e-Infrastructures Opportunities List

e-IRG
e-Infrastructure
Reflection Group

Introduction, purpose of the document

The term e-Infrastructure indicates the new generation of ICT-based Research Infrastructure in Europe. The purpose of this document is to outline the necessary steps Europe should take in the field of e-Infrastructure to realise the vision of the Lisbon agenda and to list the most promising opportunities for targeting EC funding from the cost/benefit point of view. The document was produced by the e-Infrastructure Reflection Group.

In the 20th century Europeans produced the first computer, invented packet switching (the basis for the technology operating the internet) and more recently conceived the world wide web. However, Europe has fallen behind in reaping the benefits of its own innovative force. Now, with grid technology as a strong catalyst and with the parallel deployment of a world leading networking infrastructure Europe has a major chance to regain its former leading position. The current technological leadership must be capitalized on while there is still very much a greenfield situation that allows for "colonisation" of the new research and business opportunities. The e-Infrastructure is seen as the cornerstone of the European Research Area being the "forerunner" infrastructure driving the testing and the first deployment of new innovative technologies. This means that Europe must accept the challenge to develop and build the e-infrastructure required for the information age now. With new contenders – like the fast-growing Asian economies – already looming on the horizon, Europe needs to join the front ranks again if it wishes not to be marginalized.

Key components of the e-Infrastructures are currently: **grid infrastructures**, **networking infrastructures** and **supercomputing infrastructures**. In the next decades science and research will change fundamentally in the way they operate, so the scope of thought should surpass the current situation and needs. In order to support Europe-wide communities that are able to interact in a global environment as equals, it is important to encourage sharing of electronic infrastructure resources as a way to create suitable conditions for cross-disciplinary interaction, providing fertile ground for innovation and eventual industrial exploitation.

This combination of distributed resources is desirable due to several factors:

1. It allows for combining resources that necessarily are in a different place.
2. One can "redistribute" temporary local overcapacity or special capabilities – either selling, bartering or giving it away.
3. It will allow a decentralised set-up to profit from the benefits of centralisation without (many of the) penalties. An example is flexible just-in-time federation of information coming from many data sources being maintained by many independent (even competing) parties.
4. It will greatly simplify the process for people to contribute resources suitable for more general use, and give them direct profit from instantaneous interoperability with a very large pool of other resources.
5. Diversity both in type and age of the available machines on the grid is actually an asset, e.g. for legacy applications that are dependent on certain hardware or software platform.

There is no doubt also that the impact of new infrastructures will be far beyond science, as was witnessed with both the internet and the world wide web. Possible uses of the new infrastructures outside of the research and education communities include commercial services, security and disaster management, digital libraries, entertainment (digital television, rich media distribution, gaming) and e-learning. Enhanced competitiveness in these areas positively impacts the vast majority of the European economy and offers tremendous opportunities. Collaboration and information exchange with industry and the rest of the world is necessarily a part of the entire approach (both in the process as in the resulting environment).

Grid infrastructures for Europe - opportunities

Grid technology is an evolutionary step in the way we can work with computers and everything connected to them. We will use the word Grid to describe a group of machines within a network – e.g. the internet – which use so called «middleware» to allow them to be used for complex combined tasks. The word machine in this context covers literally everything from ordinary computer, large storage facilities, telescopes, satellites, special physics equipment, weather balloons and other large sensor networks to large data streams and data collections, artificial intelligence agents, visualization means – and even people as support organizations that can be shared between institutes. The only requirement is that the machines (from supercomputer to cellphone) can at some point exchange the necessary information through standardized interfaces.

One of the major benefits of the European Science Grid is the possibility and support for more intense collaboration between various research centres and their researchers. The Grid is a major enabler for world-wide partnership and teamwork, catalysing interaction between major and minor European players in the e-Science field and beyond, while providing new ideas and possibilities to apply best practices learned from others to develop services in each research centre. In addition, initiatives for specific application-driven projects can start from Grid collaboration.

Grid Infrastructure-related Opportunities for Europe:

- Populate the Europe Science Grid with pre-funded **shared large scale facilities** (very top end capability computing, extreme storage and data redundancy/back-up) in order to allow researchers from all disciplines to do the most demanding research, improve the scientific competitiveness in Europe and lower the reaction time to actual technical developments.
- Provide a **funding scheme to grid-enable** the wide diversity of special scientific measuring equipment (large and small), sensor networks (such as seismic measurement points, radar equipment, weather stations, air traffic control data), satellites, telescopes and especially databases (from GIS to chemistry, from biochemical to linguistic); Europe can become the data hub of the world, but a well developed internal information ecology is necessary to fulfil these ambitions.
- Provide high-grade **real-time collaboration environments** to research communities working together in grid environments. These include not only video conferencing facilities, but also state-of-the-art problem solving environments and visualisation facilities to allow for multidisciplinary and distributed collaboration.
- Establish **training centers for grid developers** to optimise software for use in grids, to better utilise costly grid resources and enable researchers to realise the full potential of the grid paradigm.
- Create a **European centre of expertise for reprogrammable logic** (e.g. FPGA's) and other dedicated hardware, as these will play an important role in the grid as targeted coprocessors capable of delivering extreme performance in selected tasks.
- Facilitate the creation of competing open (potentially global) **exchanges and/or marketplaces for grid resources**. Thus, the European economy will have incentives for privately-owned resource providers. A European broker platform to rent Grid services, based on a functional accounting layer, is the next step to achieve this. The organisation for allocation of resources should be decentralised and self-organised. In parallel, look into pre-funding models for shared production facilities.
- Support and extend the **federation based authentication and authorisation infrastructure**, enabling consortia and commercial players to provide e-Science resources to virtual communities based on members of accepted identity federations. We need to create an **open trust hub** that allows the influx of large amounts of scientists and SMEs.
- Support the creation of a **mature and open grid protocol stack** for grid-like functionality by supporting the relevant standards bodies GGF, W3C, IETF and WS-I; Europe does not want a commercial monopolist hijacking the grid with proprietary exchange protocols, as this will stifle innovation and competition.
- Provide **reflectors and buffers/storage for useful data streams** from satellites, sensor networks and other data sources currently left unguarded.
- Create a **normalisation institute** that will contribute to standardised access, validated aggregation processes and good interoperability of distributed scientific data.
- Support for digital libraries and other means to take care of **data curation**, **software curation** and **semantic metadata**.
- Create a body overseeing **Quality Assurance** activities allowing to identify different levels of reliability for grid service providers – for instance for user communities with sensitive data or real-time requirements (e.g. medical use).
- Initiate an **organisation to register parameters and formats** in scope and activities similar to the functions that IANA performs for the internet (a bookkeeper for grid parameters). This organisation should operate on a global scale, and not exclude commercial parameters.
- Support the creation of a **European Federation of middleware repositories**.

National grids have their own economic reward in increased competitiveness among the member states, and need not to be supported from a European perspective – except in cases where support is required to help grid-nascent countries participate in pan-European initiatives. Funding policies for Research Infrastructures need also be aligned among European partners to enable better shared use of resources.

Networking infrastructures for Europe - opportunities

The research networking infrastructure forms the basis for not only Grid and distributed supercomputing applications, but it also supports the general purpose scientific communication and collaboration. Europe should continue to play the leading role in building both European-wide and international global connectivity for the Research and Academic community towards the Global Terabit Research & Education Network (GTREN). Due to this diversity of requirements, it is likely that for cost-effectiveness sake, building dedicated network of links between major computing centres would be a good investment. This is due to the very advantageous cost/performance ratio of so called «dark fibre» solutions, when compared to the current standard (routed wide-area network infrastructure).

Networking Infrastructure-related Opportunities for Europe:

- A next generation optical pan-European network platform needs to be deployed **integrating advanced IP-based routed services with lower layer manageable end-to-end optical connections** for the support of e-Science initiatives (e.g. Grids, collaborative research etc.) These parallel networking flows will better serve the diverse requirements of the European Research and Academic community.
- **Global end-to-end connectivity** is a key issue. In order for end-to-end services to work in the hierarchical European backbone consisting of the campus, national, regional and European spans, the European networking infrastructure should universally deploy interoperable protocols. In addition, a human network consisting of the corresponding network administrators should be formalised to exchange views and ideas.
- Europe has heavily invested in the next-generation **IPv6** protocol currently deployed in the pan-European backbone GEANT and the majority of national networks. Europe should keep this leading position and cooperate with other interested regions (e.g. China) to build upon its expertise.
- The future incarnations of the trans-European network GÉANT should proceed in **dynamic provisioning** of production quality seamless connectivity – unless the need for dynamic bandwidth falls short, in which case dynamic bandwidth should give its position to fully meshed-type **permanent** bandwidth.
- There should be support for the Research Networking community to work closely with industry in order to define and develop standard protocols and interfaces for **network control and management planes** coping with a **multi-administrative, multi-technology** and **multi-equipment domain environment**.
- In addition, in developing a European Science Grid the ability to provide **interoperability** of the grid middleware with the above network control and management planes is of key importance.
- The pan-European networking infrastructure should aggressively **cover all the European member states** with world-class connections *before the end of the decade*, and be prepared for new member states entering the European Union. If the necessary amounts of Dark Fibre (DF) are not available in time, action must be taken. This dark fibre vision includes areas of South-Eastern European countries (SEEREN) along with Belarus, the Ukraine and Moldova, in an effort to ease the "digital divide" in Europe.
- Although outside the direct scope of the Research networks further expansion of broadband and optical networking, ultimately covering the **last mile** to doorstep of households, companies, governments, institutes and organisations is a key factor of the development of the e-Infrastructure on the long term. Research networks should be seen as a technology enabler and catalyst for the proliferation of ICT usage in Europe. Solving the many (often legal) issues on a European level concerning the last mile is crucial.
- There are significant opportunities for **mobile networking**. A European organisation should be founded that investigates these opportunities and if necessary subsequently starts up a pan-European mobile networking environment.

The use of dark fibre acquired from the "new market" implements a new model of "ownership" of the networking resource, as it decouples the provision of the network from bandwidth provision - and the related pricing - by traditional carriers. This opens a completely new and innovative perspective for applications (like grids), as the cost of bandwidth is no longer a serious bottleneck for network provision. Longer term strategic issues not directly dependent on current practices and cutting edge technologies must drive e-Infrastructure planning, including research & education networking. The emerging business model should resolve fundamental questions like ownership of infrastructures, sharing policies, foresight of capital investment, consequences of technology driven choices etc.

Supercomputer infrastructures for Europe - opportunities

Supercomputers of the highest performance class are indispensable for the numerically oriented branches of the sciences. Well-known examples are climate and earth system research, water management, fluid dynamics, biophysics, theoretical chemistry, astrophysics, quantum chromodynamics, nanostructure physics and high-energy physics. Both the increasing progress on mathematical models and the complexity of simulations cause the demand of these subject areas for computing cycles to be almost without limits. The same goes for storage needs.

Until now, supercomputers and storage are funded almost purely on a national level: each country in Europe develops and executes its own strategy with respect to supercomputing. This leads to a wide variety of powerful supercomputers, spread over Europe. An example of collaboration on a European level is the European Centre for Medium-Range Weather Forecasts (ECMWF), which is a co-operation of 24 European national meteorological institutes. ECMWF operates a state-of-the-art high-performance computer, which always belongs to the top systems in Europe. Other, more loose collaborations involve DEISA (Distributed European Infrastructure for Supercomputing Applications), which interconnects their individual systems to a distributed terascale supercomputing facility by using grid technologies.

The above model is quite different from the one in the USA, where individual supercomputing centres have unified themselves in two partnerships, launched by the National Science Foundation (NSF): NCSA and NPACI.

Supercomputer Infrastructure-related Opportunities for Europe:

- There is an urgent need for the European Commission and the member states to **review and bridge the gap** between the fragmented European supercomputer facilities and what is available to the rest of the world.
- The ever-increasing computing requirements of numerical science enforce the **establishment of supercomputers of the highest performance class** in Europe, comparable to the leading sites in Japan and the USA, in order to provide an adequate supply of computing power. *Computers of this kind cannot be funded by one country alone; they have to be financed on a European level.*
- Capacity computing typically fills the needs of scientific disciplines which do not need a low-latency, high-bandwidth interconnect architecture between processors. It is not expected that combining these national resources into a few capacity superclusters in Europe will benefit from the economy of scale. **Capability computing**, on the other hand, in general requires access to many processors in parallel, large memory, and low-latency, high-bandwidth interconnects. The availability of significantly larger capability systems than available now is urgently required for increasing the scientific opportunities for many application areas, since more processors and memory will be available in a low-latency, high-bandwidth interconnect architecture. Hence, the availability of a number of capability systems will facilitate economically important fields of science and research in Europe in their efforts of running increasingly accurate simulations.
- Within the class of capability systems, there are still trade-offs to be made. Actual architectures range from purely shared memory vector computers, through clusters of SMP systems (connected by state-of-the-art interconnects, like Infiniband) to NUMA systems, with a large amount of processors with direct access to a single address space. Since specific application areas **run more efficiently on specific capability architectures**, it may be desirable to **represent these types** of systems in a European Supercomputer infrastructure.
- Europe can profit (in efficiency, flexibility, security, availability and scalability) from having a **shared approach to the increasing storage needs** by establishing a distributed shared network of large storage facilities. This will also yield the redundancy required for advanced data recovery.

Annex: Facts and figures

This annex contains current figures and some estimates for both networking infrastructures and supercomputer and storage infrastructures. Since grids are such a new technology and are less clearly definable, no data is available yet.

Networking infrastructures in Europe

It is foreseen that by 2010 all national networks will be provisioned using dark fibre spans supporting gigabit speeds. Thus backbone capacities (e.g. between two main city nodes) will probably start around 10 Gbps for less developed countries and reach 40-100 Gbps (n*10 or n*40 Gbps) or even more in advanced ones. Capacities in the order of 100 Gbps or more will be the case among GÉANT3 central nodes, while multiples of 10 Gbps for peripheral ones. As explained above, these capacities will comprise the routed IP network, and multiple switched point-to-point 1-10 Gbps connections until approximately 2007 moving to 40Gbps and beyond later on. International (global) connectivity requirements will grow towards a Global Terabit Research & Education Network (GTREN).

The European Research Networking community is gradually adopting the wide usage of dark fibre (DF) infrastructure with long leasing terms. Several DF national infrastructures have been deployed, while currently the corresponding integration effort is developed as part of the trans-European GÉANT2 network. GÉANT2 will be based on a wide Dark Fibre (DF) footprint, as a combination of national, cross-border and international fibre paths, complemented with leased wavelength- and SDH-based provisions for regions where DF solutions are currently not available, or are prohibited from a techno-economic viewpoint. It is expected that DF provision will be available in all European regions by 2010, thus the near-term vision of the NREN community is the deployment of ubiquitous e2e services.

Supercomputer infrastructures in Europe

Since the first computer was built over half a century ago Europe has struggled with the downsides of its diversity. The United States of America together with Asia have taken the lead in providing the fastest machinery to their researchers. Europe is – with the exception of a few isolated national initiatives with limited life spans – very much underrepresented in the current top 25 supercomputers of the world. Supercomputers of the highest performance class are indispensable for the numerically oriented branches of the sciences.

At the moment, in 2005, current large supercomputer procurements in Europe tend to require around 50 Tflop/s peak performance for installation in 2005. As a rough average, large European countries are involved in one such procurement every two years. In 2005, a large European country has installed an estimated aggregate 100 Tflop/s, for an aggregate total of around 400 Tflop/s in Europe. **History has shown that the rate at which the peak performance requirements grow, is more than Moore's law**. If one estimates this rate conservatively (disregarding potentially demanding new research areas) at a factor of 2 annually, one would see the the following peak performance requirements:

Year	Peak Performance in Tflop/s
2005	400
2010	12800
2015	409600

With respect to data storage, there is a growing amount of disciplines which develop into data-intensive sciences. Examples are high-energy physics, astronomy, bio-informatics and genomics, and medical imaging. Large scientific instruments, like the Large Hadron Collider at CERN, are expected to generate enormous amounts of data, which need to be stored to allow for detailed analysis. As an example, the amount of data expected in 2008 for LHC is around 25 PetaBytes (PB) disk space, and a similar amount of secondary storage at tape devices. This is the reason that CERN has to distribute the data over various so-called Tier-1 centres in the participating institutes. It is very difficult to include requirements from other data-intensive sciences, and also from "regular" numerical sciences. If we make an attempt to cover these areas on a European level, we could use a factor of 10 more in 2008, which leads to 250 PB storage on disk. With a growth rate of again a factor of 2 annually, this would lead to:

Year	Disk Storage Capacity in PB
2008	250
2010	1000
2015	32000

These are very conservative estimates, based on the European status quo (which is a poor third to both the United States and Asia). **If the ambitions are for Europe to be a world leader in e-science, one would expect these figures to at least triple.**