webforms during the Dutch Presidency. Nevertheless, thus far input has been taken only from current practises.

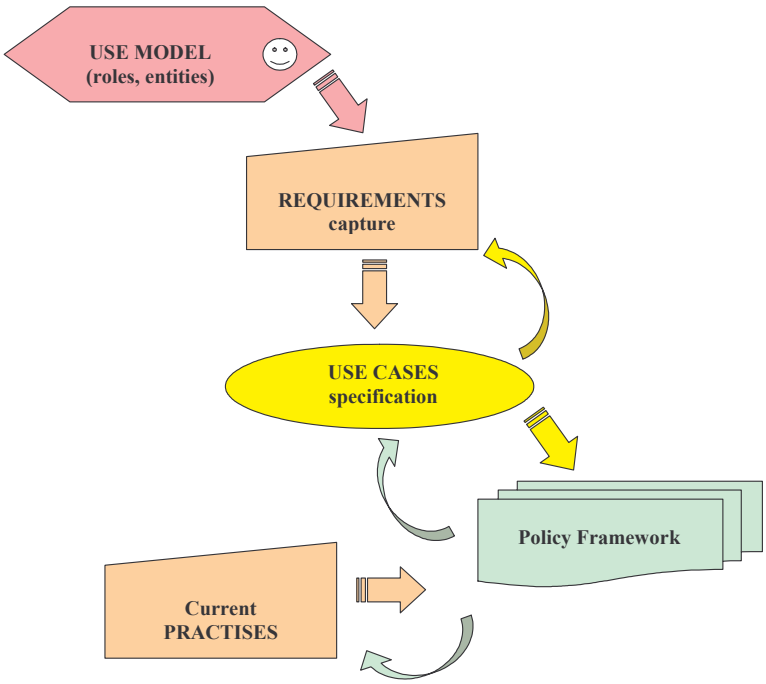


figure 3-1 – White paper methodology

4 Use Model

4.1 Introduction

This section provides a draft policy architecture and use model (or business model) including all entities (or actors) of the eInfrastructure, e.g. the users, Virtual Organisations (VO), middleware providers and operators, resource providers and operators, network providers and operators, etc. Their interactions are also depicted in a corresponding diagram. At a later stage the administrative domains should also be visible. In this section the roles of the different entities and their responsibilities will be highlighted. Finally some basic use cases will be elaborated (to be refined after a detailed requirements capture and analysis following a questionnaire). At this first attempt significant input has been received from the corresponding work of on-going projects and external references (such as the LCG Security Group on *Security and Availability Policies for LCG)*. Other projects or national initiatives are encouraged to participate.

4.2 Entities – Roles and Responsibilities

4.2.1 Policy Architecture

Different frameworks for a "production-quality" operation of Grid-aware distributed computing systems have been discussed in many projects and in dedicated sessions at workshops. These include the experience in current large test-beds, like the DataGrid, CrossGrid, and LCG projects; the GridStart inventory document; the London workshop on VOs in May 2003, the Brussels concertation meeting in June 2003; the Production Grid Management, Grid Economic Services Architecture and SAAA GGF working groups; and the chapters on "Grid resource allocation and control using computational economies" in the book "Grid Computing"[1].

Grids are defined as frameworks enabling coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations. The framework is expressed with a layered Grid architecture, as is the case in the Lightreading report on the "Architecture of the Grid"[2]. The specific architecture has been preferred since the different layers correspond to discrete physical components and can be easily linked with appropriate entities and roles. The basic layers of a Policy Infrastructure are the following from bottom to top:

- The network layer, providing the interconnectivity for the components using the respective networking equipment (routers, switches, etc.)
- The resources layer, providing the actual fabric resources excluding network resources belonging to the network layer, such as computational, storage or other (e.g. sensors, telescopes, instruments and information resources etc). The resources are part of a resource center (RC).
- The middleware layer, providing all the protocols and components enabling the sharing of resources. The basic middleware components facilitate among others information services, resource allocation and scheduling, security emphasizing on authentication and authorization, monitoring and discovery services etc. The middleware layer encompasses the "connectivity", "resource" and "collective" layers according to most other layered architectures as the in "Anatomy of the Grid".
- The application and service Layer, providing the actual applications of different scientific or business fields, along with supporting portals and development toolkits.

[1] http://www.grid2002.org/
[2] http://www.lightreading.com/document.asp?doc_id=33405&page_number=4

The architecture should be enhanced with the distinction of **collective** vs. **non-collective** middleware layers, since this is closely related with the likelihood of different administrative domains providing these two services. In essence, non-collective middleware is used by a single entity to publish services of its local resources and even charge by itself, while collective middleware is used to publish services and manage multiple resources across domains.

According to the definition of the eInfrastructure, the 3 bottom layers are those constituting the **eInfrastructure.** The users or consumers interact with the applications that sit on top of the eInfrastructure, either directly or through appropriate Virtual Organisations.

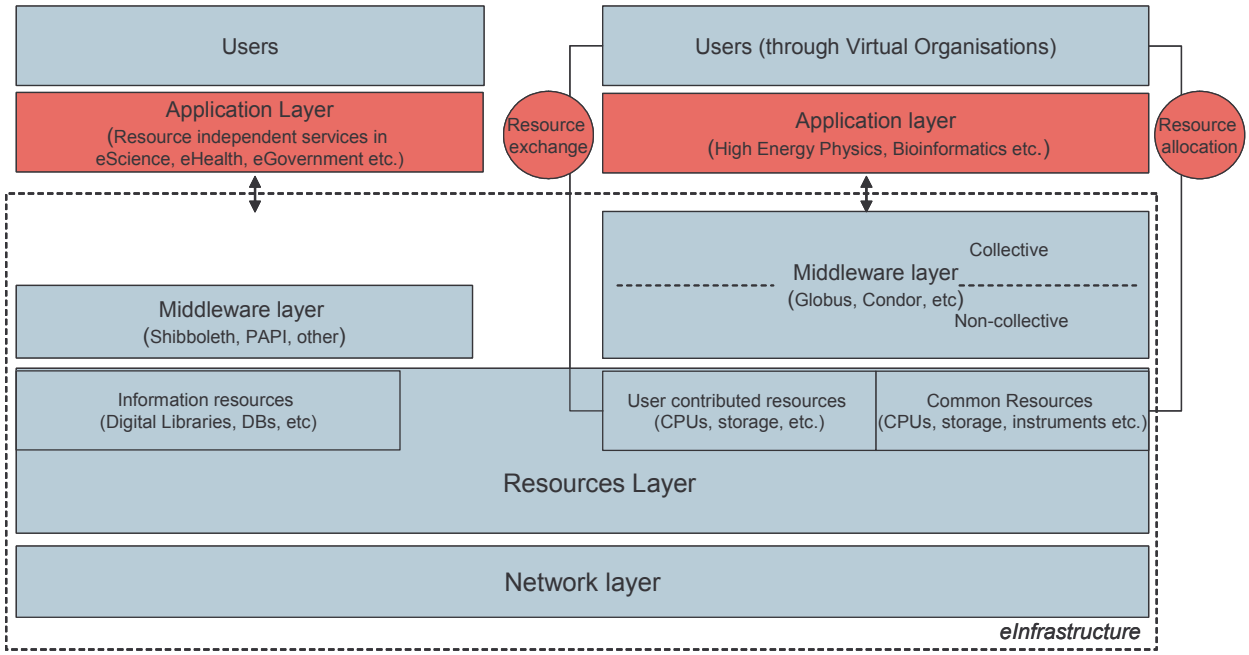


figure 4-1- Policy Architecture

4.2.2 Entities-Roles

The distinction between "users" or "consumers" and "providers" is a basic principle for discussing the sharing and access to resources. This is important because the "computational economy" is usually blurred - especially in an e-Science environment where parts of the grid computing resources are directly operated by the user communities themselves. Even when this is not the case, these resources are usually funded at national or institutional level, with the primary objective of satisfying the needs of their national or institutional community. In the academic framework, there is also a long tradition of sharing resources between different projects.

Administrative domains (national, local, etc) must also be taken into account. The current experience on the operation of network infrastructures can provide an indication of the path to follow.

Keeping in mind the above architecture and basic principles, we can identify the following roles:

- **Application Layer**
  - **Users** (or end-users or consumers**),** interacting with the application and service layer, running their applications on the Grid and actually utilising the eInfrastructure. The applications are usually related to electronic Science (e-Science), but there are some first examples of electronic business or government (e-Business, e-Government, eHealth), where the use of a Grid will be expanded at a later stage. In research and education environments the purpose of eInfrastructures is for e-Science. Users should be able to obtain suitable authentication credentials containing their identity, usually signed by one of the Certification Authorities recognised by the eInfrastructure, and then be authorised to use the eInfrastructure resources.
  - **Virtual Organisations (VOs),** each of which is a (sub-)group or association of users collaborating in a common experiment, project or other a joint venture. The Virtual Organisation has one of the most important roles in the eInfrastructure but adds great complexity, since VOs are formed as a selection of users belonging to different administrative domains. VO users are legally bound to the institutions where they work (in an e-Science framework usually working in different projects through **collaborations**, with a financial framework defined through a MoU or Annex). The legal responsibilities require the use of individual authentication, with certificates explicitly including the legal Institution, while resource usage is contemplated at the Collaboration/Project level.
- **Middleware Layer**
  - **Middleware providers (or developers),** implementing, testing, supplying and maintaining bundled releases of the necessary software. Besides the functional characteristics of the software, special attention has to be paid to non-technical properties such as robustness and reliability as well as to the security aspects. Currently there are multiple middleware packages as well as different versions of the same package. Thus, an important aspect of interoperation among eInfrastructures is the use of standard interfaces and/or common middleware packages. A European Open Middleware Institutes Initiative (OMII) and/or similar national efforts are of key importance for eInfrastructures and could greatly assist towards appropriate services reengineering and interoperability purposes. As mentioned it is important to distinguish between collective and. non-collective providers, since this might involve different administrative entities. However, this is much more important for the operational aspects:
    - **Non-collective** middleware providers: This includes the publication of the available resources, the reservation and access by authorized users, as well as related accounting services;
    - **Collective** middleware providers, including among others meta-directory services, co-allocation/co-reservation/brokering services, replica location services along with monitoring and diagnostics services.
  - **Middleware operators,** responsible for operating the eInfrastructure middleware, either as part of their Grid Operation Centers (often referred to as GOCs, although the consistency of terminlogy could be improved) or as common services like Authentication and Authorisation Infrastructures (AAIs) or metadirectories. These operators must guarantee an agreed quality level. It is obvious that multiple components are mandatory for scalability and redundancy reasons; however different providers can focus on the provision of different services. From an administrative point of view, they can be assigned to a single legal entity, or can be organized in a distributed way with different partners offering different services, under a common project or a signed MoU.

As an example, in the EGEE project there is a hierarchy of GOCs. The non-collective services are offered by the Resource Centres, while regional Grid Operation Centres (ROCs in EGEE) provide a collective service to support middleware installation and user support for a set of Resource Centres that are geographically close. Note that this cannot be defined as a "traditional grid service" that can be invoked inside an application. Another layer above in the hierarchy are the Core Infrastructure Centres, which are centrally located, and provide most of the collective services, including the Information and Replica Index, or Resource Brokering, and also collect the accounting information from the set of Resource Centres they serve. At a higher level, the Operation Management Centres (OMC) provide services like the coordination of Grid operation, definition of Service Level parameters, coordination of security activities and also monitoring of service performance levels.

Several NRENs are currently acting as middleware operators, most notably in the areas of directories and AAIs. Several ongoing international efforts (within Europe and beyond) are active in the definition of common syntax and semantics for directory attributes. With respect to the AAI arena, there are a number of approaches for federated mechanisms that allow local authentication to provide access to remote resources. Among these are Shibboleth (the system developed by Internet2), PAPI (the system created by the Spanish NREN RedIRIS), A-Select (developed by the Dutch NREN SURFnet), and FEIDE (developed by the Norwegian NREN Uninett). New developments are envisaged in the GN2 proposal to allow inter-realm authorisation, and maybe inter-realm technology gateways. Concurrently new standards are being developed and set, e.g. SAML, XACML, etc. These need to be assessed in order to make the existing solutions migrate to the new standards, thus enhancing world-wide interoperability. NRENs recognize Grids as a fundamental domain and seek for enabling seamless interactions among middleware components in the whole eInfrastructure.

- **Resource Layer**
  - **Resource providers (or resource centres)** supplying the Grid fabric resources. The grid resources are either part of an organised resource centre (RC) (with computer clusters or supercomputers), or they are individual resources, perhaps part of a desktop Grid. The resource centres will allow access to all or part of their CPU and storage resources (as computing centres), through grid-specific network transactions. Resource centres can range from small (departmental) to large multi-function facilities. From an administrative point of view, they have a well-defined legal status and identity. From an economic perspective they can also be seen as a basic entity on the “production” side.

  - **Resources (or fabric) operators or administrators,** responsible for operating the Grid fabric in the organised resource centres. In case of desktop resources, the operation of the resources is outsourced to the resource centre operators or to the administrator of the department where the resources are located. In case of organised resource centres a designated resource centre manager is the interface with the rest of the world.

- **Network Layer**
  - **Network providers** supply the underlying networking connectivity. The adopted model in the research networking community is hierarchical. The resource centres are usually connected to a campus LAN (inside a University or Research organisation), the campus LAN is interconnected with the National Research and Education Network (NREN) and the NREN is connected to the pan-European Research and Education network (GEANT). End-to-end connectivity relies upon multiple administrative domains, i.e. campus-NREN-GEANT-NREN-campus, and in some cases there is also an additional regional network, either inside a province in a country or among multiple countries. The network providers rely mostly on leased capacity from national or international carriers, but in some cases they own their networks (having acquired Indefeasible Rights of Use in optical fibres from third party carriers or building their networks from scratch. Service Level Agreements (SLAs) exist between the research networks and the carriers in terms of service availability, Mean Time Between Failure (MTBF), Mean Time To Restore (MTTR), etc. SLAs between GEANT and the NRENs do not exist yet and will be studied during the GN2 lifetime.
  - **Network operators and administrators**, responsible for operating the networking part of the eInfrastructure as part of their Network Operation Centers (NOCs). Due to the multi-administrative domain environment, it is obvious that there are multiple NOCs, each responsible for its own domain. In other words, GEANT, Regional networks, NRENs or Campus networks operate individual NOCs, which cooperate with one another under agreed procedures. Service level agreements are used to guarantee an agreed quality level. Network operators, in cooperation with the corresponding administrative structures of each NREN and of GEANT, are responsible for safeguarding their existing Acceptable Use Policies. Note that GEANT is content with the national Acceptable Usage Policies (AUPs) of each NREN and does not possess its own AUP. However, in order for an NREN to connect to the GEANT network a list of high-level requirements must be met, see http://archive.dante.net/geant/connect.html.

4.2.3 Responsibilities

Based on the above entities and roles we can identify the following tentative responsibilities per category (using input from the LCG Security Group), which constitute one possible Policy. Given the subject it is difficult to avoid a legalistic use of language.

- Users:
  - Safeguard Credentials and Private Keys: Users must ensure others cannot use their credentials to masquerade as them or usurp their access rights. The holder of a private key will be held responsible for all actions, whether carried out by the holder personally or not, carried out using credentials generated from that key. No intentional sharing of credentials is permitted.
  - Observe Access Controls: Users must be aware that their jobs will often be running on equipment and using resources owned by others. They must observe any restrictions on access to resources that they encounter and must not attempt to circumvent such restrictions.
  - Observe Limitations on Use: Resources may be used only for legitimate professional purposes connected to the purpose of the eInfrastructure. Personal use of any nature is expressly forbidden.

- Applications: Applications software written or selected by Users for execution using the eInfrastructure Resources must be directed exclusively to the legitimate purposes of the latter. Such software must respect the autonomy and privacy of the host sites on whose Resources it may run.
- Respect for Others: Users must be aware that their work may be utilising shared resources and may seriously affect the work of others. They must show responsibility, consideration and respect towards other users in the demands they place on the eInfrastructure.

- Virtual Organisations:
  - User Registration: The resources centers' institutes and the Virtual organisations are required to set up and operate a set of Registration Authorities and associated procedures for approving requests for accessing resources. Approval must be restricted to individuals who are recognised as having legitimate rights to membership. RC institutes and VOs are subsequently required to maintain the accuracy of the information held and published about their members, and to promptly remove membership from individuals who lose their right to membership.
  - Controlling Access to Resources: Some resources will be restricted to all members of a certain institute or VO or to certain individuals within the institute or the VO. VOs will provide access to information about their members as necessary to enable such controls to be implemented and maintained accurately.
  - Applying Sanctions to Users: RC institutes and VOs are responsible for investigating reports of users failing to comply with the provisions of this Policy, and for taking appropriate action to ensure compliance in the future. This action may include the notification and involvement of the User's home institute. The ultimate sanction to be exercised at the discretion of the institute or VO is the removal of membership, and hence the withdrawal of rights of access to the eInfrastructure resources.

- Middleware providers
  - Facilitating Security Controls: The software should implement appropriate security techniques to control access to resources of all types.
  - Maintaining the Integrity of Services: Before distributing replacements, upgrades or patches to existing software, developers must ensure that adequate testing is carried out to ensure the functionality and reliability of existing Services will not be jeopardised. When carrying out tests, developers will follow current best practice. This requirement may be relaxed if it is imperative that a security-related patch be distributed urgently.

- Middleware operators:
  - Contact details: The GOC is responsible for maintaining contact details of security personnel at each participating resource centre and for facilitating eInfrastructure-related intercommunications between them.
  - Monitoring SLAs: The GOC is responsible for monitoring the operational performance of the eInfrastructure services and for publishing details of its findings for comparison with the published SLAs of those services.
  - Security Expertise: The GOC and appropriate support teams are responsible for establishing and maintaining expertise in eInfrastructure-related aspects of security in order to provide detailed advice and guidance to the community on avoiding and responding to internet security incidents.

- Resource providers
  - Quality Services: Resource Centers (RCs) accept the responsibility for providing quality services to their users.
  - Risk Assessment: RCs providing resources to the Grid acknowledge the risk of intrusions and host compromises and are responsible for assessing and minimising the risks. RCs should take the necessary measures to safeguard their resources.
  - Cooperation: In case of security incidents, RCs accept the duty to cooperate with the other structures in order to investigate and resolve the incidents, taking the appropriate actions and sanctions.

- Resource operators
  - Site Policy: Resource operators must ensure their implementations of services comply with both their RC policies and this Policy.
  - Notifying Site Personnel: Resource Administrators are responsible for ensuring that all appropriate personnel concerned with security or system management on their site are notified of and accept the requirements of this Policy before implementing any services.
  - Resource Administration: The Resource Administrators are responsible for the installation and maintenance of Resources assigned to them, and subsequently for the quality of the operational service provided by those Resources. This quality will be defined by the Service Level Agreement for each Resource as published by the Administrator of that Resource.
  - Service Level Agreement (SLA): The Administrator of each service instance must maintain an assessment of the risks inherent in their particular Service design or resulting from local services or operational practice which might affect that Service's Availability, Reliability or Performance, and publish the expected values of these service parameters in accordance with the GOC procedures for Resource Administrators.

An additional issue in all the security-related aspects is the need to develop a taxonomy for different levels of security that different application domains or resource providers require (e.g. strength of the authentication mechanisms, level of intrusion detection mechanisms, etc). A possible approach would be to take the cluster security-level classification of the Carrier Grade Linux initiative[3] as a starting point and develop a similar criterion for the eInfrastructure components and organizations providing services based on them. A summary of the OSDL model and a sketch of a Grid-variation is presented in the Appendix A.

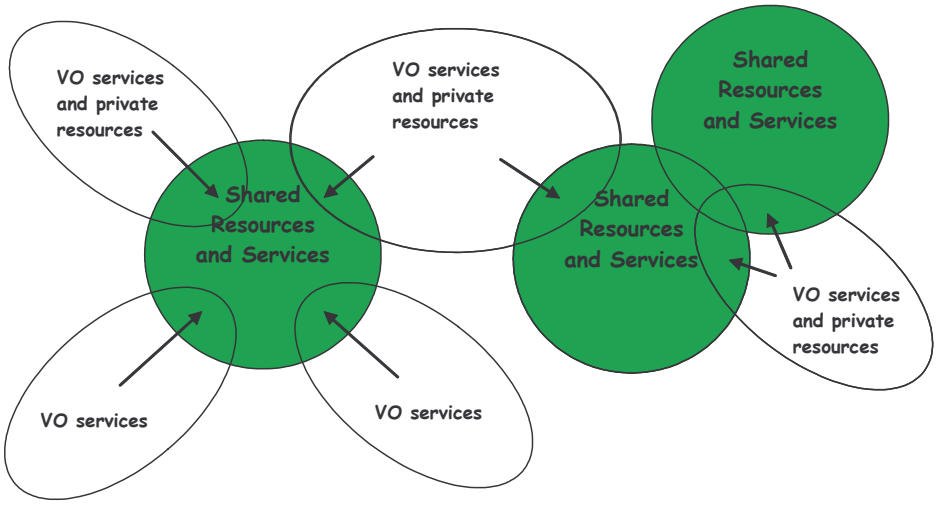
4.3 Elaborated Use Model

The production use of eInfrastructures requires the definition of new frameworks, where management organization, advanced access, sharing and accounting rules must be identified. This is the real challenge for the concept, to implement a coordinated resource sharing on a large scale for a dynamic and multi-institutional virtual organization. By way of example, let us concentrate on Grids, where the new paradigm consist of moving from "computer sharing" to "grid sharing", and from "multiple users" to "multiple VOs".

A VO stands for an international dynamic collection of users from one or more physical organizations. Usually VOs are formed to tackle large-scale scientific problems and in many cases

[3] http://www.osdl.org/lab_activities/carrier_grade_linux/documents.html

they provide their own resources or part of them. Moreover each physical organization of the VO can access its own national Grid. The general, global scenario can be depicted as below:

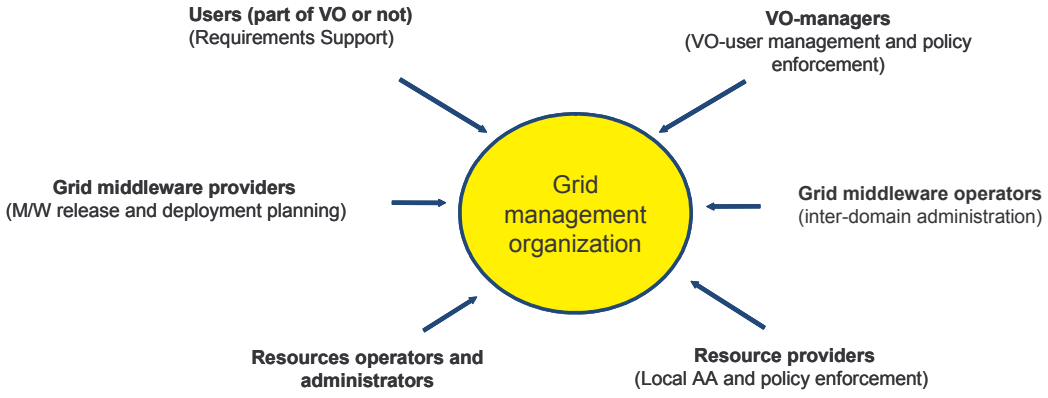


Inter-grid and multi-VO scenario

figure 4-2 Generic grid use scenario

4.3.1 Use model entities interactions

Today in most European countries national grid infrastructures for research communities are growing, and the recent European projects have demonstrated the technical capability to provide a grid computing (or other) service. But grid management organizations able to provide a so-called 'production service' are not yet in place, and policy management frameworks together with grid accounting systems are at their early stage, even technically. A tentative schema collecting all the entities of a grid user model is drafted below. Note that since it is not easy to show all the interactions of the use model entities, it is preferred for now to show all players interacting through a central management organization:



Network layer entities are not shown for simplicity: e.g. supposing a best-effort network service

figure 4-3 - Simplified Use Model

The entities and their interactions (related mainly to policy enforcement issues) are briefly summarized below:

- Users request access to grid computing services and support in case of needs. They interact with VO-manager and resource operators (e.g. authentication, authorization) and other entities if necessary.
- VO-managers handle authorization, group membership, roles of the users and moreover explicit policies for governing VO operations. They interact with the users and with grid middleware operators and resource providers, if necessary.
- Grid middleware operators deal with the grid as a whole, and interact from one side with VO managers and on the other side with resource providers, technical and operational groups. They are also concerned with policy and accounting rules.
- Resource operators and administrators provide grid fabric functionality, performance and support. They interact with resource providers, middleware providers and operators, as well as with most other entities.
- Grid middleware providers provide grid middleware release and deployment planning, middleware updates and services configuration. They interact with resource operators and administrators and middleware operators and other entities.
- Resource providers, make sure that local resources provides the computing and storage services and apply the agreed policies.

The above structure can be modelled with a hierarchical, layered framework as follows:

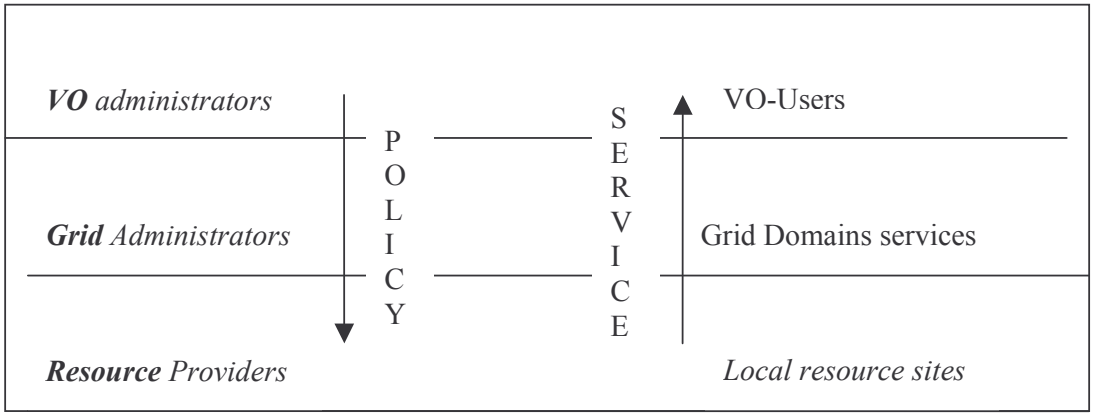


figure 4-4 – Policy enforcement and service implementation paths

Policy and resource usage contracts, agreed between VO, Grid and resource administrators, are flushed from VO servers down to resource servers and vice versa. Policy application ensures secure and efficient grid computing and storage services to the users.

4.4 Use Cases

In this section, some high-level Use Cases in the academic and research environment are given. Note that these refer to the portion of the policy architecture referring only to electronic science (eScience).

- The Scientist and the Idea. A geneticist at a conference, inspired by a talk she hears, will be able to access the key parameters of her model immediately and launch a complex biomolecular simulation with the new values from her mobile phone. The result will eventually tell her she has discovered a new drug candidate.

- The Student and the Insight. A team of engineering students in Bucharest will be able to remotely run the latest 3D rendering and CFD programs from their laptops using an eInfrastructure. The result is a sudden insight into how to improve the engine they are designing.
- Spontaneous Collaboration. Two scientists meet at a conference and notice that they could collaborate much more efficiently by combining their datasets. Instead of trying to convince the computing departments of their universities to give accounts to each other, they will form a two person *Virtual Organisation* by using the appropriate tools, which together with standardized data access interfaces, will allow them to view their separate datasets as if they were a single entity. The ontologies needed in the integration of their databases can easily be published and be reused by the other researchers in the field. This will speed up the adaptation of common terminology in the research field and related fields of study who will share data with them.
- Mobility of students. By switching to a federated service based on virtual organisations as a means to control the access to the basic infrastructure services (network connection to laptops, access to temporary storage etc), the administrative costs associated with student exchange will go down. By deploying advanced virtual organization solutions able to preserve user privacy, this model can also be extended to the access and update of student records. This will enable teachers to select student eligible to participate on courses and mark grades after the completion of them in uniform manner independent of the home institute of the students.
- Rationalizing the use of academic infrastructure. By pooling the resource from several universities into a single eInfrastructure the participating institutes can optimally leverage various economies of scale benefits in procurement by concentrating acquisition to a single entity and managing cost sharing based on the use of each of the institutes. Even inside one academic institution benefits can be obtained, e.g. departments sharing their resources.

5 Current Practices and Achievements in Resource Access and Sharing

5.1 Introduction

This section summarises the current achievements and best practises on policies concerning resource access and sharing at pan-European and International level across administrative and national domains. The best practices and achievements include issues in the area of resource access and sharing focusing on standard schemas mainly for AA, i.e. Authentication and Authorisation. ***As decided during the Rome meeting, initial emphasis will be given to Authentication and Authorisation, while the Irish presidency will pursue Authentication issues first***. In addition, in view of the urgent needs of the FP6 projects such as EGEE, DEISA and SEE-GRID, there is a special focus in this Presidency on the existing PKI-based Grid Authentication established by DataGrid. It is likely that alternative schemes for other constituencies will be examined in depth in future Presidencies.

5.2 Resource Access and Sharing Schemas

5.2.1 Introduction

In this section the technical schemas are analysed in the different areas, starting from the AAA, i.e. Authentication, Authorisation, Accounting, etc. The following diagram shows a simplified view of the steps the user needs to go through before (s)he can use Grid resources. First, the user obtains a certificate from a Certification Authority (CA). Obtaining a certificate does not authorise users to do anything; the user may ask resource administrators to be authorised to access their resources using the new identity, but this process is not scalable and does not preserve the user's privacy: a scalable and privacy-preserving authorisation is obtainable through a VO.

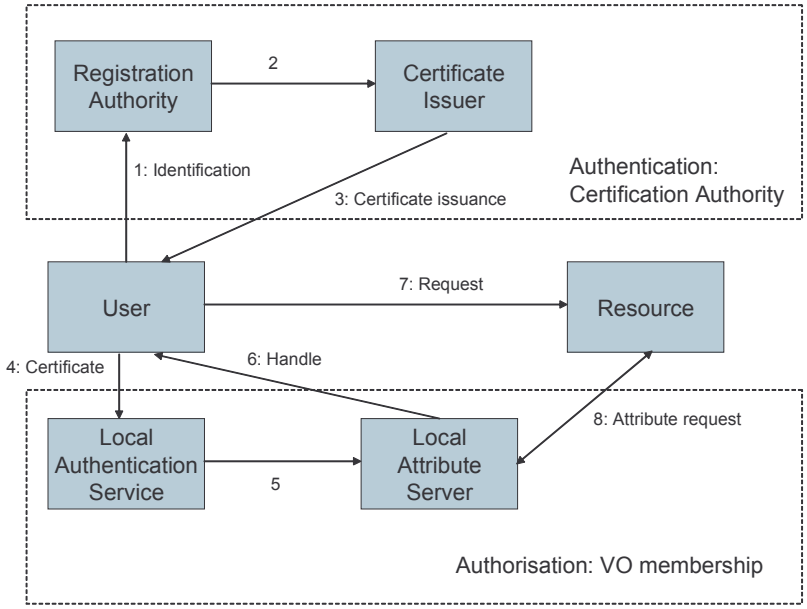


figure 5-1 – Authentication and authorisation interactions

The user identifies at the local authentication service using the certificate and obtains a opaque handle provided by the attribute server. Once the user tries to access a resource, it passes the handle that allows the resource to verify with the attribute server the user's VO membership state. It is important to realise that the CA does not authorise access to anything. The user must explicitly ask for authorisations, either from the resource administrators directly, or via one or more VOs. The description above only gives an outline of the process. Further details can be found in following sections.

5.2.2 Authentication

Usually, access to resources requires one to authenticate and give out information regarding identity to the resource provider. Traditionally within a single administrative domain (a university campus, an institute, or a corporation) there have been a wide variety of non-interoperable means of authentication, ranging from extremely light-weight (e.g. supplying an e-mail address), to somewhat secure (username and password combinations), to very strong (involving the use of biometrics or multi-factor authentication based on smart-cards and one-time access codes).

When talking about authentication is important to establish a distinction between authentication procedures and authentication schemas. Authentication procedures deal with the process of establishing a user's identity and identifying the appropriate identity attributes. Authentication schemas determine the components providing the authentication service, the interactions between the user and the service, and the interactions among the different components. The experience of the European National Research and Educational Networking organisations (NRENs) has shown that most (if not all) of the authentication procedures currently in use are compatible with any authentication schema. A recent survey of authentication practices amongst the European NRENs conducted by the TERENA task force on authentication and authorisation coordination in Europe (TF-AACE), reported on the procedures and commented on specific authentication schemas:

A Classification of Authentication Schemas
Independent criteria can be established to classify the authentication schemas that European NRENs currently use, depending on whether the authentication procedures are applied by specific middleware or application services, or at the network access points. In the first case, identity attributes are usually provided to the authorization elements by means of middleware components known as the AAI (Authentication and Authorization Infrastructure). In the second case, user's identity is usually linked (by some automatic procedure or by means of log entries) to the address assigned at the network access point.

A special situation being considered by several NRENs and within the TF-Mobility group) is the case of mobile users employing wireless network access. Several initiatives to unify authentication procedures in a single middleware layer are under development. One issue is whether the user's identity (and/or attributes) are determined by means of local sources of data, or the user's identity is based on external resources. The latter case introduces the need for trust models at the authentication phase. Another issue is whether the exchange of user's credentials must take place in well-defined and controlled points, or it is allowable to use any (possibly registered) point to request these credentials and then forward them to the authentication service.

As said above, any of the authentication procedures discussed in the following sections can be applied to any kind of schema (as described by the above criteria). For example, it is possible to establish an authentication service for network access based on either smart cards and only using corporate directories and access points, or on an LDAP-based service that uses institutional

directory servers and allows applications to ask users for their username/password pair, forwarding it to the authentication service.

Although the degree of deployment and maturity of the different infrastructures varies very much from one country to another, it is possible to identify four different approaches to authentication procedures inside the institutions served by European NRENs. All of them try to keep authentication data management within the scope of the institutions, while they differ in the way in which data are exchanged to authenticate users, and in the formal requirements and procedures to grant users for authentication accounts. There is a distinction between specific authentication procedures (such as smart cards and the like), more practical approaches using ad-hoc pre-existing methods (there is a great variety of them), and those that combine local ad-hoc methods with some central service. The four approaches are:

- Central Authorisation Servers: a central server holding all user data (identifiers, passwords, and, possibly, access rights);
- PKI-based Authentication: PKIs are the solution cited by many of the NRENs participating in the survey, although they are not well deployed yet, and their use is often mentioned as a future approach. In those cases (a couple of them) where PKI is mentioned as an actual alternative, its deployment is not reported to be very advanced. A special case in this respect is the Grid authentication PKI described below;
- Authentication Based on Pre-existing Services: The common approach to user authentication in those institutions not running a operative PKI has been to employ a service that requires user identification (typically, by means of an username/password pair) and that was already offered at a institutional level;
- Central Directories plus Local Identity Services: A solution that has been described in several cases consists of the combination of a central directory with local identity services (as described in the point above). The central directory may be constructed as an actual centralized directory service, or by means of a combination of indexing and searching procedures on locally managed directory servers.

More details on the authentication mechanisms are to be found in the report[4]

Public Key Infrastructures and the distribution of trust anchors

With the increased need of inter-organisational trust building, and the wide availability of cryptographic techniques supporting the concept of key pairs with both a public and a private part, the concept of a public key infrastructure (PKI) has gained a certain level of acceptance in controlled environments. Well-known examples are the PKIs used to secure transactions in electronic commerce – although the security based on the PKI usually works in one direction only: the web site is authenticated using PKI, but the user generally still has a username/password combination.

The general public implicitly trusts the PKIs used to secure e-commerce transactions, since their roots of trust are shipped with the popular web browsers. Although the software distribution itself is usually not secured (e.g. downloads via the web from unsecured sites), the implicit nature of the trust relationship escapes most users of these PKIs.

PKIs are also being used for authentication purposes in the academic community, to secure access to networks, to send authenticated electronic mail, and recently to identify end-entities and services on the Grid. In this community, the concept of implicit trust as seen in the web world is neither attainable nor desirable. Many academic institutions (traditionally supported by their academic network provider) have established national academic PKIs with specific policies and practices that

[4] http://www.terena.nl/tech/task-forces/tf-aace/Del/B.5/TF-AACE-B5v3.pdf

are not necessarily equivalent, and moreover such new "roots of trust" are not distributed by default by the common software vendors for commercial reasons. The certificates issues can be used for many different purposes, depending on national (or site) policies and practices.

Within the Grid community, a single common trust domain has been created between major 5th framework projects (DataGrid, CrossGrid), many national projects, and key external parties from the US and the Asian Pacific region. In this trust domain, organised via the DataGrid-hosted Certificate Authority Coordination Group (CACG), a single minimum specific is adoption by many different CAs so as to establish a domain of equivalent trust for Grid authentication only.

One of the key problems linked to the cross-domain use of any Public Key Infrastructures (PKI) is how to get all the different trust-anchors into user's browsers and other applications in a practical and cost-effective manner. A possible solution that can be applied within the academic community is the use of a process for gathering and verifying academic root-CA certificates, allowing publishing them in one easily downloadable and importable trusted file.

It is useful to note that this trust-anchor problem is common to all PKIs and PKI-inspired infrastructures, i.e. both the traditional academic PKIs sponsored by the national research networks as well as the ensemble of CAs that constitute the common trust domain for Grid authentication. This commonality is exploited by maintaining a common repository of trust anchors from which all PKI roots in Europe (and those for related communities and countries) can be obtained, and that is operated by a widely accepted reputable body like TERENA.

Grid Authentication for e-Science infrastructures

Given the youth of grids, grid security implementations are remarkably developed, but still fall far short of maturity. Most grids rely on the Grid Security Infrastructure (GSI), which is loosely based on a Public Key Infrastructure for authentication. Unicore security also integrates the PKI schema.

Here it should again be noted that the certificate is just a passport that identifies the entity to the grid. It does not authorize the entity to use resources in any way; one can consider authorisation attributes like "visas" stamped into a personal passport.

In comparison to traditional PKIs, like those used for e-commerce and in the academic network community, a PKI for Grid authentication has several characteristic distinctions:

- The primary goal of GSI is to provide a 'single sign-on' capability to the grid users. To this end, the validation process that enforces basic constraints expressed by the trust provider (CA) is modified. This in effect constitutes a violation of the policies for any traditional PKI. The CAs that join the trust domain for grid authentication explicitly accept these documented breaches of policy;
- There are technical considerations that require a specific mode of operation of the PKI used for GSI purposes. The middleware commonly used for GSI does not readily allow for hierarchical structures within the PKI – a concept that has been central to many regular PKIs. The distributed and dynamic nature of the grid precludes the use of directories for chain retrieval, since that would easily create a single point of failure in the Grid. Also the technical profile of the certificates issued is subject to implementation-defined constraints. Certain fields must be interpreted or set in one specific way for the Grid software to operate correctly. This may violate the commonly accepted interpretation of the PKIX standards;
- Since the PKI for Grid is part of the protection scheme for valuable resources, the validity of the certificates issued by the PKI must be accurate to a high degree. To this end, the timely distribution and enforcement of "certificate revocation lists" (CRLs) that state which

certificates have been revoked is an important part of the PKI operation. Resource administrators must acquire these CRLs and bar those entities listed in them from using their resources;

- A Grid PKI is one of the select PKIs where end-entities (scientists, grid services) are issued with their own personal certificates. This makes a widely available and pervasive PKI a prerequisite for any Grid operations. Given the value of the resources protected, the identity vetting rules and the quality of in-person authentication of these end-entities largely determined the granularity of authorities and delegated authorities (CAs and Registration Authorities). Nevertheless, the ensemble of CAs should appear as a coherent PKI issuing equivalent assertions in as far as Grid authentication is concerned.

Given these characteristics, a number of observations can be made:

The operation of certification structures by CAs involves social management as much as anything. A CA must be known and trusted by other CAs, so personal interaction is necessary. This does not scale well globally unless the number of CAs is limited. A single CA cannot alone handle grid certification for a whole country, so devolution of authority to Registration Authorities (RAs) is essential. The role of the CA is to root the chain of trust, establish policies and practices, and be guardian over the root security. The CA must select RAs, who become the real certifiers of trust, so that their geographical distribution reflects that of the users.

A hierarchical root for all the national CAs has proven less workable than a forum, a Policy Management Authority (PMA), where trust can be established and maintained. The GGF is debating a global PMA infrastructure. Europe leads by example in this arena: the DataGrid CA Group evolved to become a PMA in all but name, including members from Russia, USA, Canada and Taiwan.

Such a coordinating body is needed to coordinate end-entity identity providers and to foster trust relations for authentication purposes within the context of inter-organisational resource sharing within the European Scientific community and related communities. Such a body must establish the requirements and best practices for national grid identity providers to enable a common trust domain applicable to authentication of end-entities in inter-organisational distributed resource access. The PMA itself does not provide identity assertions, but instead recommends that – within its scope– the certificates issued by the participating Authorities are considered equivalent.

In the context of the DataGrid and CrossGrid projects, this body was called the "Certification Authority Coordination Group" or CACG. Within the Irish semester of the eIRG, this body has been transformed into a chartered body spanning several Grid projects in the 6th framework programme (EGEE, DEISA, SEE-GRID) as well as related projects such as the LHC Computing Grid (LCG). This new body has been named the EUGridPMA.

It is also understood that the requirements for grid authentication depend amongst other things on the financial and legal implications of a breach of trust. The scientific community, both identity providers, end users and relying parties, have together established a common set of minimum requirements in the context of the PMA. It is unlikely that this same set of requirements – and maybe not even the model of broad community trust – is applicable for a community where actions on a Grid could represent significant monetary value, or have severe societal implications (like e-commerce or e-government). The EUGridPMA therefore focuses only on a common trust domain for science.

Issues discussed in the Grid authentication PMA for e-Science

The PMA will debate grid authentication policies, and for example a healthy debate regarding the current reliance on PKI continues. In general, this body should host discussions between identity providers, subscribers and relying parties on acceptable policies and practices for grid authentication. For this body to be effective, and to establish a well-founded trust domain, it should consist of recognised technical experts in authentication, PKI and GSI. In particular, the identity providers (CAs) and the major multi-national grid projects should be represented in this body. Some current threads of discussion are singled out below as examples:

User key management and security

A crucial issue is the security of a user's private key. Whatever the mechanism, a user should not be able to compromise this by lax practices such as writing passwords in obvious places. Many CAs and relying parties feel that alternative mechanisms may secure the private key more effectively than PKI, although possibly rooted to a PKI CA, and granting only short-term credentials. The propagation of trust relationships from users to resources, and policies from PMAs to resources, is a debate waiting to happen. Issues of fine detail, such as automatic mechanisms for CRL propagation, also require debate.

Issuing and renewing certificates

Authentication on the Grid requires robust procedures for establishing and confirming identity. It is often impossible and frequently undesirable to require individual users to register at each and every Grid site. The Grid identity credential, today an X.509 certificate and its associated private key, therefore plays an important role in that this is used as the primary authentication of the user. In turn, the authorization of access to resources is granted by Virtual Organizations (VOs) and granted or denied by resource owners via the association of Authorization assertions to the Grid identity credential.

The model established in the EU DataGrid and related projects is that individual users obtain an identity credential from one of the approved CAs, which is then used to authenticate them with the Grid. The long-term aim is that this one credential can then be used as the identity basis for Authorization in multiple Grid projects and/or multiple VOs. The use of the one Grid identity across many sites and projects required the definition of policies and procedures of sufficient quality and robustness to be acceptable by all.

To achieve this, the DataGrid CA group involved active participation of CA managers from several other Grid projects across many different countries. Some of the EU CrossGrid CAs were founder members of the group early in 2001, with the US DOE Grids CA joining soon afterwards. Further expansion of the group, driven mainly by the identity requirements of the global LCG project, resulted in the approval of the remainder of the CrossGrid CAs together with Grid Canada and ASGC Taiwan, a current total of some 20 CAs. Many other national CAs have since joined and are working towards approval. The current list of new CAs includes Armenia, Belgium, Estonia, Hungary, Israel and Pakistan.

The DataGrid CA group defined, and the EUGridPMA maintains, minimum acceptable standards for the operation of member CAs and their associated RAs, and for the CA/RA policies and procedures. More details are available on the DataGrid WP6 and EUGridPMA websites, see: http://marianne.in2p3.fr/datagrid/ca/ and http://www.eugridpma.org/.

The experience gained building the DataGrid PKI and inter-Grid authentication with projects such as EU CrossGrid, US DoEGrids and Grid Canada was valuable input to the two GGF groups tackling Grid Certificate Policy and CA Operations issues and related work to establish worldwide

trust via multiple PMAs. Information is exchanged between the PMAs in conjunction with GGF meetings and via the GridPMA.org website. It is very important that the EUGridPMA continues this work in a global context thereby allowing cross-authentication between Grids across the world.

Revocation of certificates
The Certificate Revocation List (CRL) is an important component of the PKI. Each CA maintains a CRL placed at a published URL. This list, digitally signed by the CA to confirm integrity, contains the serial numbers of previously issued certificates which are now revoked and therefore no longer valid. Reasons for revocation, which are specified in the CA policy and procedure documents, include the loss or compromise of a private key or the fact that the entity is no longer entitled to hold the certificate. DataGrid and LCG established procedures to ensure that these CRL's are updated promptly and regularly and that all sites copy them frequently, and it is expected that the FP6 projects will enhance these.

A common repository of trust anchors
As stated earlier, an important issue in the operation of any PKI is the secure distribution of the CA certificates containing their public keys. In the Grid, these "roots of trust" are self-signed, so there can be no in-band digital signature confirming the veracity of the information. Alternative methods of distribution have to be used. The certificates of the major commercial CAs, for example, are distributed as part of the web browser software. In DataGrid, CrossGrid and LCG, the list of approved CAs and their public keys are stored on project operated web servers and distributed with the Grid middleware as part of the standard software distribution mechanisms. Detection of compromise of such a repository requires redundancy. Guaranteed detection of any compromise requires at least 100% redundancy (i.e. replication) – a common third-party repository (in addition to the project repositories) would minimize the complexity of managing this replication. Recently the EUGridPMA and TERENA have agreed the use of TERENA's TACAR repository as a common repository for storing and validating CA root certificates and policies.

The TERENA Academic CA Repository (TACAR)
The TERENA task force on Authentication and Authorisation Coordination for Europe (TF-AACE) has recently created a repository for storing the certificates and policy documents of NREN CAs. This is aimed at facilitating the use of PKI via easy access to secure information about participating CAs.

The idea of setting up an on-line repository hosting the NRENs trust anchors was proposed and discussed within the TF-AACE community and has immediately gained a lot of community support. Over the last months of 2003, the TF-AACE group decided to formalise the application procedure and a policy document was prepared and discussed on the TF-AACE mailing list. The original policy exploited the fact that TF-AACE community is a small community and therefore personal trust relationships were already in place. With a second revision of the policy, it was decided to extend the range of applicants to include National Academic PKIs in the TERENA member countries and non-profit research projects directly involving the academic community.

The first time an applying CA asks to join TACAR a face-to-face meeting between a TERENA representative and the CA is required in order to establish a personal trust relationship. Due to the fact that the certificates collected by TACAR can be used for several purposes, the policy does not define minimum requirements for applying CAs and TACAR do not evaluate their CP/CPS against these requirements, but only establish which CAs can join TACAR.

Collection of root CA certificates
Each applying CA must fill in a registration letter that provides information about the legal status of the CA and defines the representative of the CA who will hand out the root CA certificate. The TERENA representative will present a valid identification document as authentication, will check the identity of the CA representative against a valid identification document, and will collect the root CA data. Due to nature of this procedure, whenever anything changes in the CA status a new face-to-face meeting is required.

TERENA acts as third party guaranteeing that the applying CA belongs to an institution that is either a TERENA member NREN, or belongs either to a National Academic PKI in the TERENA member countries (NPKIs) or to a non-for-profit research project directly involving the academic community. TERENA does not have any control over the certificates revocation lists, so it is CA's responsibility to inform TERENA that something has changed.

Accreditation Procedure
The applying CA can also follow an accreditation procedure (optional, but recommended) that allows the applying party to use an accreditation letter to designate a number of individuals allowed to act as the contributors to the repository. A responsible person (who will be in charge of maintaining and updating the accreditation information as stated in the letter for the applicant) and at least two CA administrators (who will be in charge of registering, maintaining and updating the root certificate information on behalf of the applicant) must be designated as well. If the accreditation procedure is followed, then the PGP personal keys of the people entitled to represent the CA must be provided. The face-to-face meeting will allow the TERENA representative to 'register the keys' of people listed in the accreditation letter. The PGP keys can later be used to update the information of the CA via e-mail.

Limitation of this approach
The model describe above has of course some limitations. Because of the fact that the trust model is based on face-to-face meetings, this implies that the number of people involved must be limited. Moreover, the certificates collected and the related CA information are not supposed to change rapidly. The level of trustworthiness assigned to a particular CA depends on the way the CA policy is evaluated by the person relying on that particular CA. This means that a CA can be suitable for some purposes and less suitable for others. Today, the policies are provided in their original languages rather than in English, which makes widespread understanding difficult. In the near future a translation into English will be expected.

These limitations do not affect the EUGridPMA, since it is a small community with relatively static information content and a specific constituency. Hence it has been agreed by TERENA and EUGridPMA that the latter will adopt TACAR as a common repository of trust anchors.

Privacy preserving infrastructures
Many of the authentication infrastructures described above have not been primarily concerned with privacy preservation. Such considerations require the use of some alternative scheme, as averred to in Section 5.1. In recent work by the NREN community, an Authentication and Authorisation Infrastructure (AAI) has been proposed that is able to support seamless and location independent access to networking services has been proposed. This infrastructure will be based on the basic principles of federated administration and privacy preservation.

Federated administration allows for the decoupling of authentication and authorisation procedures, connecting them by means of trust links established within the federation. This allow for the coexistence of different local authentication and/or authorisation systems. Privacy preservation is

oriented towards avoiding unnecessary "data leakage" when performing AA interactions, providing users with the ultimate control over what information about them is exchanged for what transactions. Traditional authentication, using the Distinguished Names (DNs) embedded within certificates, may leak too much user data. On the other hand, it may also convey too little significant information about the user rights. Again, here is where the AAI will come into play.

5.2.3 Authorization

Authorization describes the process by which an authenticated user gets access to resources and is very different from authentication although the information about the user captured in the authentication process is often used in the authorization procedure. In the classical example of the authorization of an employee to use the company computer, the company director authorises the system manager to create an account with a password for the user and to tell the user about what she may and may not do while using the company computer. This simple example shows the two processes which make the authorization difficult for more complicated examples like grids: the delegation in the authorization process and the existence of an acceptable use policy. In current implementations of grids a user necessarily is a member of a Virtual Organisation and can be authorised to use grid resources. However the delegation chain in this may be very long as the relationship between that user and the owner of the resources may be a long chain of agreements. Moreover the rules of the game described in the acceptable use policy of the resource owner may not be accessible easily for the user. It is therefore important to try to define on the same international scale as resources become available on grids how to deal with delegation of rights and policies for proper use.

For the centralized super-computing case authorisation is much simpler based on computer accounts permissions. For the distributed supercomputing case (i.e. a grid of supercomputers), each partner will provide part of its supercomputer resources for use by the partner users. This case probably will fall into the same category as far as authorisation is concerned, provided similar grid middleware is used to enable the resource sharing.

Concepts

First let us explain some basic concepts that are involved in authorization. Authorization is a general concept which is not only used for grids, although further below some example implementation detail for grids will be presented.

One must realize that the term authorization may mean one of following:

- the process of issuing a proof of right;
- the proof of right (or reference to it) itself (i.e. an authorization token);
- the process of making an authorization decision by checking a proof of right, e.g. by rendering user attributes against access control policies.

An authorization decision can be made at a number of places:

- At the entrance of a service point (authorization means access control in this case);
- At a (central) point outside the service point.

To avoid this confusion one should always make a reference to the context.

Basic entities

In principle authorization decisions are made based on authorization information provided by authorities. These authorities must have a direct or a delegated relationship with either the authorization subject (e.g. a user or organization member to which the authorization is issued), or

with the resource that is the target of the request that prompted the authorization (e.g. owner or administrator of a resource), or with both. These relationships may be implemented using a trust mechanism based on some cryptographic method or may be implemented completely off-line (i.e. by some other trusted delivery mechanism).

This observation leads to the definition of the three basic high-level entities involved in authorization. This terminology will be refined in this paragraph for the specific use within this White Paper.

Subject: An entity (e.g. a user or process) that can request, receive, own, transfer, present or delegate an electronic authorization so as to exercise a certain right. Informally, a subject is any user of a service or resource. The subject may be identified as an individual user or as a member of a group of users. A Subject may also be a process that acts on behalf of a user and as such holds access rights that were delegated to it from the user. The subject may define a set of policies that determine how its authorization is used.

Resource: A component of the system that provides or hosts services and may enforce access to these services based on a set of rules and policies defined by entities that are authoritative for the particular resource. Typical resources in Grid environments might be a computer providing compute cycles or data storage through a set of services it offers. Access to resources may be enforced by a Resource itself or by some entity (a policy enforcement point, gateway) that is located between a resource and the requestor and thereby protecting the resource from being accessed in an unauthorized fashion.

Authority: An administrative entity that is capable of and authoritative for issuing, validating and revoking an electronic means of proof such that the named subject (a.k.a. holder) of the issued electronic means is authorized to exercise a certain right or assert a certain attribute. Right(s) may be implicitly or explicitly present in the electronic proof. A set of policies may determine how authorizations are issued, verified, etc. based on the contractual relationships the Authority has established.

Authorization is frequently split into three distinct processes:

- defining an authorization policy at a high-level by a person or organization;
- implementing the high level policy into a certain executable form;
- evaluating the executable policy by a process which subsequently decides to issue a specific authorization to a subject or take a specific action.

Each of these three entities may implement a **set of policies** that determine the handling of an authorization. The policy handling function may be implemented as a hard coded piece of logic or it may be implemented by means of a flexible policy language. The component performing the evaluation of the executable policy by computing an authorization decision on behalf of the authorities is sometimes referred to as an **Authorization Server**.

All concepts and definitions above within the area of authorization have been generic and not only valid for grids. They are equally applicable to fields like network authorization as well as mobile communication. In the following subsections there will be more focus on authorization within the context of grids and on the specific items for this white paper.

Domain Considerations

An administrative domain is a definition of the scope of authority. In many distributed authorization scenarios there are at least two administrative domains: that of the user (Subject) and

that of the Resource. In Grid environments we frequently see scenarios where there are separate domains for identity, subject attribute, resource policy, and community policy authorities. In a simple Grid use case the Subject is in one administrative domain, its home domain, and the Resource is in another (the home domain of the resource). In more advanced scenarios a community or Virtual Organization (VO) domain is present. A VO domain can provide Authorities that perform privilege management for all the members of a VO. A typical Grid scenario is one where a user needs to use services from several domains. Sometimes this is accomplished by a Resource in one domain using a Resource in another domain on behalf of the user. Grid Service Providers may provide resources to users in multiple VO or home domains.

A Virtual Organization is a dynamic collection of distributed resources that are shared by a dynamic collection of subjects (users) from one or more physical organizations. Many of today's VOs are formed to tackle large-scale scientific problems. Large computing centers provide the resources and domain scientists are selected as users. The emerging approach in grid computing is essentially to define a VO as a particular set of users whereby the equivalent of a VO server issues tokens to humans attesting to their membership in the VO. These tokens are then presented to the individual resources. However, as VOs grow in scale, their creators will need to define their VOs in more complex and comprehensive ways than via low-level membership descriptors alone. There are many types of policies in Virtual Organizations of which one of the most common ones, Authentication, is described elsewhere in this White Paper. For the next most commonly policy, Authorization, the concepts and definitions have been described above. In the restricted scope of a supercomputer or computer center (the Resource) authorization policies have been defined and applied. For grids many initiatives have been taken to generalize the authorization model to be able to cope with the distributed and dynamic nature of the resources and the subjects, the users. Today's large scientific collaborations are almost all international and the Authority which collectively owns the resources and defines the use policies has to become a supra-national institution. To define this supra-national Authority on a European scale is the charter of the eIRG and the goal of this White Paper.

5.2.4 Accounting

Usage of grid resources is of interest to many parties. Consumption of resources is very important to the administrators of the resources. It is potentially an item chargeable to research funds and can serve as an instrument of policy for institutions, funding agencies and governments. Up to now, no cost charges have been applied at any Grid resources (which is different from the super-computing cases), since the accounting background is a prerequisite for billing. Accounting policies are also in their early stages. Accounting is of key importance for the integration of super-computing centres in the Grid. No super-computing centre or large cluster can just give resources to the Grid community without taking care of revenue for their users. As long as the demand does not saturate the resources (typical in research test-beds) the problem does not come to the surface; however when there will be lack of resources, accounting and billing will become an important issue. Accounting is also important for statistical purposes, providing history parameters in grid resource markets.

In the Rome eIRG meeting it was decided that accounting for the grid is technologically immature and related policies should be deferred to future presidencies. Thus, only a short presentation of an example prototype is given here. Such a prototype is DGAS, which was developed in the DataGRID project. The DGAS mechanism is based on economic transactions between users and resources. Both users and resources are seen as entities capable of exchanging Grid Credits. The most usual case for this exchange is that the user receives the amount of Grid Credits from the VO

it belongs to, and then the user will pay the resource a well defined amount of Grid Credits in order to get their job executed. In a grid framework where the VOs own part of the resources, these resources will earn credits by executing user jobs. These credits can then be redistributed among the users belonging to the VO controlling these resources. In other words this accounting mechanism allows for a VO's users to utilize as much of the grid resources as they wish provided they "repay" by providing access for other members to resources they control. The geographically distributed service can be seen as analogous to bank-branches where the accounts for both users and resources reside.

The advantages of this approach are:

- It is user friendly, since everybody is familiar with basic economic concepts. In fact while it can be difficult to understand at a glance the usage consumption of a job analysing info such as CPU, resident memory and disk storage utilizations and similar technical details, it's easy to understand if the effort required by a job was small or big in terms of the amount of "money" required to get that job executed. This approach still permits the user or the system administrator to know the details of the computations;
- It's easy to avoid indiscriminate usage of the grid by single users. Since every user has a well defined amount of credits available, they won't submit more jobs than they can afford, thus minimizing the risk of individual users saturating the grid with job-trials (or worse with deliberate abuse of grid-resources);
- It motivates the organizations involved in the Grid to share their computing resources. This is because the more resources you have, the more credits you earn;
- It makes it feasible to involve third-parties such as big computing centers or industrial partners, since it is easy to exchange computing power by means of credits;
- It can be used to help the Workload Management process by means of natural economic feedbacks. This aspect is also known as Economic Brokering. The idea is that an economic accounting environment naturally creates an "exchange market" where the users wants to maximize their computing capabilities while minimizing their expenses; on the other hand their counterparts, the resource owners, want to maximize their earns while minimizing their expenses. So the resource owners (or a suitable automatic system) will manage the resource prices in order to maximize the utilization of their resources (usually lowering the prices for idle resources), while the user will usually seek for the cheapest resources. This is expected to generate a feedback that, if well tuned, should result in an equilibrium state where there are few idle resources and generally increased throughput.

This economic accounting system approach can facilitate the policy choices for resource sharing. However, it is an example prototype, and much more experience is needed.

5.2.5 Resource Sharing

Sharing policy is in its infancy, with no clearly expounded existing practices. However, some aspects of future practice can be stated. The principal actors are VOs and the different layers of grid administration domains down to the single resource administrator, supported by the policies of their associated funding agencies and governments. The challenge of the sharing policy administration and enforcing systems is to provide flexible, efficient, secure tools to negotiate, propagate and enforce policies agreed between VO administrators and grid administrators. Different VOs have the responsibility of contacting different grid administrators and agreeing the details on how to share the resources of the grids. The agreements on paper should be translated into formal language and implementations on the grid.

The definition of a sharing policy architecture assumes a grid administration model where the actors have specific roles, as discussed in Section 4.3. For example, Figure 5-2 shows a possible VO-driven case:

- VO administrators contact different Grids administrators in order to negotiate the CPU and storage capacity they need;
- A Grid administrator negotiates with the resource administrators the availability of CPU and storage capacity for the Grid, and then negotiates with the VO administrators the grid CPU and storage allocation;
- Resource administrators take the final decision, but they might delegate the grid administrators to interact with VOs and negotiate policies.

This is a model that establishes a hierarchy that inserts the Grid Administrator in the middle, the real new character in the transition from "computer sharing" to "grid sharing" and from "multiple users" to "multiple VOs".

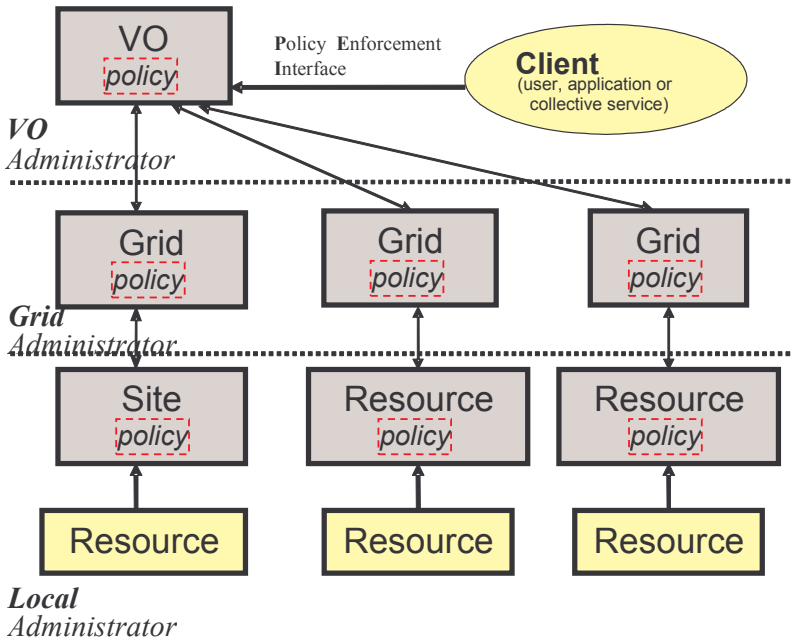


figure 5-2 VO-driven grid administration model showing policy levels and their interactions

VO-driven policies will reflect the way the VOs plan to use different grids or specific grid resources, where new types of policies must be defined in order to optimize the service delivery by the Grid to the VO. Policies driven in other ways would have different, but related, needs. A common policy language between VOs and Grids is a precondition for any interoperable policy system.

6 Policy Framework for Resource Access and Sharing

6.1 Introduction

According to the White Paper methodology a policy framework for resource access and sharing at pan-European and international level, plus a list of policies, will be drafted after capturing and analysing the multidisciplinary entities' requirements, taking into account the current experience and best practices of the major Grid projects and Super-computing centres. In addition, a roadmap will be defined for the future development of a political and administrative framework in Europe and internationally to allow a real effective exploitation of the eInfrastructures.

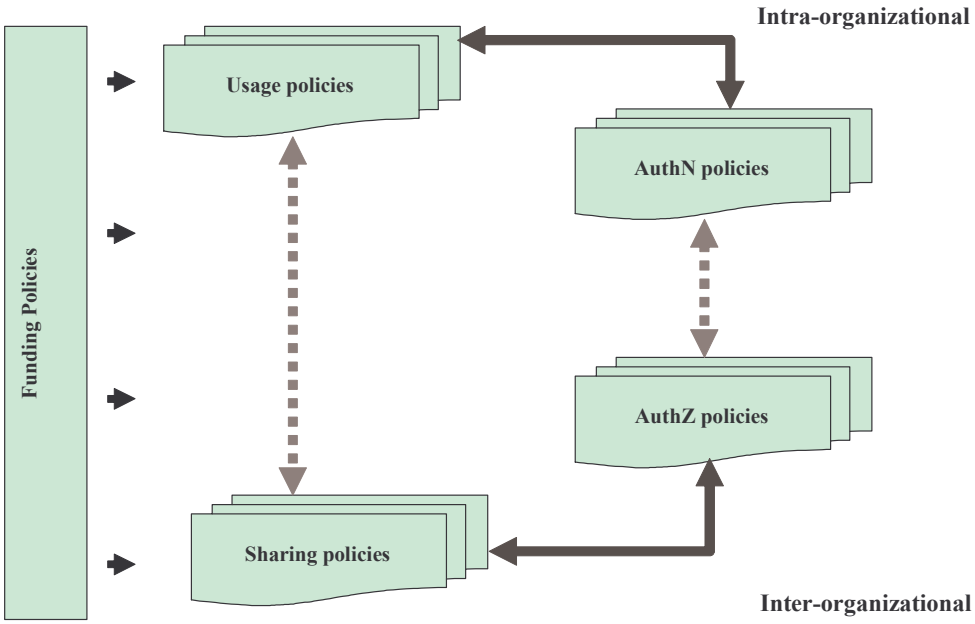


figure 0-1 – Policies interrelations

The above figure indicates those policies more related to the intra-organisational realm (and the non-collective middleware), and those more related to the external inter-organisational realm (and the collective middleware). There are direct relationships between policies applied in each corresponding realm, as well as interactions between related areas (usage and sharing, authentication and authorisation) across the different realms. Obviously, funding policies influence all other aspects.

6.2 Authentication Policies

To restate from Section 5.2.2, the requirements for authentication depend amongst other things on the financial and legal implications of a breach of trust. On the other hand it is essential for researchers in Europe that they can work across a wide variety of national and European projects bearing only a *single identity*, so as to enable a common European research area. Consequently, for operation and deployment within a single European e-infrastructure, it is beneficial that there is maximal communality amongst the relying parties in their acceptance of authenticating authorities

for access to resources. These things are independent of a user's discipline or the particular form of the e-infrastructure.

Let us take the concrete example of authentication for resource access and sharing in a scientific context using Grids. We may conclude that:

- International (European) Grid projects, and those national Grid projects that encompass organisations that also join in European projects, should be able to use identity providers that are part of a common trust domain for grid authentication in Europe;
- This domain should be distinctly for e-science, separate from other resource access and sharing efforts.

Corollaries are:

- Authorities that provide inter-organisational identity assertions should register with a common European trust anchor repository;
- There should not be a needless proliferation of authorities that provide authentication services for the scientific community.

The EUGridPMA and TERENA TACAR repository provide a sound solution to these requirements for the domain of e-science. The repositories and trust domains that have already been established are a solid foundation for the common repository for trust anchors in the academic community and for the common trust domain for grid authentication in e-science. Each is proposed by and will be maintained by the relevant bodies. Both have a wide acceptance in the community and encompass the key technical experts in their respective domains. It is suggested that an endorsement by the eIRG of the EUGridPMA and the TERENA TACAR repository would be a concrete first step towards common EU authentication policies for authentication for resource access and sharing for e-science.

6.3 Authorisation Policies

As mentioned in paragraph 5.2.3 there are two main issues involved in Authorisation: acceptable use policies and delegation.

Acceptable use policies are in general defined by the resource owner and are often influenced by local rules or national legislation. As described in paragraph 6.6 the GEANT project has developed a scheme which works for access to networks without defining an overall AUP by incorporating the policies set by its members, the national research network organisations. In the European DataGrid project a common AUP was defined which was acceptable to all its partners and it is interesting to see how the problem will be solved in the much larger LHC grid project from particle physics, where the security group has made a proposal. It is of utmost importance that the eIRG follows these developments and stimulates a development towards an AUP which is not only acceptable for all partners on a European scale, but also covers all resources like CPU, storage and networking.

For delegation of rights it is important to define a framework which can describe all forms of delegation and which is independent of the technology which is used. At a national level this typically goes from the Ministry via the national science funding agencies through to the computer centres and individual users or groups. The framework ought to describe each national scheme and should not strive to make those schemes the same. However the scheme must support authorisation across national boundaries and lead to a sharing of (CPU, storage and networking) resources in Europe.

Several technologies for authorisation are in use at present but none of them have reached a high level of maturity and further development is needed on each of them. An agent solution is used in VOMS, which has been developed in the European DataGrid and DataTag projects and is now applied for the first time in LCG. Shibboleth, which is a project to share web resources, implements authorisation in a pull model, while the DataGrid GridMapFiles are an implementation of a push model. Like the acceptable use policies, developments should be stimulated which make these technologies interoperable rather than striving for one solution. Moreover it would be ideal if the same scheme could describe all resources, CPU, storage and networking.

New ideas in the field of Authorisation should be welcomed and stimulated, as VOMS, Shibolleth and GridMapFiles all have their shortcomings. New approaches and ideas, such as authorisation federation, have been described in papers and project proposals but have not succeeded in any implementation so far.

As there are sometimes significant differences between laws and use policies in the various European countries, the eIRG should support a group of people to produce a set of use policy documents covering all issues of grid resource usage, politically as well as technically. As these use policies have to be acceptable to and followed by all partners providing resources, this group faces much the same challenge as the EUGridPMA. A Common Use Policies group should therefore be organised very similar to the EUGridPMA, and could possibly be an extension of it. The outcome of the Common Use Policies group should be presented to the eIRG and be proposed as a European policy in a later version of this White Paper.

6.4 Funding Policies

In this section two characteristic methodologies and policies of funding allocation are presented, the first being related to the US NSF policies towards the materialization of the TeraGrid project, and the other related to European funding practises, as exemplified by a Hungarian government funding policy to support a national desktop Grid.

6.4.1 The US experience – The creation of the NSF TeraGrid

Historically there have been two major programs at the National Science Foundation to supply infrastructure (computing and networking) to America's academic community. With the emergence of a national community which deals with networking (Internet2), the NSF is no longer the "supplier" of this networking capability. However, in the area of high-end computing, there is a role that needs to be fulfilled for non-mission-oriented research in academia. This role is fulfilled by a division within NSF that is currently called Shared Cyberinfrastructure. This division provides the funding for leading edge computing systems (and other necessary infrastructure, see below).

NSF, in the Spring and Summer of 1998, held a series of workshops to survey the state of the art on applications, software, and tools for high-end computing, and make recommendations concerning support for these areas. Coincident with this, the White House convened a panel, the President's Information Technology Advisory Committee (PITAC), to investigate much broader issues in information technology. In its preliminary report in the Autumn of 1998, and its final report of February 1999, the PITAC report recommended that NSF acquire "the most powerful high-end computing systems to support science and engineering research". All these recommendations were taken under advisement by NSF, which proposed a three year Terascale Computing Initiative to Congress. The first solicitation requested a large centralized system (the 6 TF HP/Compaq machine was the winner of that competition), while the second competition focused on the new computing paradigm of distributed systems, and resulted in the Teragrid project, which consisted of four sites

with an aggregate capability of greater than 13 TF all connected by a 40 gigabit backplane network dedicated to connecting all the sites.

In 2002, NSF was also considering the impact of what has come to be termed cyberinfrastructure in the United States. The TeraGrid, as the antecedent of a national cyberinfrastructure, was the base upon which to build. There already existed within America many facilities that would add benefit to the existing funded partners in the TeraGrid. In 2003, as the pieces of the TeraGrid were being developed and assembled, the NSF issued another call for proposals for an enlarged TeraGrid using the concept of an Extensible Terascale Facility to augment the TeraGrid. Thus, the foundation of a truly national cyberinfrastructure was laid.

So, funding policies used in the US can be summarized as follows:

- For high-end computing infrastructure necessary for non-mission-oriented research areas in the US, high-level advisory panels and committees are essential for providing recommendations for developing such infrastructure. Technology evolution needs to be constantly taken into account, e.g. this was the case when moving towards a distributed and grid computing model. The above structures are fundamental to enable political support and funding;
- A series of workshops on infrastructures, applications, software and tools are essential for surveying the state of the art and making recommendations concerning related funding support for these non-mission-oriented research areas, as is the case for high-end computing in the US;

In this way, as identified in the US, both a top-down and a bottom-up approach are used to develop the appropriate funding policies in order to generate support for the creation of the cyberinfrastructure

6.4.2 European paradigms

Funding policies should be based on a *vision*. This vision should be transformed to a detailed *execution plan*. A *professional organization* is needed to elaborate the elements of this plan. In order to realize the plan a requirements capture and architecture specification cycle is needed followed by the related *actions* implementing the plan.

For the development of electronic infrastructures in Europe the *top-down* European-wide vision should influence the bottom-up national plans and vice versa. This way the national plans are expected to be integrated into the appropriately adjusted European-wide plans. The professional organizations in each country are usually supported and advised by strategic committees consisting of national delegates and representatives of the European Commission.

Funding in Europe is provided from three main sources:

- The European Community (through the Commission of the European Communities);
- National governments;
- The private sector.

For the research eInfrastructure realization in Europe, the infrastructural part (vertical) has been based on national government funding and/or European structural funds supporting the national governments. The pan-European integration is supported through European Union Framework Programmes projects making the EU eInfrastructure initiative a reality. Such projects are the GEANT (GN1 and GN2), EGEE and DEISA (along with other sister projects), providing pan-

European research networking connectivity, grid middleware operation and support and distributed supercomputing cycles. The above projects are complemented with the corresponding national (National Research and Education Networks and National Grid Initiatives) or private sector efforts. EU financing requires national or private sector co-financing.

An interesting national example of this paradigm is the funding policy that gave rise to the Hungarian ClusterGrid. The ClusterGrid was initiated by the Ministry of Education of Hungary as a tender for PC laboratories in higher-educational institutions. Within the project 100 new PC laboratories were established, each containing 20 high-end PCs and a server plus a firewall. These laboratories are connected via a virtual private network on the Hungarian Academic and Research Network.

During the day the laboratories are used for educational purposes, while during nights and weekends the PCs "switch" to the Grid mode and are part of a desktop Grid. The "switch" to the desktop Grid is obligatory for the educational institutes, as foreseen by the contracts, and as such is an example of encouragement in return for sharing responsibilities. This homogeneous Grid has three organizational levels, i.e. entry points, local servers (in each laboratory), and client computers. The total investment was 4 M Euro, a modest sum in comparison to the price / performance ratio of this dual-function system.

6.5 Sharing Policies

Sharing policies are the policies that enable the sharing of resources across resource centres that span multiple administrative domains, as well as inside and among different virtual organisations. For fair sharing of grid resources Europe-wide, a framework is needed where political and if necessary also technical issues can be addressed. Ideally the end user should not be confronted with money at any stage of doing her research on the European grid infrastructure. Yet in the participating countries the resources have to be procured and purchased. The framework should therefore include the people or institutions from the participating countries which own the resources. Usage of those resources on a European scale has to be made possible on a fair share basis. Policies have to be developed to define this fair share and techniques have to be put in place to instrument those policies following the guidelines and standards described in 5.2.3. Regular meetings will have to be organised where all resource owners in Europe adjust the policies to the latest developments and discuss the balance of the sharing. The accounting should be the last step at this level; payment should not be an issue as this would enormously complicate things at this international level.

For the end user of the grid this scale up of accounting should be transparent. In general a researcher is used to submitting a request for resources needed for the research to the corresponding funding agency. The proposal is processed through some selection procedure and the researcher is assigned a budget which can be used to do the research, i.e. in the simplest case to submit jobs. The above framework should elevate this procedure to European scale without increasing the complexity of the procedure for the end user. The researcher should, however, be able to use a much increased pool of resources on the pan-European grid with her assigned budget. Any accounting between resource owners must happen beyond her horizon.

For the distributed *super-computing* case, each partner will provide part of its supercomputer resources for use by the partner users. There is a distinction between resources for internal users, who are users that already use resources of one of the individual sites, and for new external users. Not every partner necessarily makes the same commitment. For example, if only the core partners make the same commitment, and the commitment is a percentage of their own resources, then the

absolute amount of resources available for a project is different for each site. As a sharing policy this will fall into *Policy 2* below with the non-equal variation. This means that accounting information of each site relevant to infrastructure users must be available to the management of the infrastructure, and also the total usage of the site must be available.

In current projects, such as DEISA, there is as yet no clearly defined policy for how resources are assigned to users. So currently only the total usage of DEISA users at each site will be relevant, however it is to be expected that in the future some insight into usage per user will be necessary in order to be able to assign the DEISA resources in a fair way to its users (more details of resource management of the DEISA pool of resources will be presented at the Dublin eInfrastructures Workshop).

Since limited information on actual (running) sharing policies has been reported in the different Grid projects, the White Paper editors have tried to imprint the state of the art, referring to recent publications in conferences and journals. Practices and approaches adopted in other fields where policy-based sharing of resources is done (e.g. networking resources) could be also extended to cover the Grid environment.

A related paper that has been presented in Grid 2003 in Phoenix by Glenn Wasson and Marty Humphrey, University of Virginia, is quoted below, analysing the obligations of Physical Organizations (POs) to the VO.[5] The policy definitions are quoted as they appear in the paper:

Policy 1: *Each PO member opportunistically gives what it can to the VO* [***the you-give-what-you-can (ygwyc) policy***]
We believe that this is the dominant policy implicit in many of today's scientific VOs. But we also note that this is probably not the desired policy, but rather the only policy that is easily implemented (primarily because there is no required enforcement for this policy).
As mentioned in the paper, this is in fact a best-effort policy and thus cannot be considered as an explicit policy.

Policy 2: *Resource utilization is divided equally among member resources* [***the 1/N policy***]
We refer to this policy as the "1/N policy" because each resource in the VO is to perform 1/Nth of the total work of the VO (non-equal variations of this theme exist as well). This policy can apply to any resource that is distributed throughout the VO's member organizations: cycles, disk space, or other specialized resources. A 1/N policy is a common implicit desire in VOs where a PO's users are allowed to join a VO because it is assumed that the PO's resources will "pull their own weight". Typically, this policy is neither explicitly stated nor enforced.

Policy 3: *Each PO member receives VO utilization credit for the resource utilization their PO provides to other VO users outside the PO* ***[the you-get-what-you-give (ygwyg) policy]***
Instead of requiring an equal distribution of resource utilization throughout the VO, this policy allows for users to utilize as much of the VO's resources as they wish provided they "repay" the VO by providing access for other members to resources they control. We contend that, arguably, this policy and policy 2 (1/N) are the desired policies in many emerging VOs.

In a second paper to be presented at the next TERENA Networking Conference[6] the authors propose a model in which the users can be roughly divided in those that *only* use resources and those that *also* share resources on the Grid. The mechanism that establishes a simple relation between those users relies on the Internet Service Provider as a common entity between them. This

[5] http://www.cs.virginia.edu/~humphrey/papers/Grid03.pdf
[6] http://www.terena.nl/conferences/tnc2004/programme/presentations/show.php?pres_id=67

helps to simplify authentication, authorization and accounting issues especially in situations where there are users and users/providers that don't have or don't want to have any relation with others; it also this helps in enforcing Service Level Agreements.

Although the architecture that is proposed in the paper addresses the needs of commercial users (basically confidentiality and anonymity) in a particular service, it could be adapted to environments where there are entities that group several users (for example NRENs) that could later balance accounts between them or through a specific clearing house. The existence of a third party can ease the verification and enforcement of SLAs that should appear once the experimental or best-effort phase finishes and when commercial offers begin to appear.

6.6 Usage Policies

Given the infancy of Grid deployment within production environments there is not enough experience on the legal and regulatory aspects of operating an eInfrastructure. At this first attempt input has been received from the related experience of the research networking community (NREN) Acceptable Usage Policies (AUPs) and actions-sanctions in case of violations, as well as for corresponding work of the LCG Security Group on *Security and Availability Policies for LCG*. Other projects or national initiatives are encouraged to participate.

In http://archive.dante.net/geant/connect.html there is a list of the AUPs of those European NRENs connected to the pan-European GEANT backbone. Although there is a variety of AUPs covering different issues, a common basis for all AUPs can be identified. The main areas that the AUPs cover are the *eligible user communities*, their *rights and liabilities* including admissible usage and possibly some exceptions or extreme cases (disasters).

"Eligible user communities" for research networking infrastructures (i.e. communities that are allowed to connect and use the research network) are those aiming at education and/or research, rather than commercial profit and the greater "market". Although there is a variation of eligible users in the different NRENs, the user communities can be categorised in three rough categories:

- Education and research institutes (Academic and Research institutes and schools)
- Supporting public organisations (research ministries, libraries and sometimes hospitals or cultural organisations)
- Other institutes not belonging to above categories but which use the network for research or education purposes (e.g. companies or other organisations), and usually have temporal connections.

Under "**rights and liabilities"** a long inventory of articles usually appears, stating the admissible and inadmissible (prohibited) use of the network. In case of **violation** of the above-mentioned articles and prohibited use by the connected users, the NREN has the right, with or without a proper formal warning, to perform the following actions or sanctions:

- to delete the relevant offending data (e.g. pirate copies, messages with unlawful content) with or without prior notification
- and/or to suspend a specific network service or port that causes the violation
- and/or to suspend access by the user (if feasible) or the connection of the user's organization or company, with or without prior notification to the user or his organization or company and without the user or the user's organization or company accruing any claims to damages as a result. The connection of the offending network will be restored when the latter conforms to the rules.

However, this Acceptable Use Policy approach, as implemented and operational at the NREN access level, is biased by the fact that it is funding-driven, and rather defensive with regards to the telecommunication commercial sector. As the NREN infrastructure and services are addressing a rather large user community, the use of these networks is required to be properly controlled to avoid the possibility of telecommunications operators claiming unfair competition from entities that use taxpayer's funds. Hopefully the recent years have demonstrated that the services provided by the NRENs to their users are increasingly different from the services of commercial operators. Nonetheless, it is fully legitimate to keep these infrastructures in their support for a dedicated community.

One must also consider the effect of large scale project demands. The NRENs are designed to serve all research and education communities. Even if the infrastructures are designed to fulfil all possible requirements, the current policy rules to allocate specific resources are not in place to answer big requests, like those currently coming from HEP, Biology, Astronomy or even GMES communities. There is a real need to address, at a global strategy level, the issues of large scale demands, since they may absorb a very significant fraction of the network infrastructure resources and may eventually be subject to other scientific constraints and the standard funding problems.

The LCG Security Policy specifies the following policy compliance and sanctions in case of policy violations:

Policy Compliance
LCG proposes that Resource Centres conduct a self-audit of their compliance with their Policy following a procedure dictated by the appropriate central GOC. Self-audits or GOC independent on-site audits will be required for the continued recognition of the service being operated.
Legislation Compliance
Since not all countries have uniform or consistent legislation, LCG proposes to apply policies uniformly across all sites without violating local legislation wherever possible. If this is not possible, country-specific exceptions or extensions will be made to this policy and its associated practices and procedures described explicitly in an Annex.
Exceptions
In exceptional circumstances LCG accepts that the emergency actions taken may violate their policies provided it is for the greater good of pursuing or preserving legitimate LCG objectives. Still the exception should be minimised, documented, time-limited and authorised at the highest level commensurate with taking the emergency action promptly, and the details notified to the GOC at the earliest opportunity.
Sanctions
According to the LCG policy document, resource providers and their operators or administrators who fail to comply with the established policies, or its associated procedures and practices, may lose the right to have that service instance recognised by the project until compliance has been satisfactorily demonstrated again. The test of compliance will be an independent Audit. The same applies to the different entities such as users, administrators, developers or VOs.

Development of such policies requires extensive consideration.

6.7 Policy Framework Roadmap

In this section basic bullet points constitute the policy framework roadmap with immediate and future propositions:

- Promote interoperable authentication and authorisation infrastructures enabling seamless sharing of eInfrastructure resources, from network access to Grid interactions.

- The EU Grid Policy Management Authority (EUGridPMA) www.eugridpma.org, as a group of mutually trusted Certification Authorities (CAs), is instrumental for the security infrastructure of current GRID projects in the global arena. An endorsement by the eIRG of the EUGridPMA will be a concrete first step towards common EU policies for authentication for resource access and sharing for e-science.

- EGEE, DEISA and SEEGRID intend to sustain and use the Public Key Infrastructure coordinated by the EUGridPMA group. EGEE, DEISA and SEEGRID are also committed to be a proactive relying partners of the EUGridPMA group.

- A single common repository for authentication (and especially Certification Authorities) in the European Research Area would assist promoting further the trust anchor among the different research communities.

- The TERENA Academic Certificate Authority Repository (TACAR), see www.terena.nl/tech/task-forces/tf-aace/, will serve as the common repository for storing and validating the CA root certificates for the EUGridPMA constituent Certificate Authorities. An endorsement by the eIRG of the TACAR will be a further concrete step towards common EU policies for authentication for resource access and sharing for e-science.

- Enable the use of federated solutions, decoupling local authentication procedures at a user's origin organization from local authorisation at the target resource, where origins and targets are connected by the trust links built by the federation.

- Apply techniques for privacy preservation, oriented towards avoiding unnecessary data leakage when performing AA interactions, providing users with the ultimate control over what information about them is exchanged for what transactions.

- Identification, harmonisation (by means of relevant eIRG endorsement actions and recommendations) and further elaboration of policies on Authentication, Authorisation, VO resource provisioning and Accounting and enforcement of VO resource provisioning is essential.

- The above activities have to be supplemented and supported by the creation and continuous update of a registry of European Information and Communication Technology (ICT) resources and policies for e-Science and beyond, and the promotion of international collaboration on resource sharing policy aspects.

Appendix A - Security taxonomy

In order to do an optimal match between the cost effectiveness and security requirements for a specific application, one would ideally like some kind of taxonomy for the different security levels provided by the eInfrastructure. One could base this taxonomy on the work performed in the Carrier Grade Linux initiative or the OSDL[7] by modifying their 4-level Linux security classification to be applicable to the Grid environment. Thus one could propose the following taxonomy:

- **Default:** No additional security methods added to typical Grid implementation. Users not on the CRL list and having a certificate from a trusted CA will be mapped to a pool account and can execute jobs. The network traffic is not encrypted and communication with storage elements may expose data to entities with access to the network;
- **Perimeter defense:** Computation and storage components are contained within a site protected by a firewall preventing outsiders from monitoring activities or data after the job submission. Operations are logged by the local administration and can be audited, but no guarantees against attacks from inside the network will be provided;
- **Secured Communication, System Integrity Protection:** In addition to Perimeter defense, software version tracking (signatures) and methods for logging of activities in permanent manner (i.e. even an attack from the inside will leave a trace, due to logging to a remote site or medium that cannot be erased). Adding a certificate to the CRL list is an atomic operation accross all the sites in the VO;
- **Intrusion Mitigation, Insider Controls:** Procedures for shutting down a site and locking sensitive data immediately in case of intrusion are in place. In addition to always being able to know if the system has been compromised, the amount of data or algorithms exposed will be minimized.

Level 4 would be appropriate for medical data containing information that identifies individual patients, level 3 would probably be enough for anonymous medical data, and for HEP applications level 2 would quite likely be enough.

[7] http://www.osdl.org/lab_activities/carrier_grade_linux/documents.html