

COMMUNICATION FROM THE EUROPEAN COMMISSION

ICT INFRASTRUCTURES FOR e-SCIENCE



Brussels, 5.3.2009
COM(2009) 108 final



European Commission
Information Society and Media

LEGAL NOTICE

Neither the European Commission nor any person acting on its behalf is responsible for the use which might be made of the information contained in the present publication.

The European Commission is not responsible for the external web sites referred to in the present publication.

The views expressed in this publication are those of the authors and do not necessarily reflect the official European Commission's view on the subject.

Publications Office of the European Union - Luxembourg, 2009

ISBN 978-92-79-11283-6

doi 10.2759/41669

© European Communities, May 2009

Reproduction is authorised provided the source is acknowledged.

Foreword

We are confronted with an unprecedented global economic downturn which is dominating the political agendas. At the same time, we have to respond to global challenges like climate change, preserving biodiversity, sustainable energy and public health.

Of course, one can be worried about these threats and challenges. But wouldn't it be better to turn them into opportunities?

In my view, the European Union has all the potential to react to the downturn by intensifying further its effort on making the most of Europe's skills and innovative power. We need to strive for strengthening further our leadership in scientific research and reaffirm Europe's role as a competitive global partner and a hub of excellence.

This effort will, in particular, require Information and Communication Technologies infrastructures (also known as e-Infrastructures) to bring scientists together, whatever their geographical location, and give them access to top-of-the-line computational and data resources. This is what e-Infrastructures, such as Europe's world-leading GÉANT network that already serves 30 million researchers, are about: enabling collaborations, sharing resources, providing access to information. They are essential to promote innovation and support the scientific discoveries that will help the EU on the road to economic recovery.

This is why the European Commission has adopted a Communication ICT Infrastructures for e-Science. I invite the European Parliament, the 27 Member States and researchers and industrial stakeholders across Europe to use it to stimulate their reflections on our future policy priorities.



A handwritten signature in black ink, consisting of a large, stylized 'V' followed by a horizontal line and a small flourish.

Viviane Reding

European Commissioner for
Information Society and Media

**COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN
PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL
COMMITTEE AND THE COMMITTEE OF THE REGIONS**

ICT INFRASTRUCTURES FOR e-SCIENCE

1. INTRODUCTION

1.1. Objective of the Communication

This Communication *highlights* the strategic role of ICT¹ infrastructures as a crucial asset underpinning European research and innovation policies, and *calls* on Member States and the scientific communities, in cooperation with the European Commission, for a reinforced and coordinated effort to foster world-class ICT infrastructures, also known as *e-Infrastructures*, to pave the way for the scientific discoveries of the 21st century.

1.2. Background to e-Infrastructures

Innovation, the foundation for economic development, depends on rapid scientific advances. Science in turn has become increasingly based on open, cross-border collaboration between researchers across the world. In addition, it is making intensive use of high-capacity computing to model complex systems and to process experimental results.

The emergence of new research methods that exploit advanced computational resources, data collections and scientific instruments, in other words *e-Science*, promises to revolutionise the scientific discovery process, as did the ‘Scientific Renaissance’² in laying the basis for modern science. It is crucial for Europe to embrace this underlying paradigm shift in order to retain its competitive edge and to respond to society’s expectations.

To facilitate a rapid transition to e-Science, the European Commission and Member States have made significant investments in *e-Infrastructures*, including the pan-European research network GÉANT³, e-Science grids, data infrastructures and supercomputing.

Striving for world leadership in e-Science, establishing e-Infrastructures as a sustainable utility and exploiting them to promote innovation are the three vectors of a renewed European strategy to support the ground-breaking science of 2020 and beyond. This strategy calls for a major step forward in terms of the type and intensity of investment, better linking of research and innovation policies and coordination of national and Community strategies.

1.3. e-Infrastructures and the policy context

The Competitiveness Council⁴ invited Member States to ‘*encourage public and private research institutions to make full use of the emerging distributed forms of research activity (namely e-Science) based upon international research networks made possible by the*

¹ Information and Communication Technologies.

² M. B. Hall, *The scientific renaissance, 1450-1630* ISBN 0486281159.

³ The GÉANT network provides a range of services, not yet commercially available (current speeds ranging from 40 to 100 gigabit/s), to scientists across borders on a permanent basis.

⁴ Competitiveness Council of 22-23 Nov 2007
(www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/intm/97225.pdf).

availability and world-class unique quality of distributed European network infrastructures like GÉANT and e-Science grids', thus reinforcing the need for the coordination of policies.

e-Infrastructures make a major contribution to the objectives of the i2010 strategy⁵ and the vision for the European Research Area (ERA⁶), and have a key role in supporting the deployment of new research facilities, whose development is articulated with ESFRI⁷ and e-IRG⁸ policy groups in a dialogue with Member States.

Renewed emphasis was placed on support for the ERA at the Ljubljana Council meeting⁹, highlighting that a new vision should include the free movement of knowledge ('fifth freedom'), to be facilitated in particular by providing access to world-class research infrastructures and sharing and using knowledge across sectors and borders. The relevance of e-Infrastructures for innovation is recognised by the Aho Report¹⁰ of May 2008.

The report highlights the '*European added value of cross-border infrastructures, interoperability and standards*'. The ERINA¹¹ study confirmed the very high potential of e-Infrastructures beyond the research domain, in helping to smooth the transition of novel technologies and services to the marketplace.

The current financial crisis will put pressure on national budgets. Yet, as the Commission has recently underlined¹², it is now more important than ever to explore 'innovative funding for a wide range of infrastructure projects, including transport, energy and high-technology networks...'.¹³

2. E-INFRASTRUCTURES TRIGGERING A NEW SCIENTIFIC RENAISSANCE

2.1. The e-Science paradigm shift

The adoption of ICT in all phases of the scientific process will enable researchers to engage in cost-efficient collaboration with peers around the globe, while the increasing use of *in-silico*¹³ experimentation opens up new frontiers to man-machine cooperation and scientific discovery.

i2010 (Mid-Term Review, May 2008)

The contribution of ICTs to the Lisbon goals is enhanced by the development of e-Infrastructures (such as GÉANT or e-Science grids) which help build new research environments, driving productivity and increasing the quality of science.

'Aho Report' (May 2008)

"The successful development of e-Infrastructures has demonstrated the importance of European intervention [...] The e-infrastructure approach should be expanded to more application-oriented and user-oriented platforms' [...] they 'are needed in sectors such as e-Government (especially procurement), e-Health (cross-border applications), logistics and transport' [...]"

⁵ EU policy framework for information society and media (www.ec.europa.eu/i2010).

⁶ COM(2007) 161: The European Research Area: New Perspectives.

⁷ European Strategy Forum on Research Infrastructures (www.cordis.europa.eu/esfri).

⁸ e-Infrastructures Reflection Group (www.e-irg.eu).

⁹ Ljubljana Council 2008 (<http://register.consilium.europa.eu/pdf/en/08/st10/st10231.en08.pdf>).

¹⁰ *Aho Report*: 'Information Society Research and Innovation: Delivering results with sustained impact', May 2008 (http://ec.europa.eu/dgs/information_society/evaluation/rtd/fp6_ist_expost/index_en.htm).

¹¹ ERINA study (www.erina-study.eu/homepage.asp).

¹² COM(2008) 800 final: A European Economic Recovery Plan.

¹³ *In silico* is an expression used to mean 'performed on a computer or via computer simulation', coined by analogy with the Latin phrases *in vivo* and *in vitro*, which are used to refer to experiments done in living organisms and outside living organisms, respectively.

This is referred to as the transition from *wet labs*¹⁴ to *virtual research environments* and is the most visible part of the e-Science paradigm shift.

The systematisation of knowledge supported by observation and experimentation was the distinguishing factor of the scientific revolution in the Renaissance.

By scaling experimentation to unprecedented levels to tackle the very small, the very big and the very complex, we are on the verge of a new scientific renaissance.

For example: research on climate change requires complex computer simulations that access and fetch data stored in on-line repositories all over the globe; the challenge of creating individualised models of human beings for targeted healthcare requires increasingly sophisticated modelling and simulation; emulating the behaviour of hazardous phenomena, such as nuclear disasters, pandemics, tsunamis, etc., requires researchers to experiment more and more in virtual worlds rather than in expensive, high-risk real environments.

The ‘virtualisation’ of experiments enables researchers from all around the globe to cooperate and share data using advanced research networks and grid infrastructures.

Speeding up the drug discovery process

During the 2006 avian flu alarm, Asian and European laboratories used 2000 computers in the EGEE¹⁵ grid to analyse 300 000 drug components over 4 weeks — the equivalent of 100 years on a single computer. In-silico drug screening can thus accelerate the discovery of novel drugs while minimising the trial-and-error approach in a laboratory.

Scientific data factories

The Large Hadron Collider at CERN¹⁶ generates 600 million particle collisions per second. It will produce enormous quantities of data to be made available to 7000 physicists in 33 countries via GÉANT and e-Science infrastructures.

What if your peer scientist is a robot?

Robots are beginning to revolutionise laboratory practices and reduce the ‘drudgery’ of manual experiments in ‘wet labs’. They automate processes and speed up the collection and mining of scientific data crucial to understanding complex phenomena and generating new knowledge.

These changes have a transformational effect on scientific disciplines by extending their goals and scope into other domains, leading to cross-disciplinary research.

Remaining competitive in the face of these emerging scientific challenges implies collaboration between research teams and resources across Europe and around the world, the capacity to use and manage exponentially growing sets of data, and the use of high-performance computing environments for modelling and simulation.

It calls for the wide adoption of new research environments powered by advanced ICT to effectively accommodate the unprecedented requirements of today’s scientific communities for connectivity, computing and information access.

2.2. e-Infrastructure for today’s and tomorrow’s e-Science

By facilitating new scientific discoveries and innovation, e-Infrastructures are an essential tool in supporting the ‘Lisbon Strategy’ for sustainable growth and jobs.

The European Commission’s Framework Programme for Research and Technological Development (FP7) has built up a significant momentum for the deployment of

¹⁴ A *wet lab* is a laboratory equipped with appropriate plumbing, ventilation, and equipment to permit hands-on scientific research.

¹⁵ EGEE (Enabling Grids for E-science, www.eu-egee.org).

¹⁶ CERN (European Organisation for Nuclear Research)

e-Infrastructures, not only to strengthen scientific excellence, but also to promote innovation and industrial competitiveness.

While world leadership has been achieved in GÉANT and e-Science grids, more needs to be done to ensure Europe's position in supercomputing and to ensure a coherent approach to scientific data access and preservation.

The exponential growth of hardware performance (computation capacity doubling every 18 months, storage every 12 months, and network speed every 9 months¹⁷) and scientific demands (reaching the *exa*-scale¹⁸ level) poses new requirements and challenges for designing the e-Infrastructures of 2020.

Simulation in large-scale engineering

Computer simulation is key for modern engineering. The production of complex artefacts such as aircraft, cars or personal appliances relies on complex modelling and simulation, and the cooperation of researchers and engineers.

E-Infrastructures need to include a richer set of functionalities, such as new generations of system and application software, virtual machines, service delivery platforms, visualisation tools, semantic-based search engines, etc., in order to support multi-disciplinary teams in transforming bits, bytes and flops¹⁹ into scientific discoveries and complex engineering.

There is both a need and an opportunity to further develop e-Infrastructure as a strategic platform underpinning European scientific and innovation leadership. This calls for a renewed effort from the Member States, the European Commission and the scientific communities to boost investment in e-Infrastructures and to ensure the proper coordination and alignment of national and Community strategies.

2.3. A renewed strategy

Research in 2020 cannot be imagined without the intensive use of sophisticated e-Infrastructures, so Europe needs to commit to a renewed strategy to tackle the associated challenges and priorities. Three interrelated vectors are key to such a strategy: e-Science, e-Infrastructures and innovation.

- The first vector calls for Europe to become a hub of excellence for e-Science, exploiting multi-disciplinarity and global collaboration to combine complementary skills and resources in making use of computationally intensive simulations. Europe therefore needs to reinforce its research capacity in high-performance computing.
- The second vector of the strategy aims to consolidate e-Infrastructures as persistent research platforms to ensure 'research continuity'. The focus is on the delivery of production-quality services 24 hours a day, 7 days a week, and on the long-term sustainability of e-Infrastructures, which requires the coordination of efforts at national and EU level and the adoption of adequate governance models.
- The third vector focuses on the innovation potential of e-Infrastructures. The transfer of expertise to areas beyond science (e.g. e-Health, e-Government, e-Learning) and the use of e-Infrastructures as cost-efficient platforms for large-scale technological experimentation (e.g. Future Internet, massively parallel software, Living Labs) are different dimensions to be explored.

¹⁷ Commonly accepted laws governing the evolution of technology: Moore's and Gilder's.

¹⁸ Exa-scale (1 exa = 1000 peta = 1000000 tera) computing programmes targeting 2020 are emerging in Japan and the USA.

¹⁹ Flops or FLOPs — FLOating point Operations Per second — unit of measure of the performance of computers.

This strategy will be put into effect by a number of concrete actions targeting the different structural domains of e-Infrastructures. Its successful implementation requires the coordination of efforts and reinforced commitment on the part of national and EU funding authorities.

3. EUROPE LEADING THE WAY

3.1. e-Infrastructures today

e-Infrastructures are currently structured into five intertwined domains, together providing a variety of functions and services:

- **GÉANT** is the world's largest multi-gigabit communication network dedicated to research and education. In Europe, GÉANT already serves around 4000 universities and research centres and connects 34 National Research and Education Networks (NRENs). It is linked to similar networks world-wide, thus forming a single global research network (Balkans, the Black Sea and the Mediterranean regions, along with Asia, Southern Africa and Latin America). GÉANT's leading position has been achieved thanks to a consolidated governance model in which NRENs ensure the required deployment at national level and collectively coordinate the implementation of the pan-European network through the alignment of strategic and technological options as well as the pooling of financial resources at national and European level.

What is e-Infrastructure

e-Infrastructure is 'an environment where research resources (hardware, software and content) can be readily shared and accessed wherever this is necessary to promote better and more effective research'.

Such an environment integrates networks, grids and middleware, computational resources, experimental workbenches, data repositories, tools and instruments, and the operational support for global virtual research collaboration.

What is a grid?

A grid is a service for sharing computer power and data storage capacity over the Internet. It goes well beyond the connection between computers and ultimately aims to turn the global network of computers into a vast computational resource for large-scale computer- and data-intensive applications.

- **e-Science Grids** have emerged to respond to the requirements of the most demanding scientific disciplines (e.g. high-energy physics, bioinformatics) to share and combine the power of computers and sophisticated, often unique scientific instruments. With support from the EU's Framework Programmes, Europe now hosts the largest multi-science grids. EGEE today operates a multi-disciplinary grid with over 80 000 computers on 300 sites in 50 countries worldwide, used by several thousand researchers. The DEISA project²⁰ provides a persistent, production-quality, supercomputing environment across Europe, interlinking the 11 most powerful supercomputers in the continent.
- The **scientific data** domain aims to tackle the accelerated and uncontrolled proliferation of data, which, if left unmanaged, could undermine the efficiency of the scientific discovery process²¹. It is therefore crucial to develop new tools and methods to ensure the availability, treatment and preservation of unprecedented quantities of data. The landscape

²⁰ DEISA (Distributed European Infrastructure for Supercomputing Applications, www.deisa.eu).

²¹ COM(2007) 56: Scientific Information in the Digital Age.

of data repositories across Europe is fairly heterogeneous, but there is a solid basis to develop a coherent strategy to overcome the fragmentation and enable research communities to better manage, use, share and preserve data. European-funded projects in the field of scientific data infrastructures share a common vision: any form of scientific content resource (scientific reports, research articles, experimental or observational data, rich media, etc.) should be easily accessible, as a knowledge-sharing platform, through user-friendly e-Infrastructure services.

- **Supercomputing** e-Infrastructures address the data-intensive and complex challenges of providing modern science with the new computing and simulation capabilities it needs. The strategic interest of the Member States and the research community in European high-performance computing and simulation services has led to the creation of a new e-Infrastructure, PRACE²², supported by the ‘Capacities’ programme of the Seventh Research Framework Programme.
- **Global Virtual Research Communities**, anticipating the advent of research 2.0²³ paradigms, have opened new perspectives for cross-border multi-disciplinary collaboration among research communities. A cultural change is taking place in the way scientific knowledge is produced and disseminated, leading to the emergence of Global Virtual Research Communities. Europe is already contributing to the innovation of the scientific process by enabling scientific communities to use e-Infrastructures to address research challenges of global relevance.

3.2. e-Infrastructures for 2020 and beyond

Europe’s response to the long-term challenges of e-Science requires a more efficient and coordinated approach to European investment in world-class scientific infrastructures. By providing common responses to different user requirements, e-Infrastructures are crucial in order to foster scientific excellence, promote global scientific partnerships and stimulate the development of high-quality human capital, while ensuring economies of scale. E-Infrastructures are public goods that support education, research and innovation policies. The active involvement of public authorities in setting priorities and strategies is therefore essential.

GÉANT’s unique capability to enable ground-breaking research collaboration through high-speed connectivity and advanced services is one of the most prominent European success stories. For Europe to keep up its proud tradition of innovation and discovery in science beyond 2020, GÉANT needs to build upon its outstanding performance to meet the *exa*-scale dimension and contribute to the design of the Future Internet.

Today, the sustainability of **e-Science grids** depends mostly on the strong demand from scientific user communities working together in projects funded under national and Community programmes. This entails the risk of discontinued operation and is becoming an inhibiting factor for the full exploitation of grids.

National Grid Initiatives (NGIs)

NGIs are entities with a public mission aiming to integrate funding resources at national level for the provision of grid-based services. They provide a ‘one-stop-shop’ for a number of common grid-based services for national research communities.

²² PRACE (Partnership for Advanced Computing in Europe, www.prace-project.eu).

²³ Research 2.0 is a term to describe the use of web 2.0 technology to enhance creativity, information sharing and collaboration in research.

Project-based, short technology development cycles may undermine the interoperability of grid infrastructures, thus hindering cross-disciplinary cooperation and economies of scale. The EGEE and DEISA projects have already gone a long way to combining disciplines and coordinating strategies. To ensure long-term sustainability, these endeavours must evolve into truly pan-European organisation models that will open grid e-infrastructures to all scientific disciplines and complement national funding strategies in support of e-Science. Several **National Grid Initiatives** are emerging to respond in a coordinated and cost-effective way to the needs of scientific disciplines for computational resources.

The objective of **scientific data e-Infrastructures** is to develop an ecosystem of European digital repositories, combining and adding value to national and discipline-based repositories to respond to Member State requests to improve access to scientific information.

Data, data and more data...

Bioinformatics repositories are growing in size at an exponential rate. By 2012 the information added every year to a single data repository will reach 4 petabytes/year, equivalent of a 10 kilometre stack of CDs.

The emergence of ‘big data science’ has a global dimension²⁴, as it reflects the increasing value of raw observational and experimental data in virtually all fields of science (humanities, biodiversity, high-energy physics, astronomy, etc.). Europe needs to pay particular attention to the accessibility, quality assurance and preservation of key data collections. For example, European environmental policies are supported by the INSPIRE²⁵ Directive, which is intended to bring about a European spatial information infrastructure to deliver integrated spatial information services. In a heterogeneous digital data landscape, where it is estimated that only 28% of research output is managed in digital repositories²⁶, a new strategy for the management of scientific information and associated policies needs to be developed, based on the path-finding activities of key research stakeholders (e.g. EMBL, ESA, ECMWF, CERN²⁷) as well as academic institutions and libraries.

Supercomputing has been identified as a key priority for boosting Europe’s scientific performance. This requires a new strategy for industrial involvement and coordination among funding authorities²⁸. By addressing strategic, policy, technical, financial and governance issues related to supercomputing, PRACE is building important momentum in mobilising significant national funds to deploy an ecosystem of *peta*-scale machines in Europe, aiming for *exa*-scale performance by 2020.

To efficiently support e-Science and gain a lead in **Global Virtual Research Communities**, Europe needs to continue developing world-class e-Infrastructures able to support new ‘participative’ paradigms. This provides a unique opportunity to strengthen the role of European research in the evolving global context.

However, to fully exploit the potential of global scientific collaboration, a number of issues need to be addressed. These concern the clash of cultures between different disciplines, the need to rethink organisational models, and the setting up of quality assurance mechanisms and business models.

²⁴ US National Science Foundation DataNet program (<http://www.nsf.gov/pubs/2008/nsf08021/nsf08021.jsp>).

²⁵ Directive 2007/2/EC: Infrastructure for Spatial Information in the European Community.

²⁶ ‘Investigative Study of Standards for Digital Repositories and Related Services’ DRIVER (<http://dare.uva.nl/document/93727>).

²⁷ EMBL (European Molecular Biology Laboratory), ESA (European Space Agency), ECMWF (European Centre for Medium-Range Weather Forecasts), CERN (European Organisation for Nuclear Research).

²⁸ Europe has been under-represented in the world top ranking lists tracking trends in high-performance computing (<http://www.top500.org/>).

New strategies for the technological development of e-Infrastructures are also fundamental to ensure ‘future-proof’ solutions, based on open standards, which can be maintained and further improved in the long run and add value to the investment in research facilities, large and/or unique instruments, etc.

4. EUROPEAN ACTIONS

The successful implementation of a renewed strategy depends on putting into effect a series of concrete actions that target the different domains of European e-Infrastructures and on establishing synergies between them.

4.1. Consolidating the world leadership of GÉANT

GÉANT, in close articulation with NRENs, needs to continue providing persistent top-of-the-range connectivity with much higher levels of performance to researchers, educators and students in order to lower access barriers to distributed resources and instrumentation. It needs to reinforce its global perspective, encompassing both advanced and developing regions²⁹. GÉANT also needs to integrate the latest technological trends in networking and support experimentation in new paradigms that will lead to the Internet of the Future³⁰.

Member States are invited to reinforce the coordination of national and European policies in the area of Research and Education Networks.

Member States and research communities are invited to support and use GÉANT as an experimental platform leading towards the Internet of the Future.

The Commission, through FP7 and international cooperation, will continue to provide steady support to GÉANT to reinforce its capacity and global perspective.

4.2. Structuring the e-Science grid landscape

Future European e-Science grids should continue to build upon the success of current initiatives, driven by the common needs of different scientific disciplines, and to seek uptake by industry.

However, to reinforce long-term sustainability, governance models need to evolve towards a European Grid Initiative (EGI) building upon the emerging National Grid Initiatives (NGIs).

Member States are invited to consolidate and further develop National Grid Initiatives (NGIs) as a basis for a renewed European strategy.

The Commission will support the transition to new governance models for European e-Science grids as well as their efficient deployment to serve a wide range of research fields, ensuring the technological interoperability of global grids.

4.3. Enhancing access to scientific information

European and National e-Infrastructures need to address the emerging challenge of data centric science. To achieve this, Europe needs to deploy a coherent and managed eco-system of repositories of scientific information. Europe needs to define consistent policies to enhance access to scientific information (e.g. in line with the indications of ESFRI position paper on

²⁹ Building on initiatives such as ALICE (<http://alice.dante.net>), EUMEDconnect (www.eumedconnect.net), TEIN2 (www.tein2.net) supported by DG RELEX, DEV and AIDCO.

³⁰ Supporting initiatives like FIRE (Future Internet Research & Experimentation): (<http://cordis.europa.eu/fp7/ict/fire/>).

scientific data, the Communication on scientific information in the digital age: access, dissemination and preservation³¹ and the Open Access pilot in FP7³² launched in 2008).

Member States and scientific communities are invited to step up investment in scientific data infrastructures and promote the sharing of best practices.

The Commission will reinforce the catalytic investment under FP7 in scientific data infrastructure, to support accessibility and preservation policies.

4.4. Building the new generation of supercomputing facilities

In line with the ESFRI roadmap³³, Europe needs to deploy a new eco-system of computational resources so as to achieve peta-flop performance by 2010 and move towards *exa*-scale computing in 2020. This requires a particular focus on developing and upgrading software and simulation models to exploit the power of the new generations of supercomputers, and calls for reinforced research and development on enabling hardware and software technologies both upstream and downstream along the value chain, including advanced components and systems, system and application software, modelling and simulation.

To build, manage and exploit this new research capability, Europe needs to develop new organisational structures, building on the pioneering work of PRACE. Furthermore, the opportunities offered by public-private partnerships and pre-commercial procurement³⁴ need to be exploited to leverage investment in this strategic field.

To this end, European investment in supercomputing should have a clear industrial impact.

Member States are invited to scale up and pool investment in support of PRACE as well as in related research areas in full articulation with the Commission.

The Commission will launch actions to define and support an ambitious European strategic agenda for supercomputing, ranging from components and systems to the required software and services.

4.5. Hosting global virtual research communities

Europe needs to exploit e-Infrastructures to reap the high innovation potential of multi-disciplinary research and to help its researchers exploit the benefits. It also needs to ensure that scientific disciplines are structured and organised so that they can fully benefit from the services provided by e-Infrastructures. This requires enhanced training efforts to ensure that researchers can make optimal use of e-Infrastructures.

Member States and the European Commission need to ensure that future investment in research facilities is designed to fully exploit e-Infrastructures.

Member States and research communities are invited to embrace the e-Science paradigm by continuing to exploit the benefits of e-Infrastructures.

The Commission will reinforce its integrating activities under FP7 to promote the emergence of stronger European virtual research communities and to encourage them to share best practices, software and data.

³¹ COM(2007)56: Communication on scientific information in the digital age: access, dissemination and preservation.

³² http://ec.europa.eu/research/science-society/open_access.

³³ The ESFRI roadmap identifies new research infrastructure to meet the long-term needs of European research communities (www.cordis.europa.eu/esfri/roadmap.htm).

³⁴ COM(2007) 799: Pre-commercial Procurement: Driving innovation to ensure sustainable high-quality public services in Europe.

5. CONCLUSIONS

Support for research and innovation policies is crucial for Europe to cope with the huge challenges in the 10 to 15 years ahead. Science will experience major changes in the way it is performed. Researchers will be facing unprecedented levels of complexity in tackling scientific challenges with a global societal impact. Bringing together knowledge from different fields of science will be essential.

E-Infrastructures provide the underlying platforms for computationally intensive applications that enable collaboration combining knowledge from different fields of science. New forms of organisation - including virtual organisations across the globe - will emerge from the use of highly distributed network environments such as GÉANT.

The reinforced and coordinated efforts of the Member States, the European Commission and the scientific communities concerned will accelerate the pace of deployment of e-Infrastructures in order to increase their capacity and functionality by several orders of magnitude.

The renewed strategy to achieve leadership in e-Science, develop world-class e-Infrastructures and exploit research innovation potential is essential to position Europe as a hub of scientific excellence and a truly global scientific partner.

EXAMPLES OF e-INFRASTRUCTURES PROJECTS

• Linking the ideas at the speed of light: GÉANT

GÉANT2, the world's largest multi-gigabit communication network dedicated to research and education, offers the scientists the possibility to store, transmit and analyse more data than ever before.



GN2 (www.geant2.net)



With the EU's commitment to global outreach and vision in creating a single global research network for international collaboration, GN2, the e-Infrastructure project responsible for operating GÉANT2, the high bandwidth, pan-European network, has played a key role in shaping global research and education over the past four years. The GN2 project delivers high bandwidth network performance through the world-leading GEANT2 network and a range of associated multi-domain services to optimise the user experience.

By connecting researchers across Europe, through the national research and education networks (NRENs), GÉANT2 enables the best minds across Europe and the world to together pursue collaborative and ground-breaking research activities that were previously impossible or inefficient due to limited and poor quality data connections.

Together with the NRENs it connects, GÉANT2 provides the fastest connection speeds and widest geographical reach of any network of its kind in the world, with outstanding levels of reliability. GÉANT2 has also extended its reach and has narrowed the global digital divide with its links to other EU-funded initiatives in developing world regions. As a world-leading research network, years ahead of those provided by commercial counterparts, GÉANT2 provides enhanced bandwidth, point-to-point links and a range of multi-domain services to optimise the user experience and to help keep researchers at the forefront of their respective fields of research, retaining and developing academic expertise in Europe and ultimately boosting Europe's competitiveness.

GÉANT2 replaced GÉANT in mid 2004 and GÉANT3 will be launched towards the end of 2009. It allows more data to be stored, transmitted and analysed than ever before.

- **Sharing the best scientific resources: e-Science grids**

Grid infrastructures make unprecedented amount of computing power available to scientists by pooling distributed resources across Europe and beyond in a secure and reliable way. Using grids helps researchers to solve complex scientific problems more effectively and efficiently than it was possible before.



EGEE-III (Enabling Grids for E-scienceE, www.eu-egee.org)



EGEE, the world’s largest multidisciplinary grid infrastructure for research project, connects approximately 300 sites in 50 countries to produce a reliable and scalable computing resource for the global science community. By giving researchers in academia and business access to a production level grid infrastructure, with 100.000 processing cores and secure data transfer and storage, independent of their geographic location, EGEE enables research that would not otherwise be possible.

EGEE supports applications from many scientific domains, such as astronomy, astrophysics, life sciences, computational chemistry, earth sciences, high energy physics, finance and fusion. In addition, there are several applications from business sectors running on the EGEE Grid, such as geophysics and the plastics industry. The project develops and deploys the open source, business-friendly middleware distribution “gLite” that virtualises separate grid resources and allows services like data management, complex workflow tools and discipline specific components to run on top of them.

EGEE operates with two clear goals fundamental to European research infrastructures: (1) expanding and optimising its services for researchers who already depend upon the grid for their daily work, while reaching new user communities, and (2) working towards a sustainable, federated European Grid Initiative. By supporting the long-term availability of grid computing, EGEE project strengthens the ability of ICT infrastructures to transform the way science is carried out and thereby increases the impact of the European Research Area.

DEISA-2 (Distributed European Infrastructure for Supercomputing Applications, www.deisa.eu)



DEISA pulls together the power of the 12 largest supercomputers in Europe, providing virtual access to these large high-performance computing facilities to solve scientific problems of huge complexity. The DEISA Consortium has deployed and operated the Distributed European Infrastructure for Supercomputing Applications. It continues to support and further develop the distributed high performance computing (HPC) infrastructure and its services through the DEISA-2 project funded until April 2011. All HPC relevant activities and services such as Application Enabling, Operation, and Technologies are continued and further enhanced.

The DEISA Extreme Computing Initiative (DECI) regularly attracts the most challenging computational science projects Europe-wide. During the last four years, DECI has supported scientists from over 160 universities or research institutes from 15 European countries and four other continents, as such the continuation of DECI is crucial. To further enhance the impact of the European supercomputer infrastructure, the DEISA consortium has extended the single project support to also supporting European science communities globally. Furthermore collaborative activities are carried with other computing infrastructure initiatives world-wide.

The cooperation with the PRACE project which is preparing for the installation of a limited number of leadership-class supercomputers in Europe is of strategic importance.

• Accessing knowledge: Scientific data

By weaving together scientific data repositories, multidisciplinary research communities are given ready access to an unprecedented amount of data. This allows for greater creativity and productivity in research and a broader dissemination of information.



DRIVER-II (Digital Repository Infrastructure Vision for European Research, www.driver-repository.eu)



Considered the largest initiative of its kind in helping to enhance repository development worldwide, DRIVER has a clear vision: “All research institutions in Europe and world wide make all their research publications openly accessible”. This follows the Berlin Declaration, which called in 2003 for “free and unrestricted access to sciences and human knowledge representation worldwide”.

DRIVER is a multi-phase effort whose vision and primary objective is to create a cohesive, robust and flexible, pan-European infrastructure for digital repositories, offering sophisticated services and functionalities for researchers, administrators and the general public.

DRIVER has established a network of relevant experts and Open Access repositories with a fully functional, state-of-the art service, extending the network to a larger confederation of repositories.

DRIVER II is to expand, enrich, and strengthen the results of the initial DRIVER project by providing a production-quality digital repository infrastructure and establishing a European digital repositories confederation. The DRIVER Infrastructure acts as a showcase for repository developments, as a networking tool for the Driver Community and demonstrates a range of end-user services. The DRIVER Confederation comprises data and service providers, the major repository software platforms, representatives of emerging standards, protocols and data models, and NGO’s serving the repository community.



The wide range of scientists trying to mitigate climate change and its impact on society need to access climate data, in particular data from model projections for the next decades and centuries, and to assess the suitability of that data for their purposes. Currently, even if large repositories of climate data are accessible, these repositories are poorly connected and described. The data is therefore difficult to understand, compare and exploit because there is no common information paradigm.

METAFOR's main objective is to define a Common Information Model (CIM) that describes in a standard way climate data and the models that produce the data. METAFOR builds on existing metadata (data describing data) currently used in existing data repositories and addresses issues like metadata fragmentation, gaps, or duplication. In close interaction with related initiatives at the international level, METAFOR is proposing solutions to identify, access and use the climate data in these repositories. METAFOR is optimising the way climate data infrastructures are used to store knowledge, thereby adding value to primary research data for an increasingly wide range of stakeholders.

METAFOR is taking the first step in doing for climate data what search engines have done for the Internet: it is putting users of climate data in touch with the information they need. Simply unlocking existing climate data repositories for new communities will provide new exciting opportunities for scientific research, policy making, and private sector competitiveness.

• Designing future facilities: Novel e-infrastructures

Design Studies: European scientific communities need new e-Infrastructures to remain at the forefront of the global research. The studies for design and creation of new infrastructures of pan-European dimension will pave the way to unique services to scientists, thus enlarging the communities and structuring the European Research Area.



EGI_DS (The European Grid Initiative (EGI) Design Study, <http://web.eu-egi.eu>)



The European Grid Initiative (EGI) aims to develop a sustainable pan-European grid infrastructure, for optimal sharing of computing and data resources.

The EGI is a partnership between National Grid Initiatives (NGIs) and a coordinating body, the EGI Organization (EGI.org), and will begin full operation in 2010. It stems from a series of recent grid projects, such as the EGEE projects, which have contributed to the current leading position of Europe in e-Infrastructures, through the development of a large-scale production-quality grid for scientists.

The EGI goes a step further by moving from a structure based upon short-term funded projects to a long-term service, with the creation of a new organizational model that will coordinate the existing resources and ensure the development of a sustainable grid infrastructure, enabling Europe to secure its leadership in e-Infrastructures.

The EGI_DS aims to define and implement the new organization's activities, and coordinates the efforts of the participating countries as they move towards the EGI model.

EGI_DS has already published a series of Deliverables describing the individual components of the EGI infrastructure. The full organization's operational and financial framework is defined in the EGI Blueprint, endorsed by the NGI representatives on January 20th, 2009.

Construction of new infrastructures: The aim is to create a new research infrastructure in the area of High Performance Computing (HPC) in order to remain at the forefront of scientific research and to help industry to strengthen its technological know-how. This action implements one of the projects foreseen in the European Strategy Forum on Research Infrastructures (ESFRI) roadmap.

PRACE (Partnership for Advanced Computing in Europe, www.prace-project.eu)



PRACE is preparing the creation of a persistent world-leading high-end computing infrastructure for Europe.

European scientists and engineers will use the services to attain global leadership in research and development and to advance Europe's competitiveness.

Access will be granted through a peer review process. Several world-class centres will provide a variety of architectures to meet the needs of the different scientific and industrial domains and applications.

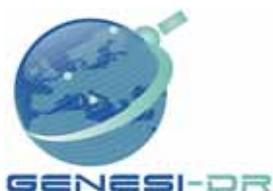
The PRACE project undertakes all legal, administrative and technical work to establish the infrastructure as a single legal entity starting in 2010. Prototypes of potential production systems have been selected based on an in-depth analysis of user requirements and existing mature technological solutions and are being evaluated. A European benchmark suite has been created for the selection of the first petascale production computer systems and researchers are being trained in the use of the future systems.

- **Innovating the scientific process: global virtual research communities**

By their ability to solve complex problems, remove geographic distance and make optimal use of expensive instruments, ICT based e-Infrastructures are changing the way research communities address key societal and economic issues, such as energy, healthcare or climate change.



GENESI-DR (Ground European Network for Earth Science Interoperations-Digital Repositories, <http://www.genesi-dr.eu>)



Earth scientists need access to loads of data in the shortest possible time to perform their research and produce trustworthy results based on all relevant sources of available information. The discovery of these data may be not trivial as data can be distributed across differently organised archives. Accessing them is even more difficult since different archives may use different interfaces. When these problems are solved, earth scientists should know how to interpret these data and have availability of applications, tools and services for their processing and analysis.

GENESI-DR aims to overcome these problems. It provides a framework where data repositories can be easily integrated and users are given a single access point for data discovery and access. GENESI-DR offers as well processing power, and the possibility to incorporate processing services so supporting the users in the analysis of the data.

The GENESI-DR e-Infrastructure is currently being validated against specific applications in the land, atmosphere and marine domain, providing initial processing services to expert users. GENESI-DR is also analysing common approaches to the preservation of data for an effective future access of the Earth Science community knowledge.

EXPReS (Express Production Real-time e-VLBI Service, www.expres-eu.org)



EXPReS makes e-VLBI available to astronomers by connecting radio telescopes around the world to high-speed fiber optic networks.

In e-VLBI, telescopes simultaneously observe a cosmic source. The data is streamed in real-time to a supercomputer which processes the data used to create the highest resolution images of astronomical objects. Historically, VLBI data was stored on hard disks and mailed to a computing facility, requiring months of waiting before results were available.

e-VLBI's real-time capability allows astronomers to react quickly to transient events by conducting nearly immediate follow-on observations.

As of March 2009, EXPReS has made e-VLBI with up to nine simultaneously operated and connected telescopes available to astronomers.

European Commission

ICT INFRASTRUCTURES FOR e-SCIENCE

Luxembourg: Publications Office of the European Union

2009 — 24 pp. — 21 x 29,7 cm

ISBN 978-92-79-11283-6

doi 10.2759/41669

COMMUNICATION FROM THE EUROPEAN COMMISSION

KK-80-09-476-EN-C

ICT INFRASTRUCTURES FOR e-SCIENCE

European Commission

Directorate-General Information Society and Media

GÉANT & e-Infrastructure Unit

http://cordis.europa.eu/fp7/ict/e-infrastructure/home_en.html



doi 10.2759/41669



European Commission
Information Society and Media