e-IRG White Paper 2011

Table of Contents

Foreword 6

Introduction and summary 7

1 e-Infrastructure governance: From management and international aspects, to legal and financial issues 8
- 1.1 Policy area and goal 8
- 1.2 Context: Current practices - achievements and limitations 8
- 1.3 Proposed approach 10
- 1.4 Recommendations. 12

2 Future of research networking 13
- 2.1 Policy area and goal 13
- 2.2 Context: Current practices - achievements and limitations 13
- 2.3 Proposed approach 14
- 2.4 Recommendations. 15

3 Authentication, authorisation and accounting 16
- 3.1 Policy area and goal 16
- 3.2 Context: Current practices - achievements and limitations 17
- 3.3 Proposed approach 17
- 3.4 Recommendations. 18

4 Energy and Green IT 19
- 4.1 Policy area and goal 19
- 4.2 Context: Current practices - achievements and limitations 20
- 4.3 Proposed approach 21
- 4.4 Recommendations. 22

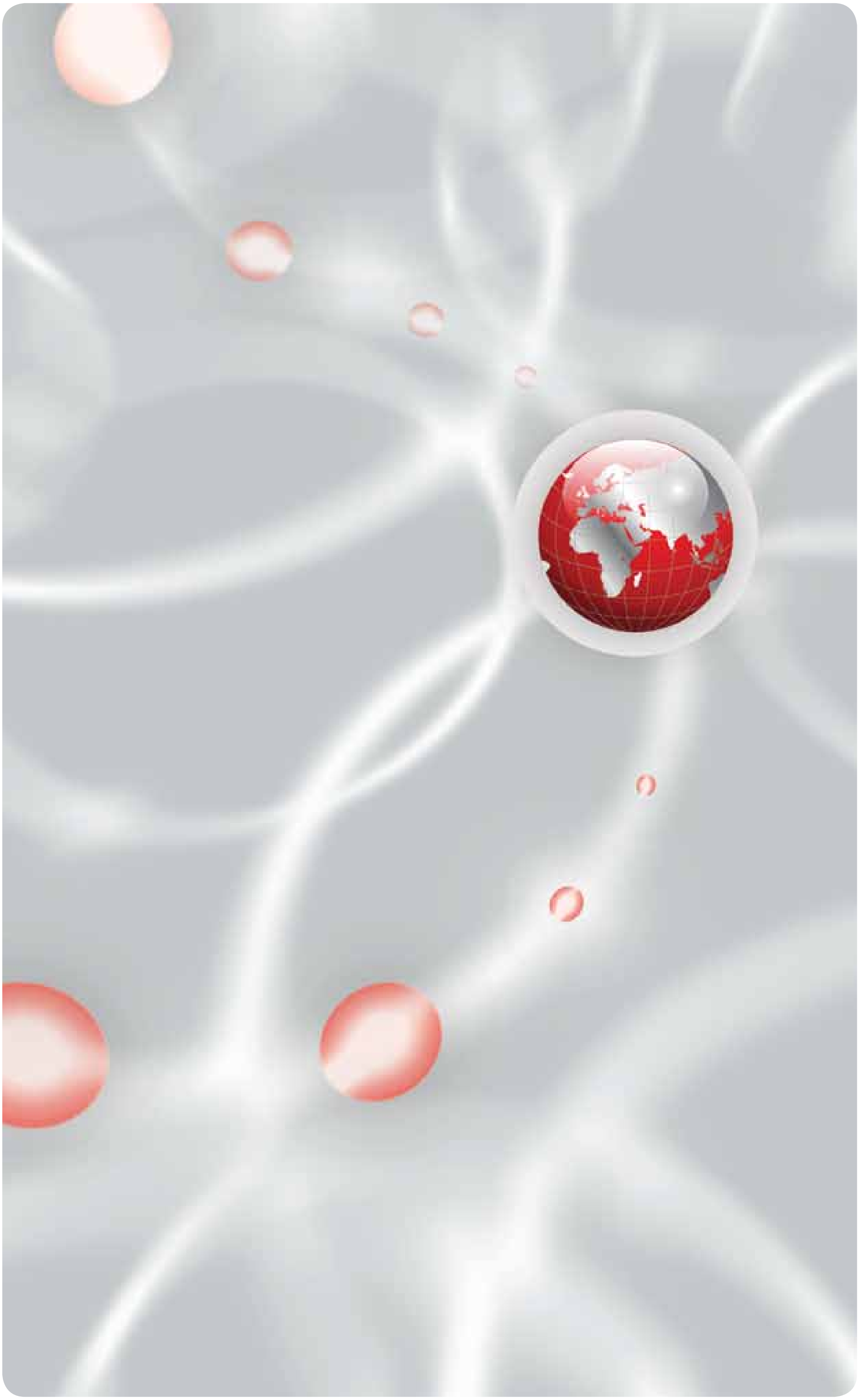
5 Exascale computing and related software 23
5.1 Policy area and goal 23
5.2 Context: Current practices - achievements and limitations 24
5.3 Proposed approach 24
5.4 Recommendations 25

6 e-Infrastructure services 26
6.1 Policy area and goal 26
6.2 Context: Current practices - achievements and limitations 27
6.3 Proposed approach 28
6.4 Recommendations 29

7 Data infrastructures 30
7.1 Policy area and goal 30
7.2 Context: Current practices - achievements and limitations 30
7.3 Proposed approach 32
7.4 Recommendations 33

Annex I - Editorial Responsibilities 34

Annex II - Terminology 35



Foreword

It is our pleasure to present the 7th White Paper from the e-Infrastructure Reflection Group (e-IRG). The White Paper series is instrumental for e-IRG in carrying out its mission to pave the way towards a general-purpose, European e-Infrastructure for research.

The case for e-Infrastructures gets stronger every day. Scientists from a growing number of disciplines empowered with simulation-based methods are increasingly being overwhelmed by the data deluge. At the same time, the e-Infrastructure landscape is changing by means of a service-oriented framework, which enables increased innovation potential and cost-efficient access from a widening range of users, thereby strengthening the socioeconomic impact. This favourable development is recognised at both national and European levels, with forceful e-Infrastructure agendas or strategies to promote an efficient governance for the research ecosystem.

The e-IRG vision for the future is an open e-Infrastructure that enables flexible cooperation and optimal use of all electronically available resources. Our vision underpins the Digital Agenda for Europe and will help enable the transition to an open innovation model where vast amounts of digital information can be harnessed for economic and social benefit. Furthermore, fulfilling our vision will result in the realisation of an e-Infrastructure that will help narrow the digital divide in Europe and support cohesion by enabling improved inter-regional digital flow of ideas and technology.

These are just some of the areas that will affect the success of e-Infrastructure, and its wide adoption by the user communities, and that will play a key role in shaping the European Research Area and the future framework programmes of the European Commission.

Topics in the present White Paper address both new and perpetual questions related to e-Infrastructure, such as:

- How do we deal with the increasing energy demands of computing?
- What software is needed to fully harness the power of future HPC systems?
- What are the appropriate governance models for e-Infrastructures?
- How can we facilitate access, discovery and sharing of large and diverse sources of scientific data?
- How can we further advance research networks, and adopt and implement new e-Infrastructure services?

We are grateful to the Spanish, Belgian and Hungarian EU presidencies who have hosted the work leading up to the White Paper, and in particular to Rosette Vandenbroucke for assuming responsibility for the editorial coordination and organisation of the corresponding workshop. The e-IRG delegates are officially appointed by their national government directly through the ministry (or ministries) responsible for science and research, or their equivalent. Each chapter of the White Paper is championed by a delegate, with contributions from others. Without the collective effort of the delegates, this White Paper would not have been realised. The work has been supported by a very competent secretariat through the e-IRG support programmes of the e-IRGSP2 and e-IRGSP3 projects. In particular, the continuity provided by Fotis Karayannis has been indispensable in preparing the White Paper.

Gudmund Høst (chair) & Leif Laaksonen (outgoing chair)

Introduction and summary

The 2011 edition of the e-IRG White Paper covers 7 well-chosen topics with "e-Infrastructure Governance" as the main contribution. The White Paper process consisted of four basic steps:

- The selection of the (independent) topics by the e-IRG delegates.
- The e-IRG Brussels workshop in October 2010, where reflections on the topics and the related issues were collected and summarised.
- The main editing phase supported by the representatives of the e-IRG member countries acting as "champions" of the selected topics. Besides the main editor of each section (champion) there were also other e-IRG delegates or external experts acting as co-authors.
- A consultation phase, where comments were gathered from a wide base of e-Infrastructure providers, e-Infrastructures users and e-Infrastructure-related projects. The comments and answers are all listed (including their original form) under **http://www.e-irg.eu/publications/white-papers.html.** The champions were responsible for providing answers to the comments.

Sustainability of e-Infrastructures is a major concern. In the 2009 White Paper, e-IRG tackled this subject for computing e-Infrastructures and recommended funding and interworking as basic foundations of their sustainability. This 2011 White Paper goes a step further in the sustainability discussion by considering the governance of all e-Infrastructures. The number and size of e-Infrastructures are growing. Are the governance models currently used sufficient to guarantee their future? Who should be the players in the governance model besides the service providers: the governments, the private sector, the user or a combination of these?

Research networks are so well established and are considered such a vital part of day-to-day work that we run the risk of forgetting to think about their future. The White Paper takes the lead by reflecting upon this issue.

The 2009 White Paper included a chapter on security. The current White Paper extends this topic by concentrating on authentication, authorisation and accounting. AAA requires the adoption of new visions to realise the interworking and hence also the sustainability of the e-Infrastructure ecosystem.

Efficient energy use, green energy, cheap energy, and energy consumption by ICT are daily topics on many management agendas. And is cheap energy compatible with green energy? Topic four of this White Paper deals with these questions at the level of e-Infrastructures.

Supercomputing is not at a stand-still and continues its spearheading innovation. Countries and companies strive to own or build the most powerful supercomputer. But can we go from T-flops to P-flops and can we really and efficiently use such a computing power? Does a software revolution need to take place before supercomputing can make a leap forward?

Services are another important part of e-Infrastructures. Users might not be interested in the pure infrastructure part but rather in the services that are provided by the e-Infrastructures. Which services should be delivered and with what quality?

And finally the very intriguing "data infrastructures" are addressed. e-IRG has already dedicated a separate document, the "**Data Management Task Force report**", to the data issue and a new document will be made available towards the end of 2011. However, at this stage we also wanted to include a few recommendations in this White Paper for the setup of data infrastructures.

Rosette Vandenbroucke – White Paper editor

1 *e-Infrastructure governance: From management and international aspects, to legal and financial issues*

1.1 Policy area and goal

Wikipedia defines governance as the set of processes, customs, policies, laws, and institutions affecting the way an organisation or ecosystem is directed, administered and controlled[1]. Governance also includes the definition of the relationships between the many stakeholders involved and between the goals of the organisation concerned. The recent e-IRGSP2 "Final Legal Issues Report" draws a clear distinction between IT governance and e-Infrastructure governance[2]. This chapter of the White Paper aims to provide recommendations about e-Infrastructure governance to ensure efficient, effective, transparent and accountable operation and use of e-Infrastructures. We consider the structures as well as organisational and relational aspects together with the governance process, distinguishing strategic processes and operational management and the various functional aspects of governance, e.g. the supporting legal and financing structures.

1.2 Context: Current practices - achievements and limitations

The European Union aims to become the most competitive and dynamic knowledge-based economy in the world. This vision relies on strong interactions within the "knowledge triangle" of education, research and innovation. The goal of the European Research Area (ERA)[3] is to create a barrier-free space and to provide attractive conditions and effective and efficient governance for carrying out research and investing in R&D-intensive sectors in Europe. Maximising the benefit of the investments in research infrastructures for their intrinsic beneficiaries, i.e. the users, requires that the users must be taken into account in governance policies.

Governance policies have to be based on common goals for the further strategic development of e-Infrastructures. Examples of this development are the demand for and the necessity of Green IT, the need for massive computational power (exascale computing), the increasing amount of data, the seamless access to services for users, the internationalisation of scientific research and the involvement of the user communities in governance of e-Infrastructures. Governance policies will also have to support the free movement of knowledge across Europe, which is

1 http://en.wikipedia.org/wiki/Corporate_governance

2 http://www.e-irg.eu/images/stories/e-irgsp2_d4_3_approved_by_the_consortium.pdf

3 http://www.leru.org/files/general/innovation-union-communication_en.pdf

acknowledged as the ‘fifth freedom’ for all EU Member States, after the free movement of goods, capital, services and people.

Today, European e-Infrastructure projects and e-Infrastructures, like the European Grid Initiative (EGI), the high-performance GÉANT network consisting of interconnected NRENs and campus networks, and the partnership for a high-performance computing infrastructure PRACE, are all part of an ecosystem to serve a range of different users: from the top-level scientific institutions in Europe, including intergovernmental scientific organisations and laboratories[4], to universities, research institutions and individual researchers. Across Europe, seamlessly interworking and easily accessible e-infrastructure services will be the foundation for the ERA which is targeted to be completed by 2014. It is not simply the resources, but rather the services to the user that will draw on a combination of various resources across Europe, which will create the capability to bring researchers from across Europe together in international, virtual teams and organisations. A holistic view on this e-Infrastructure ecosystem is therefore needed in order to meet the challenge of governing such a system effectively, efficiently and in a way that is future proof. This requires a deliberation of the strategies (e.g. concerning procurement, societal demands or industry involvement) and early involvement of all relevant stakeholders in order to realise a solid basis for the further development of e-Infrastructures.

The shift from mere resource provisioning to a system of infrastructure services will have a considerable impact on how such infrastructures are funded and financed. Different financing schemes may be appropriate to allow for the specific situation of given e-Infrastructures. Different modes are currently in use. An accepted approach in international e-Infrastructure cost-sharing between countries is based on GDP[5] (such as in GÉANT and EGI). On the other hand, PRACE confines the investments and the operating cost, including much of the control, to a smaller group of countries. In the financing, some countries make a clear distinction between funding for the network innovation, which is centrally funded on a project base, and funding for the use of networks for which the connected institutes are being charged on a use basis. In many countries, High-Performance Computing is funded via the resource owners, but in some cases partly via the user budgets. In some projects central funding is used to compensate for users in disadvantaged regions. If e-infrastructures are to be operated in the medium to long-term in a sustainable manner, then users will need to be given a choice for the best available services regardless of national boundaries (ERA). This includes the choice between the current e-Infrastructures and the commercial market for commodity services.

To guarantee that timely innovation takes place in e-Infrastructures for science and higher education so as to fulfil the legitimate demands for advanced services ahead of what commercial market offerings can provide, the active participation of users with leading edge service requirements in strategic governance decisions concerning e-infrastructures is essential. In EGI, European organisations such as CERN and EMBL are members of the EGI Council and user representation is also provided in other policy groups within EGI. TERENA has three different membership categories. Besides the NRENs as national members, international members such as CERN and ESA also have voting rights. Commercial providers can participate in the discussions as associate members without voting rights. In NRENs, user participation is arranged in several ways. In some NRENs universities and scientific institutes are directly in charge of the multiyear strategic plans, whereas in others, user involvement is organised through a council of users. Some NRENs are associations of user organisations and several are fully relying on user funding for both innovation and the provision of services.

Users must be able to influence the strategy and decision-making process for the related investments if the infrastructure is to be kept sustainable and healthy. However, it should be remembered that there are always different levels of user groups: the so-called ’leading edge

4 Such as the European Organisation for Nuclear Research (CERN), European Space Agency (ESA), European Organisation for Astronomical Research in the Southern Hemisphere (ESO), European Molecular Biology Laboratory (EMBL) and others.

5 Gross Domestic Product

communities' versus the 'masses'. At any given moment in time these two groups may have different requirements. As many cases in the past have shown, however, the requirements of today's leading edge user communities will be the needs of all users tomorrow. Any model should therefore be able to keep timely innovation alive.

1.3 Proposed approach

The essence of the proposed approach is that e-Infrastructure governance will have to show a shift towards a user-driven approach. There will be **no one size fits all** solution. Different technical, political and commercial developments, such as the virtualisation of services, the emergence of cloud computing, the ambition of establishing an ERA and the ever increasing need of leading edge user communities for services far beyond what the commercial market can offer, will drive this process.

Changes in the governance system, its legal base and its financing mechanisms, should lead to strategic user empowerment, should guarantee the flexibility and internationalisation of the supply chain, and should safeguard the crucial role of e-Infrastructures in bringing ICT innovation to the users in its earliest stage. Such changes should include provisions that allow users to switch to services with different funding and charging mechanisms, e.g. moving from a centrally funded service to one that is partially funded from the user budgets.

More specifically, a future **governance system** for e-Infrastructures and the related **financing mechanisms** and **legal structures** should allow for:

- ***User-centric approach:***
  - Users and user communities, in particular the leading edge users, will acquire a key role in setting the strategy because the quality and continuity of their research depend on the long-term availability of the right e-Infrastructure ecosystem.
  - Timely e-Infrastructure innovation to serve user communities ahead of what the commercial markets can provide will remain a public responsibility at both the national and European levels. It will continue to require a concerted effort and large-scale funding on all levels. For national e-Infrastructures with international significance, national financing could be matched with appropriate EU funding.
  - The funding of the use of e-Infrastructures services should increasingly be paid out of the budgets of users and user projects. The users and projects should explicitly budget for this. However, appropriate measures need to be taken to avoid unintentional funding cuts as a result of this change. Such a shift will not only enhance efficiency but also promote enhancement of national and international service provisioning and will encourage commercial offerings of e-Infrastructure-services whenever these are viable and more efficient. Nevertheless, commercial lock-in situations must be avoided.
- ***Strategic and financial concerns:***
  - The governance system should be supported by an elaborate system of metrics to establish the value and costs of the services and the service delivery systems.
  - Long-term financial guarantees will be necessary, coupled with continuing benchmarking mechanisms to keep the players in the system eager and keen.
  - The system should include provisions and budgetary guarantees for research projects who want to move to services with different funding mechanisms.
  - Unnecessary distinctions between national and international e-Infrastructure services and governance structures should be abolished.

- National and community regulations or practices that inhibit open and equal public and private cooperation in the use of e-Infrastructures should be reviewed so as to allow equal treatment of users in public and private research. This will also be necessary for cooperation in PPPs like ESFRI projects or the EIT initiative.
- The different territorial and functional roles of the various e-Infrastructure resources, like networks, computing resources and data storage facilities should be carefully considered. This should also allow for management of the different life cycles of the various components of e-Infrastructures and for the investment of the underlying IT resources.
- Policies to overcome the disparities between different regions of Europe should be paid for by the European regional development funds.
- National and European funding bodies should support the strategic alignment of the governance policies.

- ***Openness, neutrality and diversity of resources and services:***
  - The best practices and lessons learned from Research Infrastructures show that opening access to RIs for the full European user base, catalyses the development of an EU knowledge market.
  - Cross-organisational service level management will become an important issue, which will have to be supported by governance structures. Formalising the quality and management aspects of service provision practices and complementing these with tools and procedures from the established IT service management discipline would have many benefits.
  - Activities for the development of open standards for using the e-Infrastructure services should be promoted and supported on all functional levels and in all application areas. IPR on ICT products and services should not pose a barrier to such openness. This would make e-Infrastructure a flexible and living system and would avoid the occurrence of legal and financial roadblocks for later developments into unforeseen directions.
  - In planning and subsidising the innovation and rollout of e-Infrastructure, careful consideration needs to be given to the multi-domain and ecosystem nature of many e-Infrastructure components, and sufficient space should be given to the development of alternative strategic options.
  - Innovation and the digital divide are both important contrasting components of any e-Infrastructure policy and can be best addressed in separate mechanisms, both financially and organisationally.
  - Coordination of national, regional and European e-Infrastructure policies is required at the operational level and at the various regional, national and European levels.

- ***Legal structures supporting the strategy***
  - Legal structures, like ERIC, have to be effective. They should be easy to implement and should enable international cooperation on a non-governmental level. Governmental representation in those legal structures should be minimal or even not included. The structures should also follow the trend towards service provisioning with dispersed and virtual resources for computing or data storage and data management.
  - Appropriate arrangements for IPR and copyright should be in place, enhancing the value of public investments for all European user and service providers and avoiding barriers to interoperability.

1.4 Recommendations

- Establish a user-community-centric approach in strategic e-Infrastructure governance. This should include the appropriate funding mechanisms that make a clear distinction between the funding of service provision and the funding of innovation activities.
- Define the long-term financial strategy for e-Infrastructures aimed at a sustainable operation of services in a flexible and open environment that includes offers from commercial service providers.
- Address the problems of barriers to cross-border service delivery and quickly remove as many of these as possible.
- Introduce governance models that provide efficient and effective coordination mechanisms at all levels: regional, national, European and – where possible – global while providing the possibility for public and private research and cooperation via public-private partnerships (PPPs).
- Encourage important players in the use of e-Infrastructures, like ESFRI, Virtual Research Communities, … to investigate the impact of strategic changes in e-Infrastructure governance and financing on the operation of and access to international Research Infrastructures.
- Investigate the effectiveness of legal structures, like ERIC, for e-Infrastructures.

2 *Future of research networking*

2.1 Policy area and goal

Networks are an essential element of the e-Infrastructure. The goal of this chapter is to make recommendations for national and international policy makers for the further advancement of research networks, given the expected technological developments, the needs of the user communities and the evolution in the markets for information and communication services.

2.2 Context: Current practices - achievements and limitations

The current organisational structure of European research networking was defined and implemented in the late 1980s/early 1990s, hence well before the telecommunications liberalisation and before the Internet developed as a worldwide communications infrastructure.

In Europe, the 36 NRENs already work together via the GÉANT network infrastructure and its associated technical developments, as well as through the TERENA collaboration. The European Commission facilitates GÉANT via the GN3 project. Beyond Europe, GÉANT also procures international links to provide broader international connectivity. DANTE is a NREN-owned corporation that acts as a coordinating body for the GN3 project. A number of NRENs also provide additional advanced regional and international connections and developments, such as participation in the Global Lambda Integrated Facility (GLIF) that supports worldwide developments for using lightpaths. TERENA is the association of NRENs: a forum to collaborate, innovate and share knowledge in order to foster the development of Internet technology, infrastructure and services to be used by the research and education community.

The direct help from national governments and the European Commission supports innovation and helps to overcome the hurdles of the lack of a single market for telecommunications. This setup has worked well and has allowed Europe to play a global leading role in research networking by the continuous development of the national research networks, GÉANT and GLIF.

NRENs and other stakeholders are aware that the world is moving fast, that networking has moved from 'nice to have' to an indispensable service, and that the time has come to adapt and progress towards achievement of the EU goal of global competitiveness in the 2020 agenda. The only way to achieve this goal is to facilitate world class science via world class e-Infrastructures. Advanced research networks, ahead of what is commercially available, have proven in the past to be drivers of innovation and in the future these will also be an essential element in achieving the Europe 2020 ambitions.

The challenges ahead are manifold. The implementation of the ERA by 2014 is one of these. In view of the increasing global cooperation and competition in research, the networks must keep innovating, and measures must be taken to ensure that all researchers fully benefit from the "Fifth Freedom"[6] across the ERA, i.e. free circulation of researchers, knowledge and technology. Disparities between users and countries, the so-called digital divide between countries, and in many cases between different users in the same country, should be overcome or at least be eased. This is particularly important because advanced e-Infrastructures also act as innovation engines for research activities and for the ICT sector as a whole, and are recognised as catalysts for the progress of the economy in general. Another strategic challenge is that energy saving and environmental protection are becoming increasingly important issues within ICT and gradually in networking as well. Moreover, the data deluge[7] will change the relations between e-Infrastructures and their end users, demanding further internationalisation and coordination of the research area (e.g. ESFRI, EIT) and increasing cooperation between public and private research, i.e. public-private partnerships.

However, the main strategic thrust in networking is the user demand for integrated access to the various international e-Infrastructure services, which allows them to focus on performing their actual research. As a user-driven sustainable ecosystem, e-Infrastructure services need to be application-oriented, easily accessible, open and flexible so that they can continuously adapt to technological changes and evolving user needs. Research networks are already made available as a service, but the drive towards seamless access to all services, including the connected e-Infrastructure elements for computing and data storage as a fully integrated ecosystem, is new and calls for a continuous and global review of e-Infrastructure policies, with particular emphasis on continuous innovation and good governance.

Likewise, the availability of new technologies also calls for the innovation of the networking infrastructure and its services. This is complemented by the emergence of new commercial entrants in the research arena, creating a more competitive environment in the ICT sector and a market opening for innovative actors such as commercial brokers or the combined power of users with similar interests. Users need to be able to influence the networking strategy and, if possible, the decision-making process for the related investments, which is a challenge for the current governance system of many networks in Europe. The challenges described above will, at the very least, require a serious assessment of the present governance structures and the prevailing legal and financial arrangements currently in place.

2.3 Proposed approach

The first issue related to the future of research networks is drafting and maintaining the technical and functional innovation agenda for the next 5 years with due consideration to the consequences of the challenges described under 2.2. Such an agenda should be a living document and a common reference guide for all stakeholders in research networking and should cover all R&D activities of international significance in this area. The planned limited foresight

6 "Free movement of knowledge", a proposed addition to the Four Freedoms in European Union law.

7 Ref. 'Riding the wave'. How Europe can gain from the rising tide of scientific data'. Final report of the High Level Expert Group on Scientific Data. http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

study under the GN3 project that will be led by TERENA starting in April 2011 will be used to elaborate such a common agenda, and will include aspects like organisational consequences, the funding of development programmes and the interface with the EU Framework Programmes. The resulting agenda could become a tool to drive open innovation through competition, cooperation and flexibility, to promote research and broad international participation by the different stakeholders, to facilitate funding and to ensure knowledge sharing. It should, however, draw a clear distinction between the various types of research on networks[8], on the one hand and the development of advanced networks for operational services on the other.

Openness, neutrality and diversity should be the guiding principles in developing the networking infrastructures. Networking is inherently multi-domain and should be built in a federative and open approach, supplying worldwide connectivity based on globally accepted standards. This will not only facilitate the natural evolution of the e-Infrastructure ecosystem, but also simplify the interconnection with commercial providers, create opportunities to use commercial services by NREN end users, and enable open innovation with private research organisations and public-private partnerships (PPPs). Network services should be made available via a common user interface to allow integrated access to different e-infrastructure services.

The digital divide issue, both between and within countries, should be rigorously tackled, starting with an investigation of what this divide means exactly geographically, how it might affect the ERA, and what its causes are, so that the right remedies can be chosen. However, funding instruments should be used correctly: innovation funding should be used for innovation while structural funds could be used to address the digital divide. Furthermore, increasing bilateral, regional and multilateral cooperation between NRENs might help to narrow the digital divide, and it could also generate sufficient political pressure to enforce a single market for electronic communication services in the EU.

The most challenging issue will be getting the users to participate effectively in network governance and network innovation. A pressing argument for the users themselves will be the data deluge. However, user participation is mainly a general issue for all e-Infrastructure elements and will be dealt with by the governance chapter of this White Paper.

2.4 Recommendations

- Innovate in network provisioning and network governance to satisfy user demand and stay competitive at the global level.
- Use the planned GN3 foresight study led by TERENA to draft an Innovation Agenda for research networking to be used by all stakeholders.
- Build the networks as a federative and open system, giving flexibility and worldwide connectivity to public and private researchers and with seamless integration with other e-Infrastructure service providers.
- Rigorously investigate the causes of the digital divide between European researchers and combat this with the appropriate instruments.

8 E.g. research on real experimental networks deploying technologies not available for commercial operators (e.g. beyond 100 Gbps, open source hardware, support crash tests, etc.)

3 Authentication, authorisation and accounting

3.1 Policy area and goal

Authentication allows entities (usually users) to establish their identity within a specific e-Infrastructure. In a highly distributed and collaborative environment that crosses multiple administrative domains and national boundaries, valid authentication is of considerable value, not only in the researchers' home institution but also in a large part of their collaborative environment. Authorisation establishes the rights of individual users to perform certain operations within that specific infrastructure, where those rights are decided according to defined resource access policies. These policies are defined by the resource owner, or delegated to a trusted third party (e.g. a virtual organisation).

The overall objective is to establish and maintain the level of mutual trust amongst users and service providers that is needed for an open ecosystem to function. As an e-Infrastructure matures and its user community grows, requirements for aligning authentication and authorisations grow as well. This must translate into:

- Improved usability, lowering the threshold for researchers to use the services.
- Improved security and accountability, which often conflicts with the usability requirement.
- Leveraging of existing identification systems, such as that of the employing organisation.
- Enhanced sharing, allowing willing users to minimise the burden of policy enforcement.
- Reduced management costs, freeing resources for other service or research activities, and providing a sound basis for accounting.
- Improved alliance with the commercial Internet, which also improves interaction between scientists and society.

3.2 Context: Current practices - achievements and limitations

The development and deployment of authentication and authorisation infrastructures (AAI) has taken place in different research and education environments, as well as in the private and public sectors. In grids, for example, access control is based on standard X.509 certificates and a set of well-established global procedures and policies, supported by a widely deployed software base as part of grid middleware. The International Grid Trust Federation (IGTF) and its European branch, EUGridPMA, work to assure harmonisation and supervision policies for the certification authorities, which provide researchers with internationally accepted electronic identities that support real-time and asynchronous access to international resources. However, there are two factors that hinder a widespread adoption of X.509 certificates. Firstly, X.509 certificates have proven to be difficult for the user to handle properly and securely. Secondly, the current organisational setup is not suited to issuing certificates to a very large user base.

For the past five to ten years, Europe's NRENs, under TERENA's coordination, have operated identity federations, and provided services to a very large number of users within the academic and research community. Based on open standards, these national identity federations initially focused on providing access to web-based resources (such as data repositories and e-Journals). Although in such cases, the user typically acts as a consumer only, a full e-Infrastructure should also allow the user to act as a producer of information, which is supported by SAML. In this context, clear and simple mechanisms for accessing and managing authorisation policies are a key requirement. Within Europe, the GÉANT3 (GN3) project is about to deploy eduGAIN as a service within the GÉANT3 infrastructure. eduGAIN is intended to connect the different national identity federations into a seamless common identity space that supports real-time access to web resources across Europe. However, the maturity of the national AAIs differs substantially between individual countries. This fact currently hinders the widespread adoption of eduGAIN-enabled services across Europe. Nevertheless, interoperation between several federations such as the Kalmar Union is already a reality.

Third party players outside academia include providers of user-centric identity management models, especially in the field of Web 2.0 applications (e.g. OpenID), as well as governments offering identity infrastructures rooted on a legally recognised and authoritative framework (such as the STORK11 project).

3.3 Proposed approach

There is progressive convergence in grid and NREN-operated identity federations, which is promising in terms of future services that can be offered to growing user communities.

REFEDS, a group sponsored by TERENA, aims to include representatives of each of the organisations providing identity services in the global research and academic sector. The group is taking a very active role in fostering technical and policy alignment, supporting the adoption of new technologies and best practices, and seeking interaction with identity service providers and consumers in commercial and other sectors.

A sound development of AAA requires specific attention for the following topics:

- Not every European Member State operates a fully production quality AAI that spans its entire research and education sector.
- Interoperation between existing national AAI has been successfully demonstrated (eduGAIN) but has not yet reached maturity.

- Full support for the management of distributed dynamic Virtual Organisations (VO, such as ESFRI projects) is only being tackled now.
- Robust and open accounting solutions for the e-Infrastructure are needed to monitor the services and allow for comprehensive service level management.
- Integration of user-centric and governmental infrastructures with academic AAIs is a largely undefined area at this moment.

Although at a technical level these limitations are actively being addressed and initial solutions will become available in the short to medium term, a lot of work still needs to be done at the policy level.

3.4 Recommendations

- Continue to improve national infrastructures and their alignment with agreed standard procedures for identity management, accounting and assurance, with the objective of technical interoperability between all national AAIs.
- Accelerate the continued integration of different identity technologies, through supporting active collaboration between the IGTF, GÉANT and relevant European and international working groups.
- Require that, wherever possible, future pan-European e-Infrastructure and ESFRI RI projects define their access control policies and mechanisms from the beginning, in accordance with the standards and best practices adopted by the community.
- Draw up a roadmap to book progress for all stakeholders in the unified integrated approaches to replace existing authentication and authorisation infrastructures based on national AAIs.

4 *Energy and Green IT*

4.1 Policy area and goal

Green computing or Green IT refers to environmentally sustainable computing and more widely to IT in general. It covers the study and practice of designing, manufacturing, using, and disposing efficiently and effectively of computers, servers, and associated subsystems like communications systems with minimal or no impact on the environment. The general goal of the Green IT initiative is to reduce the use of hazardous materials, maximise energy efficiency during the product's lifetime, and promote the recyclability or biodegradability of defunct products and factory waste. For the development of the e-Infrastructure in Europe this means paying particular attention to the energy consumption of its IT components. On the one hand, the total energy consumption should be reduced and on the other hand, the energy efficiency has to be maximised, e.g. re-using the heat generated by the computing infrastructure.

Efforts directed at improved energy efficiency in recent years have also focussed on reducing the cost of energy consumption at the hardware level. Yet this efficiency can also be realised by looking for cheaper sources of energy, although that whole process requires an analysis of the situation before the location is chosen. However, buying cheap energy alone will not be enough to realise the Green IT objectives. We also need to consume as little energy as possible, preferably from renewable sources or without any carbon footprint, and use the energy we consume as efficiently as possible. It appears that the issue is multidimensional.

4.2 Context: Current practices - achievements and limitations

Energy and Green IT are very broad topics. This section will focus on those subjects that are important in the context of e-Infrastructures and will not dwell on details related to energy production, transfer or usage in other applications. The energy consumed by ICT and e-Infrastructures is currently just 3-5% of the global power consumption. However, since information is becoming a crucial resource, this ratio will increase in the following years.

Data handling, processing, transfer and storage all generate heat. In fact, nearly all of the power provided to data-processing systems is converted into heat. A large data centre currently consumes megawatts of energy and unless the efficiency of these plants increases, they will consume tens of megawatts of power with a corresponding increase in the amount of heat that has to be transferred out of the data centre rooms and dissipated in the environment. Computing will always consume energy and produce heat but depending on the location of the data centres and their design, the environmental impact of this can be reduced. The most important aspect that should be considered is reducing the CO2 footprint of the e-infrastructure facilities.

Supplying the power

The IT resource providers require huge amounts of power that have to be transferred from a power plant to the resource location, e.g. data centre. The energy transfer is a potential source of considerable losses: 10 to 20% of the power delivered is estimated to be lost due to transmission losses. Such losses can be minimised by locating the infrastructure as close as possible to the main power source. Availability of green power in a given area may influence the choice of the location for the e-infrastructure facilities. Other limitations which may influence the decision, e.g. costs of networking or higher maintenance costs, should also be taken into account.

Using the power

The PUE (Power Usage Efficiency) is a factor that indicates how efficient the power consumption is in a given facility or device. It is the ratio between the power consumption of devices that provide so-called “core” functionality and the additional devices installed to provide a working environment for the major devices. For a data centre it is the ratio between power drawn by servers, storage, and so forth and losses caused by transformers, air conditioning, and UPS systems. A well-designed green data centre should have a PUE as close to 1 as possible. That would mean that all of the energy supplied is used solely by the e-infrastructure and nothing is lost on cooling, et cetera. In general most of the technologies used for increasing the reliability of the data centre such as diesel generators, UPS, redundant transformers et cetera are increasing the PUE of a facility. Due consideration should therefore be given to the level of reliability required and whether all of the devices installed need the same level of security. A good example is a telecommunication facility: reliability must be high as having uninterrupted connection is an essential feature, but such a facility does not consume a lot of energy. An HPC centre has exactly the opposite characteristics: the power consumption and heat density are very high and in the future may even be in the order of tens of megawatts. However, in such centres the need for uninterrupted work is far less critical and that provides space for reducing losses and striving towards a better PUE value.

The chips of PC systems are not designed for energy efficiency; they are built with price and performance as major goals as these are the factors affecting the sales. The increased power costs for average household caused by a less efficient PC working for several hours a day is barely visible, whereas better responsiveness of the system is noticed immediately. In the large IT installations where the number of chips present is 3 or more orders of magnitude greater, the installation of power-hungry but cheap processors results in enormous increase of both the computational power and the energy consumption. In many EU countries, the current funding model for ICT and e-Infrastructures treats the acquisition and the maintenance costs as two separate items that are not correlated in any way. Such policy is driving the power consumption increase, as the computer centres are assessed by the amount of computational power that is delivered to the users within the budget for new hardware.

Cheaper energy versus location

The redesign of the data centre operation on a much broader strategic scale may provide us with far more flexibility, thereby enabling a decrease in the costs of infrastructure maintenance. The change of the operation means separating the hitherto traditional bond between IT hardware on the one hand, and IT operations, competence and usage on the other. In other words, allowing for the hardware to be situated away from the competence, in places offering cheap, abundant and green energy. We can already observe some exemplary cases realised in the HPC world, like the one in Finland by CSC.[9] Nevertheless it once again needs to be emphasised that the Green IT objectives will not be fulfilled by buying cheap energy, but by consuming as little energy as possible.

Dealing with the heat

The e-infrastructures will consume more and more energy. This is unavoidable in a modern ICT community where information becomes a critical resource. There are several ways of dealing with the heat produced by the data centres. Generally, the two alternatives are efficient dissipation of the heat or trying to re-use the heat.

The most appealing solution is to re-use the heat as then we basically have a free energy source. Unfortunately, the possibility of re-using heat is limited by the difficulties in transporting heat and the relatively low temperature of the heat waste. Data centres should therefore be located near to potential heat consumers so as to maximise the gains. Despite the potential gains, the use of heat waste should be carefully considered, as the costs of implementing such installations may far outweigh the possible benefits.

In data centres currently operated, the most common way of removing the heat waste is to pump it into the atmosphere. The efficiency of this solution largely depends on the climate where the installation is placed. In some cases, it may require as much as 50% additional energy to cool the heat generated by the infrastructure. The colder the climate, the better the efficiency. When the temperature difference between the outside air and the air that transports the heat from the data centre is big enough then free cooling can more or less be obtained. Free cooling means that almost no additional energy is required to dissipate the heat waste. There are technical solutions that reduce the energy costs of the cooling, but again the price of the cooling solution rises with increasing efficiency and the air temperature it can efficiently work with. Changing the cooling medium from air to better heat conductors, such as water, can increase the cooling efficiency. Such a change may be difficult or even impossible to implement in existing facilities as these may be limited by the infrastructure of the data centres or the distance to a lake or river.

4.3 Proposed approach

The work on Green IT has already started but with no consistent vision on how to proceed globally. It is obvious that the major goals are to reduce energy consumption, increase energy efficiency and minimise the influence on the environment. However, the proposed approach should be done at several levels of stakeholders involved in the problem: policy makers, hardware vendors, hardware/services providers, and end users. The most important questions that have to be answered before defining a global approach are:

9 http://www.csc.fi/english/csc/news/news/data

- What can be done to make existing e-infrastructures greener and minimise their environmental impact?
- How can green energy be provided efficiently?
- What measures are required to force the shift to more green solutions?
- Green IT requires additional effort and money – are the benefits big enough to warrant investment in changes in such a small area? What is the sweet spot where the benefits justify the costs?
- Is the economy a strong enough factor to be a driving force for changes? Is the current model of IT funding encouraging changes towards green solutions?
- How can we analyse all the advantages and disadvantages to find a best location for the data centre where cheaper energy can be provided? Is green IT a possibility for some regions to develop? Good climate, cheap energy or possible heat consumers may be the factors influencing decisions about where to build new e-infrastructures facilities. However. the decisions should not be to the detriment of environmental protection issues, which have the highest priority.

4.4 Recommendations

- Decrease the energy consumption of all e-infrastructure components by providing a different kind of architecture and working out more efficient software management procedures.
- Develop more efficient ways of using the provided energy by increasing the efficiency of the cooling systems and reusing the heat energy for different purposes.
- Analyse the environmental impact of various approaches to energy maintenance.
- Promote R&D on Green IT topics and provide more service management procedures.
- Work out and promote Green IT standards at an international level like the Energy Star or the green grid.
- Locate data centres at optimum locations in terms of the balance between green energy and energy efficiency

5 *Exascale computing and related software*

5.1 Policy area and goal

The exascale computing facilities may open new possibilities for grand challenges applications if a good background for their use is created. There are currently many limitations that do not allow the construction of an exascale system. These are not just budgetary issues but also boundaries related to technology, energy consumption and software scalability. Nevertheless, design studies have already started in science and industry communities. Exascale computing will probably be available before the end of the decade. It is vital that Europe does not lag behind incumbent leaders in HPC, such as the United States, Japan or the newcomers from China, in the pursuit of exascale computing technologies. The e-Infrastructure community in general must decide how to prepare for the transition to exascale over significant technical hurdles. It must also consider how it will convince the public, industry and policy makers that working towards exascale computing will benefit European economies in terms of job creation, innovation, and cutting-edge research as a result of the opening up of boundaries for new grand challenges in e-Science.

5.2 Context: Current practices - achievements and limitations

Significant challenges have to be addressed if the extra exascale computational power is to be harnessed. A number of projects in the EU and elsewhere are beginning to address this. The main issues are:

- The design of new hardware and software architectures that will be efficient enough to address exascale.
- The increase in power consumption, which may be restrained by using new technologies and/or heterogeneous architectures, where instead of a single processor architecture, different architectures are used in the same machine.
- The increase in concurrency, as the increasing number of processors and cores entails a change in scale at the level of parallelism that must be exploited by the software.
- The failure of standard parallelisation techniques to scale up in an environment restrained not only by the number of CPUs but also by core-to-core respective memory to file bandwidth, non-uniform memory access as well as backbone topology.
- Resilient architectures, programming models and applications, which will ensure that the system produces acceptable results even in the presence of hardware failures, which may become common as the number of components will increase steadily.
- The development of new programming paradigms allowing the effective use of a machine that is 1000 times bigger machine than those currently in existence: better compilers, monitoring tools, hiding software complexity.

These challenges underlie a paradigm shift in software for exascale computing. The main components of this shift are:

- A new programming model, as a move beyond Message Passing Interface (MPI) programming, which has been the dominant paradigm in supercomputing programming.
- Establishing a performance indicator over differing architectures that considers all significant parameters of a configuration (e.g. cost per execution, memory usage, bandwidth), not merely PFlops or execution time.
- The leveraging of heterogeneous computing by operating systems, software libraries, compilers, toolkits, et cetera.
- Establishment of general testing procedures to verify the correctness of a highly parallel implementation.
- Setting standards (technical, logistic, legal) for community-based development, documentation and release of major software packages.
- Last but not least, a practical approach to data safety and security.

5.3 Proposed approach

Although the transition to exascale will probably be gradual (in the same way that the transition to petascale is gradual), in Europe we must start, as soon as possible, to identify solutions for the software challenges, if users are to utilise the future machines to their full potential. In particular, users, and especially those that lack a computer science background, cannot be expected to program exascale computers effectively without appropriate software tools that hide complexity, facilitate parallelism, and let them concentrate on utilising their domain knowledge. It has already been recognised that software engineering has an important role to play in computational science and engineering (see Z. Merali, "Computational science:..Error… why scientific programming does not compute", *Nature*, 467, 775-777, 2010); we can only expect this role to increase in the coming years.

The push to exascale is not solely a matter for scientists, however. Industry, the wider public and policy makers must be well informed, as significant public and private resources will have to be invested. Members of the e-Infrastructure community must therefore communicate the value of pursuing exascale computing to the lay public and policy makers, and they should view this an integral part of their project and not an onerous additional task. This should happen with the support of and in close collaboration with national and European HPC centres as well as the potential exascale end users, e.g. ESFRI projects.

5.4 Recommendations

- Encourage the development of European hardware technology in order to compete and cooperate with the current leading countries in HPC.
- Dedicate resources to the study of new programming models, algorithms and languages, porting software libraries and software tools to exascale environments, and preferring open source software solutions to leverage existing know-how in a cost-efficient way.
- Identify new grand challenges in science that are able to utilise the exascale platforms.
- The partnership between users of exascale computing, industry and computer scientists must be encouraged, and scientists should be given the opportunity to liaise with programming experts.
- Specialists must create training materials, including robust and easy to use “cook books” for users, especially for those who are not computer scientists.
- Ensure that the value of the scientific case for exascale computing is well understood and appreciated by society at large by means of knowledge dissemination, and engagement with the public, policy makers and industry.

6 e-Infrastructure services

6.1 Policy area and goal

The role of e-Infrastructures in research and education is to support the access, sharing, federation and exploitation of the collective power of scientific resources, both human capabilities and tangible facilities in a widely distributed and highly collaborative research environment that also crosses national boundaries. This multicolour function of the e-Infrastructure is well illustrated by the emergence of e-Infrastructure as a service where such an approach is requested/accepted by the users, and promises additional benefits.

In Europe, the GÉANT network infrastructure of interconnected campuses and national research networks provides excellent end-to-end network connectivity between more than 3000 campuses and up to 40 million users. On top of these networking services, a growing number of high-quality, complex computing grid, HPC, collaboration tools and data infrastructures are also becoming available as a service.

A major goal of this chapter is to analyse and describe how e-Infrastructure resources can be offered as a service to users in a reliable, scalable, customised and secure setting. The challenge is to:

- upgrade/refine the present services
- develop/introduce new services (also by rationalising the service portfolio)
- improve the governance/management of e-Infrastructure operations offered as services
- extend/intensify cooperation and collaboration in the e-Infrastructure area
- establish and gradually introduce a sustainable business model for e-Infrastructure operation and services

Particular attention will be paid to the investigation and, if advisable, exploitation of converging e-Infrastructures such as a High Performance Computing & Networking Cloud involving Infrastructure as a Service, Future Internet Services, Software (Applications) as a Service, and more.

6.2 Context: Current practices - achievements and limitations

Context

Traditional e-Infrastructure services, such as communication, security, authentication, conferencing, computing, et cetera, have been provided by the NRENs (National Research and Education Networks) for more than 20 years. Earlier service portfolios were simple and easily exploitable, due to their development (mostly for restricted groups) as separated, individual services based on dedicated equipment and unique software components. Their separate functioning also meant an almost complete absence of interoperability between different implementations or other services.

As research became more complex and distributed, the e-Infrastructure had to cope with these developments. The changing requirements, like the increasing need for shared international access to remote resources, increased security, economies of scale for shared use, and the emergence of virtualisation techniques[10] gradually led to novel federated services – the grid and cloud, service-oriented architectures (SOA), and the provision of sophisticated on-demand access to different shared resources, like hardware, software, infrastructure, platforms, et cetera. All of these as complete services rather than just tools, devices, or facilities.

IaaS and similar service-oriented solutions are emerging in both academic research and industry, e.g. by exploiting the opportunities provided by the clouds. These solutions respond to ever-changing requirements by means of on-demand provision of requested resources for a widening spectrum of diverse applications, and they also stimulate a service-oriented approach to software development and deployment. The goal is to integrate the existing software frameworks into a connected, interoperable environment that fulfils the needs of the users, and to use a modular environment of lightweight software tools so that stability and performance problems can be easily diagnosed and resolved.

A major implication of the services' shift appears in the changing division of responsibilities between the user and the supplier: the responsibility of linking the service demand to the user need is moving towards the supplier. This means widening the distance of the users to the physical resources. Another important feature of these service-oriented solutions is that most of the higher-level complex services are based on well-defined interoperable and distributed lower level services.

Current practices and future developments – achievements, challenges, objectives

More and more services will be offered on top of the GÉANT network infrastructure. And a growing number of computing and data services will be provided via the commercial world where cloud computing and virtualisation of both hardware and services are guiding factors for technological change.

If this challenge of implementing changes to IaaS is to be met then important issues will have to be resolved by setting ambitious goals:

- Heterogeneity of e-Infrastructure services necessitates a strong focus on standardisation efforts, at the level of both the users and the service providers.
- Openness as well as adaptability of the standards is a must, in view of the diverse user needs. There is a serious threat of user lock-in in captive service offerings using proprietary standards.
- Service-oriented architectures should aid interoperation of e-Infrastructure services, if and where possible.
- Virtualisation should be used to build virtual research environments (VRE) and virtual research communities (VRC).

10 See also e-IRG WP2009, Chapter 5 and e-IRG Roadmap 2010, Chapter 2.4

- e-Infrastructure users should be better served by improved friendliness of access, adapted customisation of services, and tailored support and training.
- Multi-tenancy of services should enable better sharing of e-Infrastructure resources and costs across a large number of users, improved resource utilisation, increased peak-load capacities, and installing/operating infrastructure resources in locations with lower costs.
- Special services are to be offered by establishing service portals or centres dedicated to specific user communities, specialised service providers and specific large-scale projects.
- Coordination should result in exchanging services and sharing service portfolios among co-operating e-Infrastructure providers, as well as in joint tendering or licensing by them.
- Adequate governance models of operating e-Infrastructures at the European level should be widely and cooperatively applied by diverse e-Infrastructure areas and user groups.
- Stability and sustainability of the infrastructure are to be improved by developing and gradually introducing fair and straightforward business models, business standards and charging practices.
- PPP approaches for building strong innovation chains are to be applied on the basis of shared e-Infrastructures.
- Novel e-Infrastructure approaches should aim to serve the user with commodity services and stimulate high-end research and innovation.

6.3 Proposed approach

In an IaaS environment, the primary responsibility of policy makers and research authorities is to ensure that the user communities will be heard. This will require an approach that covers many relevant aspects:

- Work on standards for e-Infrastructure services should be promoted at various levels and domains, including the multitude of international application-oriented communities. This will be a major organisational effort.
- Virtualisation and SOA approaches should be widely exploited in service development.
- There should be an emphasis on simplifying the use of e-Infrastructures by friendly access, transparent interfaces and service offerings, easy configuration, customised training, service portals and centres, and campus level support.
- Contentious governance issues that impact the adoption of IaaS (or cloud infrastructure services) must be addressed. These include transparency, privacy, security, availability and performance, compliance with data protection directives, as well as the lack of widely adopted open standards.
- Applying service level management tools and procedures in e-Infrastructure service provision practices has various benefits. These allow users, service providers and funding agencies to investigate e-Infrastructure services from a perspective of individual use cases.
- Exchanging services, sharing service portfolios, and other forms of improved cooperation by and between NRENs, NGIs and other national e-Infrastructure service providers should be exploited for better geographic and disciplinary coverage.
- The innovative development of e-Infrastructure services should be protected by involving research and education users in the development of e-Infrastructure services.
- At the same time, non-commercial e-Infrastructure providers should be proactive, rather than simply copying commodity services already offered by commercial providers. Creating alternatives is often important, especially in monopoly situations and in cases of unacceptable license conditions imposed by commercial providers.
- Use of good governance models will be an essential tool for effectively managing the transition to a service-oriented environment. Fair and transparent business models are to be introduced in order to increase integration and sustainability of e-Infrastructure

services and to guarantee a fair distinction between commercial and non-commercial services.
- Development and exploitation of e-Infrastructures should focus on the innovation chain by using transparent PPP approaches, and aim at stimulating research and innovation.
- Similarities between e-Infrastructure services for e-Science and services required by other sectors, such as government, health, et cetera should be investigated by exchanging experiences and transferring knowledge/expertise from the research sector to other sectors so that the speed of ICT innovation can be enhanced and possible economies of scale exploited.

6.4 Recommendations

- Involve the user communities in the definition and exploitation of e-Infrastructure services.
- Use virtualisation and SOA when developing and introducing new e-Infrastructure services wherever this is efficient. Apply simplified access, transparent service offerings, customised support, standardisation, improved governance models and sustainable business models in the definition and deployment of e-Infrastructure services.
- Promote cooperation with other public sectors in the e-Infrastructure arena, like government and healthcare, to exploit economies of scale and intensify the contribution of research and education e-Infrastructures in facing societal challenges at large.
- Boost innovation by public-private partnership activities through the joint creation of a market for e-Infrastructure resources and services.

7 Data infrastructures

7.1 Policy area and goal

Research is nowadays driven by data that is growing exponentially; sharing, evaluating and re-using it is becoming vital, so that several disciplines can harness it and create new knowledge. However, data is typically stored on local repositories making it hard to unify it so that research breakthroughs can happen more quickly and effectively. Data curation and selective preservation for current and future generations are equally important issues. Researchers and organisations need to move away from silo-based repositories and migrate to an environment that enables collaboration, data sharing and unification.

There is a considerable need to integrate data sources in order to build a sustainable way of providing a good level of information and knowledge. This feature is currently missing from the European e-Infrastructure. Although well-armed with other infrastructure components from networking, grids and HPC, the European e-Infrastructure does not yet have a data infrastructure constituent that would allow the operation and unification of several existing multi-domain data and heterogeneous environments. We do, however, have some prior experience and so the major goal is to unify all existing use cases and provide a global approach for Europe.

7.2 Context: Current practices - achievements and limitations

The vision for the next decade is *"a cost-effective, efficient collaborative data research environment built on an interoperable and sustainable governance model fulfilling user needs across geographical borders and disciplines"* and this is referred to as a Global Data Research Infrastructure (GRDI)[11] or an ecosystem of interoperable GRDIs. This ecosystem of interoperable research data infrastructures will be composed of regional, disciplinary and multidisciplinary elements such as libraries, archives and data centres, offering data services for both primary datasets and publications. The ecosystem will support data-intensive science and research, and stimulate the interaction between all its elements, thus promoting multidisciplinary and interdisciplinary science. The aim of EUDAT consortium (Data Infrastructures for e-Science) is to build a high-quality, cost-efficient and sustainable pan-European data ecosystem, with Europe-wide sharing of resources, services and competence.

11 From GRDI2020 project: A 10-year vision for Global Research Data Infrastructures www.grdi2020.eu

The EGI and DEISA/PRACE are operational e-Infrastructures that already support large-scale pan-European communities. Multi-institutional organisations that know how to deal with federated data resources, provide access in a coordinated manner, and manage policies/procedures process through consensus that span organisations already exist. For example, EGI has experience in dealing with large quantities of data through supporting current heavy users of the infrastructure, such as high energy physics, computational chemistry and life sciences. However, we need to identify the key areas where collaboration is needed to build up the scientific data for the e-Infrastructure.

The creation of such a next generation, European data infrastructure needs to address a series of challenges in order to translate the above vision into a reality and to act as a horizontal layer for multiple disciplines. The massive increases in the quantity of data and its complexity as well as the fragmented landscape are some of the key challenges that make it difficult and costly to collaborate across boundaries and share best practices; however this is necessary if we are to embark upon the greater global societal challenges. We therefore need to urgently chart a course for an effective, open and accessible-to-all research environment, upon which European competitiveness hinges.

The challenges involved in the creation of data infrastructures are not only technical, but also legal, organisational, managerial and social. ***Design, implementation, operations, funding, governance and sustainability*** models need to be defined to promote:

- cooperation between data providers and users:
    - exchanging information for better governance of data gathering and management or to fulfil legal requirements
    - sharing and reusing information to increase research efficiency
- new data management, exchange and protection paradigm/approach based on the following assumptions:
    - the designed data infrastructure is sustainable, robust and scalable
    - appropriate services, mechanisms and features are offered within the infrastructure that extend, improve and facilitate (automate) the data handling, preservation, curation and exploitation leading to:
        - improved service delivery to researchers by facilitating the one-stop shop delivery of data services
        - reduced complexity, e.g. by providing federated access to data
        - improved cross-disciplinary and cross-border cooperation in the scientific environment and in Europe as a whole
        - reduced costs, by exploiting economies of scale, thanks to a critical mass of resources providers, data storage and processing resources and users (communities and individuals)
        - assurance that valuable data will be accessible, protected, preserved and curated over decades according to user needs
- the process of embedding data infrastructure into e-Infrastructure (how chains of computations and data analyses are combined and handled).

The governance and design of the data infrastructure require the specification of all actors and stakeholders involved, business and cost-sharing models, the system functionality, the operational components, and the user engagement. And last but not least, funding needs to be secured to make data infrastructure a reality.

For governance, the obvious issues are whether a ***decentralised or centralised approach*** should be adopted and whether this should be ***top-down*** (service provider driven) or ***bottom-up*** (user-oriented). A combination of approaches is probably best, where service providers drive facilities at the low-level, e.g. bitstream-level data storage, exchange and preservations, and users develop the services at the high level, including data exploitation, content curation, data discovery, mining, combination.

A further question is whether data infrastructures can be ***modelled after the other e-Infrastructure organisational structures*** such as NRENs or NGIs and/or can they be integrated with any of them? One of the vital issues is to what extent and how the cross-European data infrastructure should rely on existing and emerging national and international data infrastructures (NorStore, PLATON etc.). Furthermore, the model for including domain data infrastructures (such as WLCG or EBI) must be elaborated to support multi-domain infrastructures. The strategy accepted here affects the ***sustainability*** and ***persistency***, as an infrastructure of a certain scale and scope may require a decentralised organisational and funding model.

Open access and free movement of knowledge are one of the key policy issues; the latter is already being promoted in Europe[12]. Security-related policy issues involve privacy, protection, confidentiality, ownership aspects, all of which need to be specifically addressed. As part of the Digital Agenda for Europe[13] the European Commission will be requesting data management and open content for all research projects in the next framework programme. This will not be possible without standardisation: homogenisation in naming, formats, interfaces, and query languages. Furthermore, there are ***social issues*** that need to be tackled: researchers are inclined to build their own (discipline-based) data "silos" to ensure fuller control of their data. Even the use of a common repository does not automatically ensure any sharing of data, not to mention open access to other disciplines. Thus, appropriate incentives need to be cultivated (a combination of carrot and stick actions) to ***develop a sharing mentality*** among the users. A thorough cost-benefit analysis of an open versus a silo-type approach could be beneficial in this context. The added value of the common European infrastructure should be developed and appropriately promoted, in order to encourage the users to change their approach and *invest their effort in adapting to this model*. The data stored and shared in the pan-European infrastructure needs to be interoperable, easily discoverable, and its provenance must be known and trusted. Only after that can the data be effectively used and re-used. In addition, curation and preservation services need to be available and must guarantee a long-term, multi-level, automated data protection. These are the really high-impact changes needed to make the initiative successful and collect the critical mass of users. However, providing the added values in the large-scale European data infrastructure will be a challenge.

Finally, there are ***technical issues*** related to assembling, securing, managing, preserving and making interoperable the huge amount of data that scientists are producing. A fundamental technical problem to solve is how to address the data explosion by assuring the infrastructure's scalability in terms of storage space, number of data objects stored, number of users concurrency accessing the data, and performance of data accessing and handling. Another vital problem is how to address the complexity, i.e. deal with different, domain-specific data organisations, formats, handling policies et cetera. Ultimately, reaching the reliability and robustness, which have a specific meaning in the context of data exchange, sharing and long-term preservation, in the geographically distributed and complex infrastructure is a big technical challenge.

7.3 Proposed approach

The above analysis reveals just how complex data infrastructures are. Therefore, ***priorities need to be set*** so that a feasible plan can be realised for a European data infrastructure operating in a sustainable manner.

12 EC Open Access pilot in FP7 http://ec.europa.eu/research/participants/portal/ShowDoc/Extensions+Repository/General+Documentation/Guidance+documents/Other+documents/open-access-pilot_en.pdf, ERC guidelines in FP7 and OpenAIRE www.openaire.eu

13 http://ec.europa.eu/information_society/digital-agenda/index_en.htm

A strategy for the creation of such a data infrastructure is becoming urgent. This European infrastructure needs to be coordinated with other data infrastructures around the world, and be interoperable so that users can work on a global network of research data infrastructures. A possible strategy is the **step-by-step approach** in which the most important and vital issues are addressed first, including interoperability at the system and service levels, efficiency of data exchange, data preservation (including e.g. physical bitstream replication) and data access control, while leaving more complex issues such as addressing high demands against security, privacy and legal issues for future stages. Such an evolution in the infrastructure could start by approaching the communities that accepted the open-access approach (e.g. the digital libraries), and gaining their trust by offering the reliable services, while at the same time developing the solutions for dealing with more complex models, including privacy guarantees, legal solutions for cross-border data exchange et cetera. ***From an overall perspective, a roadmap is essential for analysing which steps are required next***[14] [15] and for transforming the European data infrastructure vision into a reality.

The strategy, while being gradual, should also be **multi-level** by spanning all critical stakeholders, including resources and service providers, infrastructure owners as well as potential end users of the services. The European data infrastructure requires data storage and processing resources (storage devices and servers) as well as the network bandwidth for effective data access, exchange and replication. A close collaboration should therefore exist between the infrastructure operator and existing European policy, decision-making and advisory bodies (such as e-IRG, DMTF, TERENA TF-Storage) and existing network, computation and data storage infrastructures (such as GEANT, PRACE, EGI, WLCG, NorStore, PLATON) as they constitute a natural environment for growing and promoting the ideas, as well as developing and running solutions and sustainable, robust and scalable services.

On the other hand, a close relationship with users' communities and discipline-specific or application-specific service providers may be necessary to realise European data services that address users' real needs and that enable the infrastructure to benefit from the domain-specific knowledge and experience related to data sharing, exploitation and curation. The end-user communities and domain data centres therefore need to be given a clear role in the development and operation of the infrastructure.

Most of the challenges detailed, do not have an easy or a clear answer. However, it is obvious that ***funding*** needs to be secured for these issues to be tackled. In particular, the European Commission should impose the requirement for open science on all of its projects, i.e. open data management, content and tools as a condition for funding.

7.4 Recommendations

- Work out a step-by-step strategy for developing the European data infrastructure gradually, addressing basic issues such as data persistency, accessibility and interoperability first, and leaving complicated issues such as privacy and legal matters (like cross-border exchange of sensitive data) for subsequent stages.
- Implement strategy at different levels, including low-level services such as bitstream data storage, exchange in data infrastructures, content-related curation, preservation and data exploitation services, as well as activities aimed at interoperability and data access federation and openness.
- Involve stakeholders of the data infrastructure including resource providers, existing infrastructures and initiatives and user communities in order to build reliable and robust data services suitable to real needs.

14 From GRDI2020 project: A 10-year vision for Global Research Data Infrastructures www.grdi2020.eu

15 Riding the Wave - How Europe can gain from the rising tide of scientific data: http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

Annex I - Editorial Responsibilities

Foreword: Gudmund Høst (chair) & Leif Laaksonen (outgoing chair)

Introduction: Rosette Vandenbroucke, co-chair and White Paper editor

Terminology: e-IRG Support Programmes (e-IRGSP2 and e-IRGSP3)

Chapter 1, e-Infrastructure governance, legal and financial Issues
Champions: Germany and the Netherlands, Editors: Kees Neggers <kees.neggers@surfnet.nl>, Gabriele von Voigt <vonvoigt@rrzn.uni-hannover.de>, Contributors: Ralph Dieter <Ralph.Dieter@bmbf.bund.de>, Jan Wiebelitz <wiebelitz@dcsec.uni-hannover.de>, Frank van Iersel <frank.vaniersel@xs4all.nl>, Fotis Karayannis <fotis.karayannis@gmail.com>

Chapter 2, Future of research networking
Champion: The Netherlands, Editor: Kees Neggers <kees.neggers@surfnet.nl>, Contributors: Artur Binczewski <artur@man.poznan.pl>, Ivan Maric <ivan.maric@srce.hr>, Norbert Meyer <meyer@man.poznan.pl>, Rosette Vandenbroucke <rosette.vandenbroucke@belnet.be>

Chapter 3, Authentication, authorisation and accounting
Champion: Switzerland, Editor: Fiorenzo Scaroni <scaroni@switch.ch>, Contributors: Diego R. Lopez <diego.lopez@rediris.es>, Christoph Witzig <christoph.witzig@switch.ch>, Mikael Linden <mikael.linden@csc.fi>

Chapter 4, Energy and Green IT
Champion: Poland, Editor: Norbert Meyer <meyer@man.poznan.pl>, Contributors: Patrick Aerts <p.aerts@nwo.nl>, Rene Belso <belso@dcsc.dk>, Sergi Girona <sergi.girona@bsc.es>, Radek Januszewski <radekj@man.poznan.pl>, Mirco Mazzucato <mirco.mazzucato@pd.infn.it>, Tero Tuononen <tero.tuononen@csc.fi>, Dany Vandromme <vandrome@renater.fr>

Chapter 5, Exascale computing and related software
Champion: Spain, Editor: Sergi Girona <sergi.girona@bsc.es>, Contributors: Patrick Aerts <p.aerts@nwo.nl>, Leif Laaksonen <leif.laaksonen@csc.fi>, Panos Louridas <louridas@grnet.gr>, Norbert Meyer meyer@man.poznan.pl>, Gabriele von Voigt <vonvoigt@rrzn.uni-hannover.de>, Jan Wiebelitz <wiebelitz@rvs.uni-hannover.de>, Sebastian von Alfthan

Chapter 6, e-Infrastructure services
Champion: Hungary, Editors: Lajos Balint <lajos.balint@niif.hu>, Imre Szeberenyi <szebi@iit.bme.hu>, Contributors: Gudmund Høst <gho@forskningsradet.no>, Panos Louridas <louridas@grnet.gr>

Chapter 7, Data infrastructures
Champion: Poland, Editor: Norbert Meyer <meyer@man.poznan.pl>, Contributors: Kyriakos Baxevanidis <kyriakos.baxevanidis@ec.europa.eu>, Maciej Brzezniak <maciekb@man.poznan.pl>, Gudmund Høst <gho@forskningsradet.no>, Fotis Karayannis <Fotis.Karayannis@gmail.com>, Leif Laaksonen <leif.laaksonen@csc.fi>, Pirjo-Leena Forsström <pirjo-leena.forsstrom@csc.fi>.

Extensive background information, including known projects and experts, other actors, and some starting points is available upon request from the e-IRG secretariat (secretariat@e-irg.eu)

Annex II - Terminology

Authentication and Authorisation Infrastructure (AAI) refers to the systems used to identify and authorise users of shared resources. Authentication is the process of verifying or disproving a claimed electronic identity; authorisation is the process of deciding if a request to perform an action on a resource shall be granted or not. AAI includes authentication and authorisation services, components for identity and privilege management, and the entities responsible for these services.

Campus network refers to a network, or part thereof, that is located inside a university or research centre campus and is considered the access network for individual researchers and students.

Capability computing refers to serving at one single moment in time a coarse number of specialised computing tasks requiring an extremely powerful and tightly integrated computing system. Capability computing can be also referred to as High-Performance Computing (HPC).

Capacity computing refers to serving an extremely large number of parallel tasks on a large-scale computing infrastructure. Capacity computing can be also referred to as High-Throughput Computing (HTC) or grid computing.

Certification Authorities (CAs) are one the main authentication mechanisms used in grid computing.

Cloud computing (or simply 'Cloud') is an on-demand service offering a large pool of easily usable and accessible virtualised resources (such as hardware, development platforms and/or services) in a pay-per-use model. Clouds are usually offered commercially and currently use proprietary interfaces.

EC refers to the European Commission.

eduGAIN aims at providing the means for achieving interoperation between different Authentication and Authorisation Infrastructures. It enables the sharing of identify data between different federations over existing organisations and policies. It therefore plays the role of a confederation: a federation of federations (see also Federation).

EGI.eu is an organisation established on 8 February 2010 to coordinate and manage the infrastructure (EGI) on behalf of its participants: National Grid Initiatives (NGIs) and European Intergovernmental Research Organisations (EIROs). EGI.eu is a foundation recognised by Dutch law and headquartered in Science Park Amsterdam, the Netherlands with 20 employees.

e-Infrastructure or electronic infrastructure covers ICT-related infrastructure and services, such as networking, computing, data and software components. e-Infrastructure by default refers to research, as the term was introduced by the EC, and can be also described as e-RI (in ESFRI terminology).

EGI (European Grid Infrastructure) is a federation of national and domain specific resource providers coordinated by EGI.eu that seeks to sustainably operate a secure integrated production infrastructure for a multidisciplinary user community across Europe and their international collaborators.

e-Science is the invention and application of ICT-enabled methods to achieve better, faster or more efficient research, innovation, decision support and/or diagnosis in any discipline. It draws on advances in computing science, computation, digital communications, storage and data management.

EUGRIDPMA - European Grid Policy Management Authority is the coordinating body of the national Certification Authorities (CAs) in Europe.

Federation is a group of organisations whose members have agreed to cooperate in a particular area, such as in the operation of an inter-organisational AAI (a Federated AAI or an AAI Federation).

GÉANT is the pan-European communication network dedicated to the research and education community. Together with Europe's national research networks (NRENs), GÉANT connects 40 million users in over 8,000 institutions across 40 countries. GN3 is the latest GÉANT project, coordinated by DANTE and co-funded by the EC.

GLIF, the Global Lambda Integrated Facility, is an international virtual organisation that promotes the paradigm of lambda networking. GLIF provides lambdas internationally as an integrated facility to support data-intensive scientific research, and supports middleware development for lambda networking.

Grid is a system that federates, shares and coordinates distributed resources from different organisations that are not subject to centralised control, using open, general-purpose and in some cases standard protocols and interfaces to deliver non-trivial qualities of service. Grid computing is used by VOs.

HPC - High Performance Computing - See capability computing.

HTC - High Throughput Computing - See capacity computing.

Initiative for Globus in Europe (IGE) - is an initiative to provide continued support for the Globus Toolkit on a European scale. The Globus Toolkit is a grid middleware originating in the US and currently used by major computing e-Infrastructures in the US (Open Science Grid, TeraGrid) and beyond.

IGTF (International Grid Trust Federation) - is a body working to establish common policies and guidelines between members of its Policy Management Authorities (PMAs) in the different regions (EUGRIDPMA is the European PMA).

Intellectual Property Rights (IPR) – refer to the controlled right of use of created items, so that the creator benefits from that use. Intellectual Property is broken down into several types, which apply to different created items: copyright, designs, patents, trademarks, protection from passing off and protection of confidential information

Lightpath usually refers to an optical wavelength signal emitted by a laser inside a fibre optic cable. In fibre optic communications, wavelength-division multiplexing (WDM) is a technology which multiplexes multiple optical signals on a single optical fibre by using different wavelengths (colours) of laser light to carry different signals. This multiplies capacity and enables bidirectional communications over one strand of fibre. This is a form of frequency division multiplexing (FDM) but is commonly called WDM. Modern use of 'lightpaths' generally refers to end-to-end optical paths (using a concatenation of wavelengths) to achieve optical network connectivity between end sites.

NGIs (National Grid Initiatives) - are the entities responsible of procuring and operating the national grid infrastructure (in terms of computers and storage devices) and corresponding services to the research and academic communities. NGIs are the main building blocks of EGI.

NRENs (National Research and Education Networks) - are the entities responsible of procuring and operating the national network and corresponding services dedicated to the research and academic communities. NRENs are the main building blocks of GÉANT.

OpenID is an open, decentralised standard for authenticating users and access control, allowing users to log on to different services using the same digital identity when these services trust the authentication body. OpenID replaces the common login process (which uses a login-name and a password) by allowing a user to log in once and gain access to the resources of multiple software systems.

PRACE (Partnership for Advance Computing in Europe) is a unique persistent pan-European Research Infrastructure for HPC implementing 3-5 petaflop supercomputing systems in Europe. PRACE manages extreme computing power and a selected set of highly specialized services. PRACE is a project co-funded by FP7.

PPP stands for Public Private Partnership.

REFEDS (Research and Education Federations) wish to address the need of existing and emerging e-identity federations operating in the field of education and research in Europe, America and Asia to collaborate on policy issues.

Repository - a storage place for digital resources. Users can easily search, access and use resources collected in a repository via an online network. A digital library is a type of repository (a particular application).

RI is the common abbreviation for Research Infrastructure.

TERENA - The Trans-European Research and Education Networking Association is the association of NRENs. TERENA offers a forum in which collaboration, innovation and knowledge sharing can take place that foster the development of Internet technology, infrastructure and services to be used by the research and education community.

VOs (Virtual Organisations) are the structures implemented in the e-Infrastructure for the on-line access to resources that span multiple administrative and geographical boundaries. Often used in grid computing.

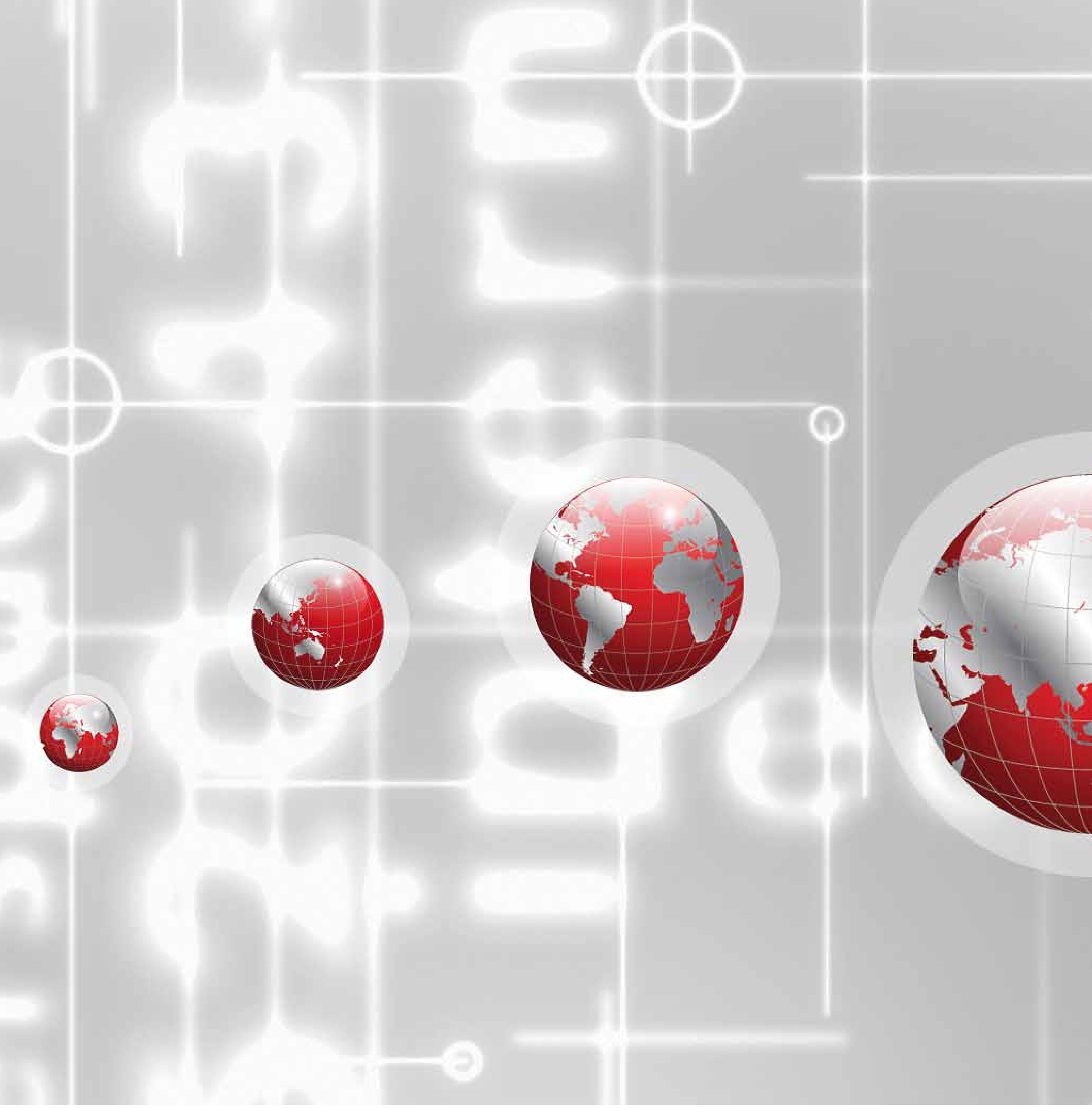
VRCs (Virtual Research Communities) are collaborations of 'like-minded' individuals that are grouped by discipline or computational model that can span multiple VOs. Effectively, they are the human representation of a community that complements the VO representation within the e-Infrastructure.

Virtualisation refers to decoupling of the resource providing a service from the hardware that service is running on. It provides an additional layer that hides the complexity of underlying technology and devices from users.

White Paper - an authoritative report that often addresses problems and how to solve them. White Papers are used to educate readers and to help people make decisions.

Web2.0 is a term commonly associated with web applications that facilitate interactive information sharing, interoperability, user-centred design, and collaboration on the World Wide Web. Examples of Web 2.0 include web-based communities, hosted services, web applications, social-networking sites, video-sharing sites, wikis, blogs, mashups, and folksonomies. A Web 2.0 site allows its users to interact with other users or to change website content, in contrast to non-interactive websites where users are limited to the passive viewing of information that is provided to them.

X.509 is an ITU-T standard for authentication. It refers to a public key infrastructure (PKI) for single sign-on (SSO) and Privilege Management Infrastructure (PMI). X.509 specifies, amongst other things, standard formats for public key certificates, certificate revocation lists, attribute certificates, and a certification path validation algorithm.



e-IRG secretariat
P.O.Box 93575, NL-2509 AN The Hague, The Netherlands
Phone: +31 (0)70 344 0526
secretariat@e-irg.eu

Visiting address
Anna van Saksenlaan 51
NL-2593 HW The Hague
The Netherlands