e-IRG "Blue Paper" 2010

e-IRG “Blue Paper” 2010

Table of Contents

Executive Summary

The e-IRG Blue Paper is a response to a request from ESFRI to examine ways in which ESFRI Research Infrastructures (RI) and their users can engage and exploit common e-Infrastructure services to satisfy their requirements. It provides an assessment of Europe's e-Infrastructure service portfolio, and identifies the opportunities and challenges involved. *This Blue Paper reports on current trends and issues and sets out policy recommendations for several key areas[1]. It is clear that the major requirement is to bring communities together to improve their mutual understanding and collaboration. This is reflected at all levels from technical to policy.*

There are many real benefits that stem from a common e-Infrastructure across the European Research Area (ERA), including:

1. avoiding diverting resources for research into ad-hoc basic ICT service provision
2. avoiding unnecessary duplication in provision of ICT solutions
3. leveraging existing expertise and experience
4. facilitating the integration and interoperation of different communities and RI
5. broadening engagement across Europe and internationally
6. encouraging and supporting open research and innovation

To encourage the exploitation of common e-infrastructure, which is fit-for-purpose, a number of key themes emerge:

There must be **co-evolution of Research Infrastructures, e-Infrastructure and user requirements.** The e-infrastructures must evolve to meet the evolving needs of the RI, and the RI and users must engage actively in this process in order to fully influence and exploit the opportunity. **Collaboration between RI and e-Infrastructures should be actively supported at all levels, to their mutual benefit.** This collaboration can be facilitated by the emerging focus on **service-oriented delivery models** that encourage the creation of well-defined services that can be exploited and developed[2] at the same time. It is clear that **ICT technology will continue to develop rapidly**. Europe must not be parochial in this development; the major research challenges are global in nature. Projects such as the Human Genome Project and the Large Hadron Collider are already global in scale. Climate change must be considered on a global scale. **E-Infrastructure tools and resources must be developed in a global context** to support researchers' global endeavours.

Looking more specifically at the different e-Infrastructure areas:

High quality networks are keys to cutting-edge RI. A sophisticated and essentially distributed governance system has so far supported network innovation, interconnection and interoperability in a multi-domain environment and made the position of European research networks in the world very competitive. **Research networks remain strategic change agents for all RI** and for other e-Infrastructure components. As the demands of leading edge RI grow and change, network policies must continue to focus on innovation, accessibility and ease of use. **New RI are recognised as 'innovation engines' in network evolution and are encouraged to participate in networking coordination bodies to define, test and use new networking services.**

1 including networking, authentication and authorisation, grid, cloud and virtualisation, high-performance computing, remote access and remote instrumentation computing, data infrastructures and persistent storage, and virtual research communities and collaboration

2 well defined services can also be delivered by a number of alternative providers

Authentication and authorisation technologies and policies, which allow users to establish their identity to access a specific e-Infrastructure and to perform certain operations, are fundamental for the successful operation of RI. As research resources become more interconnected, aligning **authentication and authorisation infrastructures** becomes a key issue. This Blue Paper also addresses several topics relevant to this area (such as usage, security, sharing, costs, etc.) and supports the integration of different identity technologies, and the **improvement of national infrastructures and their alignment with agreed standard procedures for identity management and assurance.** It also encourages future pan-European e-Infrastructure and **ESFRI RI projects to define their access control policies and mechanisms** from the beginning, in accordance with the standards and best practices adopted by the research community.

Computing is another area covered by this Blue Paper, in which **stronger collaboration between computing service providers and researchers** is needed. Today, research is increasingly dependent on computational resources, and many scientific domains increasingly require **large scale computing** for their work. This trend is driven by the growing use of computational methods, simulation, and data analysis, and is sustained by falling unit costs in this area. Better collaboration in this area is necessary to develop a more effective understanding of researchers' requirements for HPC and other computational resources and the opportunities such facilities can provide to researchers, and to tackle the so-called scaling issues in HPC software.

The Blue Paper also discusses opportunities arising from the development of **remote access and remote instrumentation,** which are important components of ESFRI RI projects across all thematic areas. Remote use of scientific equipment enables cost-effective sharing and can substantially reduce the human and financial costs of research. The development and spread of remote instrumentation techniques and technologies will also open new opportunities for scientific communities. However, full exploitation of these opportunities requires more effective **integration of data acquisition infrastructures with data processing infrastructures** through standard interfaces to sensor networks and remote instrumentation, the so-called 'Internet of Things'.

Another fundamental shift in the way research is conducted arises from the growing access to and use of data. As RI grow increasingly dependent on digital data, the challenges of **data infrastructures and persistent data storage** become more pertinent. In this Blue Paper, e-IRG supports the identification and promotion of **common (long-term) data-related services across different RI**, and encourages, through policy and facilitation, the development of community practices and standards that assist researchers in exploiting multiple data resources, within and across disciplines.

The development of **Virtual Research Communities** (VRCs) that allow a group of geographically dispersed researchers to work together through the use of information and communications technologies, is the last main area on which the Blue Paper reports. VRCs are also known as Virtual Organisations (VOs) and facilitate integration of distributed research capacities/resources, virtual mobility of physically distant researchers, better access to research results, and regional and global collaboration and partnerships. VRC developments should proceed gradually, starting with domain-specific shared access to distributed resources, expanding to integrate different research activities, and ultimately including wider support infrastructure that integrates separate organisations/finances/management.

The over-riding themes that emerge from the Blue Paper are the increasingly global requirements placed on researchers and RI, and the need to evolve Europe's e-Infrastructure through active engagement of service developers, users and providers, driven by the emerging requirements of the broader research community. Europe's continued international leadership and global contribution depends heavily on our ability to provide researchers with access to the very best scientific and research tools. e-Infrastructure is an essential part of this suite, and together, by optimising and better utilising the shared and distributed e-Infrastructure resources available in the ERA, we can look forward to an era of unsurpassed scientific endeavour and success.

Leif Laaksonen, e-IRG Chair

1 Foreword - Scope of the document

The world is changing. e-Infrastructure is increasingly recognised as an essential facilitator of research and innovation. In its role as a supporter of innovation, ESFRI has long recognised that sharing ICT infrastructure across its projects creates significant opportunities.

The first ESFRI roadmap already identified the importance of e-Infrastructure, highlighting research communication networks, distributed grids, high-performance computing facilities, digital repositories, data storage, data management, data curation, and the software required to operate, integrate and exploit all of these components (referred to as middleware). The emergence of a large number of ESFRI projects (referred to as Research Infrastructures or RI) involving different sets of the above components, albeit in widely differing fields, has further highlighted the potential for a common e-Infrastructure.

Following ESFRI, the above e-Infrastructure taxonomy is used throughout this Blue Paper.

In this paper we focus particularly upon common and shared infrastructure and services. We explore the scope of applicability of current e-Infrastructure provision, the likely future developments, and how these should engage with and be influenced by the emerging landscape of ESFRI projects. In focusing on services, we concentrate on the key requirements, remain technology- and provider-agnostic where possible, and delegate the technical details to appropriate expertise. There are many potential benefits of a common e-Infrastructure across the European Research Area (ERA), including:

- avoiding the diversion of available resources for research to basic ICT service provision
- avoiding duplication in construction and operation of ICT solutions
- leveraging expertise and experience to take advantage of previous work
- facilitating the integration and interoperation of different communities and RI
- broadening engagement across Europe and internationally
- encouraging and supporting open research and innovation

For emerging RI, many, if not all, of the barriers to achieving these benefits are structural and organisational rather than technical. We have only just begun to address these issues at a European level. Key questions for research e-Infrastructure include:

- Responsibilities: Who does what, how should it be paid for, and what role should policy requirements play?
- Balance: Maintaining equilibrium between stability of services and innovation in the e-Infrastructure itself, especially as technology and user requirements evolve
- Context: How do the needs of research and researchers fit with broader ICT and societal trends?

In this paper we aim to:

- summarise current e-Infrastructure offerings and services
- illustrate how these current provisions can meet requirements in some key ESFRI projects[3]
- propose mechanisms for improved engagement between RI and e-Infrastructure
- identify specific issues, gaps and future opportunities for e-Infrastructure in Europe

3 Indicative ESFRI projects and their e-Infrastructure requirements are described in Appendix A

It is important to recognise that many of today's research challenges cannot be solved using standard 'off-the-shelf' solutions. Europe's existing e-Infrastructure is also a research infrastructure. What perhaps sets it apart from other RI is its ability to deliver services across a broad spectrum of RI user communities. However, the ongoing evolution of e-Infrastructure is central to its ultimate success. Close cooperation across RI and e-Infrastructure will drive this evolution to their mutual benefit.

Finally, we must not forget that modern research is increasingly international: not simply pan-European, but global. Efficient e-Infrastructure is vital if Europe and European researchers are to exploit global opportunities, and Europe should play a major role in shaping these global solutions.

2 e-Infrastructure Services to support Research

The following is a brief review of the current landscape of e-Infrastructure components as identified by ESFRI: communication networks, distributed grids and high-performance computing facilities and digital repositories.

Today's researchers are driven by experience with sophisticated consumer devices and modern software (often developed over several decades); they want and expect conceptually simple, user-focused, integrated ICT solutions that provide access to a powerful range of resources while freeing them to concentrate on their own problems. These modern users increasingly demand ease of use, interoperability between different resources, and guaranteed stability or longevity, such that the effort required to adopt new ICT tools is perceived to be worthwhile.

Thus the need for user-oriented solutions is widely recognised. In response, e-Infrastructure providers are increasingly focused on the delivery of user-tailored services that encapsulate all or part of their infrastructure offering to deliver specific abilities to the users. Users with access to simple, reliable and customized services are more likely to focus on their own challenges, and to rely on and trust their service providers' expertise, without needing to fully understand the technical details of the solution. In this way, we can rapidly expand exploitation of innovative infrastructure, reaching far beyond the original developer/early adopter community.

However, the construction of effective, relevant and user-focused services is challenging and depends critically upon understanding user needs. The process is further complicated by

- poor understanding of users' requirements (on the part of both users and providers),
- a lack of common (technical) terminology,
- poor understanding of constraints and opportunities,
- competitive funding environments which often mix research with infrastructure, and
- a degree of mistrust from all parties.

Thus it is essential that policymakers facilitate cooperation between service providers and user communities, encouraging them to work together to develop the mutual understanding, expertise and trust required effectively to deliver sophisticated and powerful services across a wide range of research challenges. Such collaborative development is often described as 'user-driven', but this does not adequately describe the proactive commitment required from all parties.

2.1 e-Infrastructure layers

Network

Networks are indispensable to the development and use of RI. The current Internet is a direct spin-off from research network developments and an essential element of modern society. Still, the demands of the research community are usually ahead of what the 'commodity' Internet can provide.

Research networks have already transformed into 'hybrid' networks, providing high-quality Internet services as well as direct point-to-point connections, such as 'lightpaths', which ensure dedicated data transfer to and from specific locations (for example, from one research lab to another). This ability to dedicate a data path allows providers to guarantee quality of service, even for very large data volumes. Such hybrid infrastructures already form a solid base for many international research collaborations, such as the LHC (Large Hadron Collider) and eVLBI projects, and further promote shared RI use. However, there is room for improvement in functionality, performance and ease of use, and ongoing innovation is a key characteristic of research networks.

From a user perspective, networks should provide a ubiquitous, hassle-free, end-to-end data path for use anytime, anywhere. In practice, this path must cross through the domains of several different network providers. For researchers, these networks can include campus networks (such as those within a university or research centre), National Research and Education Networks (NRENs), and international interconnecting infrastructures.

To facilitate the seamless networking increasingly demanded (and required) by users, networks must interoperate, and thus network providers must work together to develop globally accepted standards, an ongoing process taking place in global forums like IETF, IEEE, ITU, OGF.

In Europe, the 36 NRENs already work together via the GÉANT network infrastructure and its associated technical developments, and throug h the TERENA collaboration. The European Commission facilitates GÉANT via the GN3 project. Beyond Europe, GÉANT also procures international links providing broader international connectivity. DANTE is an 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), which 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.

As e-Infrastructure components such as networking, computing and data become increasingly integrated (forming a seamless service from the users' viewpoint), additional demands will be placed on the network infrastructure. Close interaction between network providers, other e-Infrastructure providers and the user community will be crucial to the timely creation of this new integrated work space.

Computing

Whether using a single computer, large-scale distributed data processing, or the world's largest 'supercomputer', modern researchers invariably require access to computing power. They can choose from a variety of computing infrastructures: their laptop or desktop, local computer services at the department or institute level, as well as national or international computing infrastructure.

While locally controlled computing services are important and more easily tuned to individual needs, national and international computing infrastructures can provide prioritised research projects with access to resources at scales not otherwise achievable. Increasingly, the boundaries between local, national and international resources are blurring as researchers pool their resources to reach the resource scales required for their research.

Although the requirements of innovative research – including large data volumes and complex data processing – should drive development of computing resources (the so-called 'science case'), there is an increasing need to balance this with increased resource accessibility and usage. For example, technical barriers that prevent new users from accessing resources could be lowered, and new and different computing resources could be added to existing resource pools.

In Europe, two international computing infrastructures – the European Grid Initiative (EGI) and the Partnership for Advanced Computing in Europe (PRACE) – have been established to manage the development of European e-Infrastructure:

- EGI is a collaboration between a number of European National Grid Initiatives (NGIs), CERN, and EMBL; EGI is working to deliver standards-based services to support collaborative distributed computing at the local through to international scale. While European grids have historically focused on large-scale data processing (e.g. in support of the LHC, EGI will support many levels of research discipline, ranging from communities of a handful to several thousand researchers. EGI will also focus on working with new user communities to deliver and enhance services tailored to their research needs.

- PRACE is an association of European entities providing National High Performance Computing (HPC) services and is preparing for the creation of a persistent pan-European HPC service. This service will focus on the highest capability machines and will consist of several 'tier-0' centres, providing European researchers with access to cutting-edge capability computers.

Both EGI and PRACE aim to offer sustainable services that will evolve with changing technology and user demand. User engagement with these initiatives is vital to their vision.

Middleware

Within e-Infrastructure there are many types and kinds of middleware. Even the term 'middleware' is not standardised, with its meaning changing depending on the environment in which it is developed and deployed. In a distributed computing system, for example, middleware is defined as the software layer lying between the operating system and the applications. In a broader sense, middleware is computer software that connects components or applications[4]. Middleware is also used to refer to the 'glue' that enables virtualisation technology and services.

In addition, grid middleware is defined as the intermediate software between any local IT resource management system and the applications. It is built by layered interacting software packages that enable the shared usage of various ICT resources from multiple administrative domains for a common goal. The many different groups brought together in this collaborative effort, aiming in general at solving a common problem, are identified as a Virtual Organization (VO).

A central theme of middleware development is the promotion of interoperability and standardisation of networked resources through a common base of protocols and services. Examples of middleware include infrastructures and software services that:

- Authenticate and authorise secure access to resources (compute, storage, data, etc.).
- Manage software life-cycles.
- Support collaboration within members of VOs (e.g. attribution of roles and rights to VO members Instant Messaging, etc.).
- Support open standards to allow the interoperability of different implementations (e.g. middleware repositories and parameter registration).
- Train and support scientists and support personnel.

Middleware – and software in general – lies at the heart of all aspects of e-Infrastructure development and exploitation. While software and middleware services developments should

4 http://en.wikipedia.org/wiki/Middleware

be driven by clear objectives (e.g. user requirements or operational efficiencies), the need for long-term sustainability should not be ignored. Effective processes and procedures are required to improve the way research organisations and in particular the ESFRI RI communities develop and sustain the software they need, thus enabling a clear separation of generic services (for general use with maintenance and operation able to be shared between different e-Infrastructure communities) and specific discipline-oriented solutions.

As a result of a competitive development phase for some RI components, different implementations of similar services with different interfaces and API are available today. As an example, the European Middleware Initiative (EMI) project is expected to have a challenging time ahead in its effort to define and adopt the standard interfaces that will provide users with a unified access environment. New, more general middleware solutions continue to emerge, virtualising computing resources and making them available through different interfaces: local, grid, or cloud. Such middleware support user interactions through traditional local or grid requests, and also via the Open Cloud Computing Interface (OCCI) standard, making it possible to allocate compute, storage and network resources on a pay-as-you-go basis.

Data

Data management – how best to exploit the growing wealth of global data – is one of the biggest challenges we must face. Data and information are growing not only in volume but also in complexity, as researchers increasingly require information derived from multiple sources and formats. Further, user communities are increasingly diverse, and make new demands for data preservation, publication, and citation. As data is created, stored and used in new and often unsystematic ways, we must also deal with the incompatibilities that inevitably come from independent and autonomous creation and organisation of data.

There is a long history of successful management of research data in Europe, with a large number of existing data infrastructures comprising established and growing user communities. These include international research organisations such as ESO, ESA, EBI or CERN. However, most data initiatives address the needs of a specific discipline: very few serve generic multidisciplinary activities. The existing European and global RI have much in common, creating significant potential for synergy despite their many differences. Areas with potential for joint development include: shared data storage, data shipment, data access, preservation, publishing and citation[5].

Recently there have been a number of agreements on data responsibilities and requirements, including the Toronto statement for biological data[6] and the INSPIRE directive[7] to mandate common interchange standards, access rules and metadata standards for publicly funded data. These initiatives will open new opportunities for researchers to exploit data from a wide range of sources. However, they also bring a significant increase in the responsibilities required of researchers themselves to make data available over extended periods.

5 ERC Scientific Council. Statement on Open Access. erc.europa.eu, December 2007 and OECD Follow Up Group on Issues of Access to Publicly Funded Research Data. Promoting Access to Public Research Data for Scientific, Economic, and Social Development. Technical report, OECD, 2003.

6 Toronto International Data Release Workshop Authors. Prepublication data sharing. Nature, 461:168–170, 09 2009.

7 EU Parliament. Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Official Journal of the European Union, 50(L108), April 2007.

2.2 *Towards European e-Infrastructure as a Service*

The special case of research user needs

The computing needs of research user communities are often well ahead of the general ICT-market, and usually impossible to fulfil using generally available 'commodity' solutions in standard configurations. In the past, research user communities have been spurred to develop new ICT components or specific solutions (e.g. the Internet and the World Wide Web), of which they have also been the first customers. An example is the high energy physics community, who built an innovative and powerful distributed computing e-Infrastructure to cope with the data processing needs of the LHC. A key factor in the success of these endeavours was close interaction between users, developers, and service providers.

A distinguishing feature of such research-led e-Infrastructure is its ability to offer innovative services to user communities. These services are often unavailable on the open market and thus can anticipate market trends. A good example is hybrid networking (described above). Another example is the international grid infrastructure, which connects several hundred compute facilities from multiple research entities across the globe to provide a geographic distribution of services and solutions that is difficult to match in the commercial world. The e-Infrastructure is acting both as a new integrating framework or facility (from the e-Infrastructure provider side) and as a new service platform (from the users' side). Constructive and collaborative interaction between these stakeholders is key to the success of the e-Infrastructure, and of its users.

Today's computer infrastructures allow users to access and share distributed HPC and/or High-Throughput Computing (HTC) systems and enormous storage archives. These are used to run models and conduct analyses at previously unprecedented scales; such activities are increasingly important for a knowledge-based society. These solutions offer cost-effective shared use of technology and pervasive accessibility, making them available to virtually every researcher. Using the advanced software services available today, users can automatically find the computing power that best corresponds to their needs, manage their workload, search through data distributed in different sites, handle security and related user authentication and authorisation, provide accounting services and real-time monitoring of their activities, and much more. Thanks to these services, it becomes easy to access and share distributed clusters and archives.

As yet there is no single unified computing system in the sense that there is a single World Wide Web; rather there is a (relatively small) number of grid, HPC and cloud services, each with different interfaces. Global standardisation efforts are slowly improving interoperability between these components.

The move towards service orientation

In all e-Infrastructure domains there is a move towards service-orientation and away from the traditional technology- or product-orientation. This trend is fuelled by:

1) the growing diversity of user groups that increasingly include smaller groups with no dedicated IT support, and
2) sophisticated virtualisation technology that can create a virtual instance of almost all forms of physical resources. Virtualisation allows all or part of an e-infrastructure to be encapsulated and dedicated to the delivery of specific functionality to specific users. This innovative technology allows providers of common infrastructures to offer 'on-demand' provision of fully functional services for specific user groups.

Thanks to virtualisation, e-Infrastructure can now provide on-demand web-based access to shareable services, reproducing exactly the environment required by a specific user. For example, users can request and be provisioned with N computing nodes on operating system X for M days, or a data access service for K days to be shared by the users of the Virtual Organisation Z. Services available in this way include virtual compute and storage resources as well as services that are themselves hosted in the e-Infrastructure, such as web portals, mailing services, databases, VO-user management, and so on.

The functionality of such cloud services can be described in many ways, including:

- *Hardware as a Service* (*HaaS*) *or Infrastructure as a Service (IaaS)*, focusing on making hardware available via the Internet to customers.
- *Platform as a Service (PaaS)*, providing higher-level services including intermediate software components (middleware or APIs) bundled with the offered hardware.
- *Software as a Service* (SaaS), focusing on providing ready-to-use software services via the web.

The 'as a service' notion can be expanded to other resources, such as data. *Data as a Service* (DaaS), for example, focuses on web-based provisioning of access to data available from various sources. The e-Infrastructure itself can become a set of services easily bought and sold via the web.
The adoption of such a service-oriented model – in which every component is offered via the web as a service customised according to user demand – increases the sustainability of e-Infrastructure by expanding the pool of potential users to encompass all of society, and by facilitating innovative solutions not yet available commercially.

Use case examples

The relationship between service provider and service user may vary in a number of ways. For example, a company delivering utility computing may take on a contract to provide a hospital with a secure and archived computing and storage environment for a period of ten years. Another company may offer a banking service by badging and tailoring a full-scale banking service provided by another bank. In these two cases, a significant negotiation takes place to determine and specify the relationship between supplier and consumer; this relationship will often be specified in a legally binding and commercially sensitive Service-Level Agreement (SLA).

At the other end of the scale, an individual researcher can use the service of a particular provider to buy one-off access to a particular application using particular data. In this case, there is no negotiation: the user simply selects one of the standard SLAs presented in the provider's virtual shop window. Thus future provision of e-Infrastructure will incorporate a range of scales.

e-IRG has previously documented a set of recommendations[8] related to the move towards service-oriented e-Infrastructure, covering the main policy issues in the fields of *grid and cloud computing*, *Service-centric e-Infrastructures through virtualisation*, *Remote instrumentation*, and *Sustainability of computing-related e-Infrastructure*.

8 E-IRG White Paper 2009 http://www.e-irg.eu/images/stories/e-irg_white_paper_ii_web1.pdf,
e-IRG Recommendations 2009 http://www.e-irg.eu/images/stories/e-irg_wp_recomm_2009_a5_web.pdf

3 e-Infrastructure Service Areas

In discussion with a large number of projects currently listed on the ESFRI roadmap, it has been useful to describe their e-Infrastructure requirements in terms of a number of technical areas that do not map directly on to the traditional layers outlined above. This mapping has been used by the European e-Infrastructure Forum (EEF)[9] for a more detailed requirements-gathering exercise across the ESFRI projects.

All of the areas described below are interconnected. For example, data services often implicitly require significant computing for data processing, data transport services, and user authentication and management. Initial discussions with the biomedical community have triggered efforts to further refine understanding of e-Infrastructure requirements in the Biomedical Sciences (BMS), led by the ELIXIR project[10]. Such initiatives should be encouraged and should involve both research e-Infrastructure providers and users, to help develop the mutual understanding and trust that are essential to future success.

3.1 Networking

European networks are globally very competitive, using a highly developed but decentralised governance system to sustain network innovation, interconnection and interoperability in a multi-domain environment. Technology in the networking arena is continually changing, driving strategic change in other e-Infrastructure components and RI. The 'death of distance' is a main strategic driver for the way research is done. Repeatedly, increases in the capabilities of networks sparked new applications and unforeseen user needs. To meet the demands of leading edge RI, network policies must focus on innovation to overcome existing shortcomings in ubiquitous accessibility, service orientation, and ease of use. This process will be both technology- and user-driven, and should be supported by a governance model aiming at timely innovation and rapid uptake of new services by user communities.

The main technical drivers for network innovation are:

- the broad uptake of hybrid networking, which is enabling high-quality and high-bandwidth point-to-point connections besides the normal routed Internet services
- emerging technologies for broad and ubiquitous wired and wireless access to networks
- the associated need for advanced security provisions

Networks must meet increasingly heterogeneous user community requirements, such as the demand for service orientation and commodity computing, and the need to address the ongoing data deluge. RI communities are leading edge users and as such ought to play a major role in defining user requirements for the delivery of network services, as the LHC and e-VLBI communities have done in recent years. International RI can contribute by leveraging improved accessibility and end-to-end interworking for users in less favoured regions and in campus networks everywhere.

Any governance model should recognize that users see networking as a global 'end-to-end service' and that the technical reality should remain invisible. This implies that the decentralised structures for management and financing should be balanced by centralised global structures for harmonisation of the network services and appropriate structures for user support and faults restoration. The emerging integration with other e-Infrastructure compo-

9 The EEF is a collaboration of GÉANT, TERENA EGEE, EGI, DEISA and PRACE

10 29-30 March 2010 Workshop on ESFRI BMS RI Co-operation with e-Infrastructures, Hinxton, Cambridge

nents brings additional challenges into the governance system, which is currently distributed over the individual infrastructure components.

Recommendations:

- Recognise new RI as 'innovation engines' in research network evolution, and encourage them to engage with this role by defining, testing and using new networking services.
- Encourage RI to participate in networking coordination bodies to secure an ongoing exchange of information on the development of advanced networking services.
- Encourage advanced users and research network providers to ensure that national and European authorities support appropriate governance and financial models.

3.2 Authentication, Authorisation and Accounting

Authentication allows entities (usually users) to establish their identity within a specific e-Infrastructure. Authorisation establishes the rights of individual users to perform certain operations within that infrastructure, where those rights are decided according to defined resource access policies. In a highly distributed and collaborative environment that crosses multiple administrative domains and national boundaries, distributed authentication must be performed as close to the users as possible (e.g. at a researcher's home institution), while coordinated authorisation should be performed as close to the resources as possible (e.g., at the site of the desired computing resources).

As an e-Infrastructure matures and its user community grows, requirements for aligning authentication and authorisation also grow. 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.

The development and deployment of Authentication and Authorisation Infrastructure (AAI) has taken place in different research and education environments as well as in the private and public sector. 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 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 international resources.

For the last five to ten years, Europe's NRENs, under TERENA's coordination, have operated identity federations, and provided services to a large number of users within the academic and research community. Based on open standards, these national identity federations have focused on providing access to web-based resources (such as data repositories and e-Journals). Within Europe, the GÉANT3 (GN3) project is about to deploy eduGAIN as a service within the GÉANT infrastructure. eduGAIN is intended to connect the different national identity federations into a seamless common identity space supporting real-time access to web resources across Europe.

Third party players outside academia include providers of user-centric identity management models, particularly 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 STORK[11] project).

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

11 The aim of the STORK project is to establish a European eID Interoperability Platform that will allow citizens to establish new e-relations across borders, just by presenting their national eID https://www.eid-stork.eu/)

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 in commercial and other sectors.

Thus 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. Nonetheless, there are different aspects needing special attention, like the alignment of developments with the development of the Internet, and full support for the management of distributed, dynamic VOs.

The following actions will reinforce e-IRG’s commitment to foster the convergence of access control mechanisms for e-Infrastructure.

Recommendations:

- Accelerate the continued integration of different identity technologies, through supporting active collaboration between the IGTF, GÉANT and relevant European and international working groups.
- Continue to improve national infrastructures and their alignment with agreed standard procedures for identity management and assurance.
- 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.

3.3 Grid, Cloud and Virtualisation

Science is increasingly global, and the rise in distributed research teams working with distributed data sources will continue to drive the need for distributed data processing and storage. Large-scale distributed computing is here to stay, whether delivered by grids, clouds, or shared data centres.

To service this need, the research sector has developed and deployed grid services in Europe and around the world, supporting standards-based access to computers, storage, software, data and other non-IT resources, regardless of geographical location, administrative affiliation, and local management tools.

Another approach to distributed computing is that of 'cloud computing'. Relying on 'the cloud' reduces the need for users and centres to invest in their own large-scale computing and storage resources, however, this is a development that must be carefully investigated. There is confusion between users, experts and policymakers as to the definition and potential of current commercial cloud computing in terms of its ability to cost-effectively satisfy the advanced requirements of the research and higher-education sector, with much of the recent press coverage being driven by considerations of commercial competition, proprietary technology, and customer lock-in.

The interests of scientific users can certainly be served by introducing service concepts such as 'clouds' to the current e-Infrastructure (such that users can easily provision the resources and services they require 'on demand'); however, practical steps must be taken to assess the offerings of commercial cloud computing versus the needs of researchers for sustainable and advanced open infrastructures[12]. If e-Infrastructure is to continue to meet the challenging demands of RI communities, there must be continued focus on technical innovation, especially in areas in which a commercial cloud offering does not yet exist. Specifically, there is a need to overcome the existing limitations of multi-domain large-scale distributed data processing, data management and data storage.

The main drivers for future grid and cloud innovation are:

- **Standards** to facilitate interoperability and support freedom in choosing a service provider,
- **Integration** of existing grid layers with **the on-demand delivery model typical of commercial clouds,**
- **Virtualisation and service-orientation** supporting better resource utilization, increased flexibility, and enhanced provision of user-focused environments,
- **Governance** models appropriate to driving open standards-based interoperability and integrated user services, and
- **Finance models** to support delivery of user-focused services leveraging a cost-effective shared infrastructure provision.

Recommendations:

- Promote collaboration among grid and cloud infrastructure providers and users to raise awareness of the range of available technologies and how to best use them.

12 e-IRG White Paper http://www.e-irg.eu/images/stories/e-irg_white_paper_ii_web1.pdf

- Encourage RI to inform NGIs and EGI.eu of their technical requirements and provide feedback on existing and future services, with a focus on requirements and services rather than technologies.
- Support organisational models that encourage the RI community to engage with the management structures of the NGIs, EGI, and related activities such as EMI and IGE

3.4 High Performance Computing

The boundaries between HPC and other ICT categorisations, such as HTC, grid, or capacity computing, are increasingly ill-defined. What is clear, however, is that a range of research domains increasingly require different kinds of supercomputers and related competencies (such as expertise in parallel software development). These requirements are sustained by the falling cost of all but the most demanding systems: last year's world-leading HPC system is tomorrow's institutional resource.

The PRACE initiative coordinates the planned pan-European approach to sustainable services for high-end HPC. While pooling investments into a powerful computing grid may be the most cost-effective route to create a world-class system, experience shows that effective exploitation of top-of-the-range HPC systems requires adequate mid-range provision to support user and software migration onto larger machines. In the case of the PRACE service, this requires significant national and/or regional resources. Currently, such systems cannot be provided by cloud or desktop systems. To leverage the full potential of European researchers, these systems should ideally operate as an integrated system along the lines of those developed by DEISA and EGEE, integrating local, national and European scale resources. As HPC usage spreads beyond traditional computational sciences, it is increasingly important to provide researchers with an incremental path – in terms of systems and support – from the use of desktop computers through to multiprocessor systems and to highly parallel computers as required.

The technical challenges of effectively exploiting the largest HPC systems have often restricted or delayed their wider exploitation. Thus, in addition to providing the largest scale computing systems to tackle the most challenging problems, it is also important to support collaborative development between researchers and service providers, to address the so-called 'scaling issues' in software, and to develop a more effective understanding of the opportunities for exploiting HPC resources.

Recommendations:

- Improve understanding of the specific requirements of the research community (the 'science case'), and the broader economic needs in terms of driving future requirements for the largest HPC systems.
- Support the development of a balanced HPC ecosystem that integrates resources at a range of scales matched to user requirements.
- Promote specific enabling activities, such as scalable software development and user training, to ensure efficient usage of HPC resources.

3.5 *Remote Access and Remote Instrumentation*

Remote, authenticated and authorised access to laboratory equipment is an essential enabler for the ERA. Access to unique and expensive equipment is often a precondition for successful research; however, this kind of equipment is increasingly unavailable locally. Shared access, regardless of researcher and resource location and instead based primarily on (peer-reviewed) research quality, is an essential component of the trust that must be built if we are to deliver the ERA vision of common, shared, research infrastructure.

The remote use of scientific equipment enables cost-effective sharing by substantially reducing the human and financial costs of research. Much research instrumentation is widely distributed or located in remote areas by necessity. In these cases, the ability to remotely access instruments improves efficient use of researchers' time and increases return on investment for large installations (such as many of those in the ESFRI roadmap). Remote instrumentation is an important component of ESFRI RI projects across all thematic areas.

There is significant international activity to standardise interfaces to and integration of modern sensor networks. This standardisation should be encouraged and better integrated with existing activities in networking and data processing. Sensor networks, often leveraging ubiquitous radio frequency identification (RFID) technology, are increasingly network-aware and can be remotely configured, managed and controlled. This in itself creates new challenges for security and large-scale distributed data processing.

The development and spread of remote instrumentation techniques and technologies is opening new opportunities for scientific communities. However, to fully exploit these opportunities we need more effective integration of data acquisition infrastructures with data processing infrastructures via appropriate standard interfaces to sensor networks and remote instrumentation, the so-called 'Internet of Things'. Standards will thus be of key importance in supporting increased exploitation of scientific equipment by diverse user groups as well as interoperability with cooperating infrastructures.

Recommendations:

- Encourage RI and sensor networks to connect to the wider networked world using standard interfaces, either directly, or indirectly, through the existing e-Infrastructure.
- Formalise the responsibility of European RI to support remote users.
- Champion the user- and broader societal perspective for development and deployment of new RI and sensor networks.

3.6 Data infrastructures and persistent storage

Europe is estimated to be home to 250–300 different RI highly dependent on digital data. For several decades, research organizations including CERN, EBI, ESA, ESO, and more recently ESS, have been running advanced data services, many of which are international in scale. Thus Europe has accrued significant expertise in providing data-related services, primarily focused on the support of discipline-specific communities. In addition, national data services and IT centres across Europe host data infrastructures and provide services to individual research groups.

However, initiatives to foster and exploit existing expertise to develop generic services of value to existing providers, or to develop generic international services, do not generally exist at a European level.

The worst possible scenario is that all new RI build their own data infrastructure resulting in 250–300 data silos of limited compatibility and interoperability. Trying to bring the deployed solutions closer together after the fact will be difficult, if not impossible. Thus future projects and activities should be encouraged to build new services in collaboration with existing expertise and infrastructure. Existing data services already possess vitally important pools of expertise that are highly tuned to their user community’s needs. These specialised data-service providers should be regarded as potential customers for pan-European data services that could make their work easier and more economic. Such customers would be expected to migrate aspects of their work onto common platforms, just as they migrate onto their latest purchases from ICT manufacturers today. This trend will be driven by a combination of economics, trust, and policy, as appropriate. There is potential to benefit from shared provision of data services, drawing on common and existing data resources, national data services, and European grid projects, while minimising disruption to users.

Many data service requirements are universal, such as the need for storage coupled with computation, bulk-data movement, archiving and replication, and reliable data referencing. These common services may not be relevant to all RI, but may be tuned to match the needs of a subset. For example, initiatives across the social sciences and humanities sectors or within biomedical sciences may require common specialised solutions. RI in thematic areas should be encouraged to exploit common requirements wherever appropriate.

Recommendations:

- Identify and promote common (long-term) data-related services across different RI.
- Encourage, through policy and facilitation, community practises and standards that assist researchers in exploiting multiple data resources, within and across disciplines.
- Raise awareness of the responsibilities set by the Toronto statement, INSPIRE directive, and similar initiatives.

3.7 Virtual Research Communities and collaboration

A virtual community is a group of scattered individuals and/or organisations that share common interests and resources. Popular examples of such communities are YouTube, eBay and Wikipedia. By analogy, Virtual Research Communities (VRCs) are widely dispersed groups of researchers and associated research resources working together in virtual research environments.

The virtual (electronic) integration of distributed people and things through the underlying e-Infrastructure is a key enabler for supporting distributed research teams, facilitating inter- and multidisciplinary research, and driving innovative exploitation of the combined research potential of Europe. As such, VRCs are fundamental to the vision of the ERA. In practice, support of VRCs requires integration of e-Infrastructure and more general RI services into standard practices and tools used by researchers.

In addition to direct support for research endeavours, VRCs support can extend access to research and research outputs, increase openness, improve (virtual) mobility and cost sharing, and reduce duplication.

VRCs not only enable integration of distributed research capacities/resources, but also allow:

- virtual mobility of physically distant researchers,
- accessibility of intermediate as well as final research results,
- combination of distributed multidisciplinary knowledge,
- discovery of novel R&D directions,
- enhancement of innovation capabilities,
- integration of research and education,
- improvement of regional and global collaboration,
- development of high-performance, sustainable multilateral partnerships, and
- improvement of organisational as well as financial practices

VRCs developments will change attitudes and approaches within (virtual) research communities, directly supporting the wider integrating agenda of the ERA.

VRC developments should proceed gradually, starting with domain-specific shared access to distributed resources, eventually expanding to integrate different research activities, and ultimately including wider support infrastructure integrating separate organisations/finances/management. Throughout, due emphasis on monitoring and motivating of researchers' attitudes and behaviour will be essential.

Recommendations:

- Build VRCs by fully exploiting the benefits of the ERA and integrating a stimulating mixture of the involved organisational entities. Multidisciplinary and/or Public Private Partnership VRCs are recommended.
- Ensure VRC developments are incremental and application- and challenge-led.
- Exploit the educational and innovation potential of VRC tools developed and deployed by research communities.

3.8 *Generic Issues*

A number of issues are common to many or all of the different e-Infrastructure service areas covered above. In the time and space available it is impossible to fully explore these issues. They are listed here as a catalogue to be addressed in more detail in the future:

Business and governance models for the e-Infrastructure and for their relationships with European RI. What is the responsibility of individual RI to fund and deploy advanced ICT services supporting broad international access? This impacts directly on the likely engagement of RI with generic e-Infrastructures. The e-Infrastructure projects potentially have an important role in offsetting the costs to each individual RI of deploying broadly based ICT tools. However, this raises the question of how to fund e-Infrastructure services themselves. Experience with networking has shown that collecting financial contributions from the users based on their share of using the network remains difficult.
There is a strong need for user-centric solutions. In order to fulfill this need, it will be necessary for users, and in particular high–end users like those represented in ESFRI RI, to play a bigger role in the governance and management of the e-Infrastructures they need. In principle, this can be done on several levels:

1. <u>On the strategic level</u>. International users in research have to play a major role in setting strategic goals for *international* e-Infrastructures. This will help to formulate and generate organisational requirements for the way national and international RI should be organised. ESFRI itself might play a key role here.
2. <u>On the operational level.</u> Advanced users of international e-Infrastructures should support the innovation and real-life testing of new e-Infrastructure developments, for their mutual benefit. LHC and eVLBI are already good examples of this approach.
3. <u>On the level of standardization.</u> In particular, (ESFRI) RI user communities should contribute to the process of setting and implementing the international standards necessary to achieve the transition from the current e-Infrastructure service portfolio to the international service-oriented e-Infrastructure portfolio recommended in this paper.

Digital Divide issues: Access to and via e-Infrastructure for researchers in remote or developing areas is an ongoing challenge. Over the last decade, co-funding for the GÉANT e-Infrastructure by the European Commission has proven essential in maintaining stability and sustainability of network access in a number of European countries. In addition to the initial structuring effect of European funding, continued support has had a strong influence on European cohesion and national investments. There is concern that, despite the recognised importance of e-Infrastructure in supporting innovation[13], any gaps in European (co-) funding could have a serious negative impact on national investments and thus, on the services themselves, across these critical e-Infrastructure areas, by causing considerable difficulties for many e-Infrastructure operators/NRENs in Europe.

New user induction and user training have been identified as common requirements across almost all ESFRI projects. In 2008 the e-IRG Education and Training Task Force (EETF) published a report[14] highlighting the motivations, opportunities, challenges and strategies for education and training. The report recommended, among other related investments, collab-

13 Competitiveness Council conclusions on the future of information and communication technologies research, innovation and infrastructures, 3/12/ 2009 http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/intm/111719.pdf

14 Education and Training Task Force report http://www.e-irg.eu/images/stories/publ/task_force_reports/ettf_long_report_final_july08.pdf

oration across the ERA and harmonisation of policies, standardisation of material, as well as sharing of related training infrastructure. An important consideration is how best to exploit the opportunities presented by e-Infrastructure without unduly disrupting standard research practices and the need to integrate systems into existing tools and vice versa. Related to this is the issue of **user friendliness**. This remains an issue, even though it is recognised by service providers. The (lack of) maturity and user friendliness of the current grid and cloud services is one example here.

Cost effectiveness has already been highlighted as a driver for shared e-Infrastructure. The benefits here are likely to be at least as great in the shared development of the e-Infrastructure as in any sharing of physical resources. The latter may be possible to a greater or lesser extent depending upon a number of factors, including level of demand, nature of the resource, and related access policies.

Green IT is a growing issue and one we should expect to become increasingly important at both a policy and practical level. Green IT includes the dimensions of environmental sustainability, the economics of energy efficiency, and the total cost of ownership, which includes the cost of disposal and recycling[15]. For e-Infrastructures, this is linked to operation of the different components, and can be interpreted as being mostly related to energy efficiency (reducing electricity consumption/reducing carbon emissions), and the use of renewable sources either directly (e.g. for cooling) or indirectly (e.g. use of energy from renewable sources).

Software is essential to the operation of e-Infrastructure and the ESFRI RI. The cost of developing and sustaining effective software should not be underestimated. Development costs are often hidden, since researchers may spend significant 'research' time writing software. Poorly written software can bring a significant cost in terms of maintenance, which can require more investment than the initial software writing. Today, almost all software faces the new challenge of exploiting the multi-cored and multi-CPU systems increasingly deployed in systems from desktops to the most powerful supercomputers.

15 http://en.wikipedia.org/wiki/Green_computing

4 Mapping of ESFRI requirements to e-Infrastructure service areas

An extensive survey of the e-Infrastructure requirements of almost all of the projects listed in the ESFRI roadmap has recently been undertaken by the European e-Infrastructures Forum (EEF, http://www.einfrastructure-forum.eu/). The EEF is a collaboration of the Pan-European e-Infrastructure providers in the areas of Research Networking, High Performance Computing, secure data-storage and services, and the European Grid Infrastructure. The interest of the EEF in this effort is to tailor their services to the ESFRI projects' needs, to avoid parallel e-Infrastructures being set up without connection to existing or planned investments and to have links established from the EEF to the ESFRI projects. Although this work, therefore, necessarily focuses on an evolution of the currently available e-Infrastructure services, the EEF provides a valuable way of interacting in a coordinated way with user communities of a multinational nature that are interested in making use of the e-Infrastructure, and therefore helps to present a common view on European e-Infrastructures.

The EEF survey is a valuable exercise and has already led to more extensive work (e.g. an effort to understand the requirements of the biomedical science projects (see ELIXIR appendix – section A.7.)). A more detailed EEF assessment of the current state of e-Infrastructure requirements is presented in [[16]]; however, their main findings are summarized below.

Common requirements of the **Social Science and Humanities** (SSH) projects:

- Data archiving and curation
- Flexible repository system with support for Persistent IDentifiers (PIDs)
- Fine-grained Authorization and Authentication systems to cope with data sensitivity (i.e. digital rights management)
- Access to grid and cloud computing facilities
- Ease of access to all e-Infrastructure resources via Single Sign-On
- Education and training for e-Infrastructure usage

Common requirements of the **Biological and Medical Science** (BMS) projects:

- Ability to manage massive quantities of data being generated across Europe and continuously growing
- Access to a wide variety of databases which are regularly updated
- Specific sub-communities deal with patient data where secure data management is essential
- Ease of use via Single Sign-On facility
- Large-scale usage of grid and cloud computing
- Collaboration with partners in USA and Japan is essential
- Education and training for e-Infrastructure usage

16 http://www.einfrastructure-forum.eu/documents/EEF-report

Common requirements of the **Environmental Sciences** (ENV) projects:

- Mixture of fixed and mobile facilities
- Ability to gather data from large-scale and distributed sensor networks in a timely manner (some real-time constraints)
- Combining, processing and storing of large, complex data

Common requirements of the **Energy** (ENE) projects:

- Emphasis on modelling, simulation and processing facilities for single-site installations
- Access to HPC and grid systems with the ability to run complex work-flows across the range of e-Infrastructures
- Education and training – notably application code porting to grids and massively multi-core machines

Common requirements of the projects in **material and analytical facilities** and **Physical Sciences and Engineering** (PSE):

- Emphasis on modelling, simulation and processing facilities for single-site installations
- Data curation, archiving and distribution
- High-speed, reliable networks for continuous data transfer
- Open access to data and results for large, widely distributed user communities across and beyond Europe
- Single sign-on service

This extensive list reveals a number of requirements common across all of the thematic areas:

- Single sign-on: consistent access to resources
- Virtual organisations (collaboration)
- Persistent storage: long-term preservation of data and its access
- Data Management services
- Standards – web services
- Workflows – support of access to HPC/grid/network resources (compute and data) across Europe
- Training
- Global scope: beyond Europe

The EEF members have already proposed to harmonise their existing services in a number of areas in a manner that reflects these requirements with the aim of relatively quickly adding value for all the user communities. In particular:

- Access to resources (i.e. harmonising policies for Authentication, Authorisation, Accounting and Auditing),

- User support (i.e. problem handling procedures) and training,
- Security incident handling (i.e. cooperating security incident response group)
- Data management (i.e. seamless authorised access to data across the infrastructures for users).

The focus of the EEF work to date has been on identification of common requirements across projects. The relative importance of different aspects may be very different across these projects and understanding the detailed priorities and trade-offs will be an important next step for the ESFRI projects themselves. This evaluation will be facilitated by a better understanding of the services available from the e-Infrastructures, their future development and deployment roadmaps, and possible realignment of these developments in the light of the EEF work. This will necessarily be an iterative process that will only succeed with the active participation of the ESFRI RI.

5 Conclusions

In the previous sections, a number of detailed recommendations have been made. These aim to exploit the opportunities presented through more effective integration of the current and likely future e-Infrastructure developments with the ESFRI RI. A number of over-riding themes emerged during the preparation of this paper. The detailed recommendations in Section 3 are specific examples of these more general issues. Specifically:

Co-evolution of Research Infrastructures, e-Infrastructure and user requirements; Active and direct user engagement

Research Infrastructure is dynamic: it evolves and develops, driven by the technical innovation that lies at the heart of research. This is perhaps particularly true for e-Infrastructure where we are often leveraging developments and opportunities emerging in the rapidly changing commercial ICT world. Harnessing this natural evolution for the benefit of the ESFRI RI is only possible with the active engagement of these communities in the continued development of the e-Infrastructure. The responsibility of active participation placed on the RI is matched by a responsibility on the part of the e-Infrastructure itself to evolve user- (RI) focused services that can be effectively described and delivered through (a range of) SLA-like mechanisms, thereby allowing users to exploit these services with confidence over the medium and long-term. These services must also evolve to meet the evolving needs of the RI themselves.

Internationalisation of Research

Europe and European RI cannot evolve in isolation from the rest of the world. Research continues to internationalise. Projects such as the Human Genome Project and the Large Hadron Collider are, of necessity, global in scale. Climate change must be considered on a global scale. Providing local or regional tools and resources for researchers that do not integrate into their global endeavour will not provide the support required to ensure a world-leading position in research and innovation.

Exploitation of available services; Service orientation; Sustained Innovation

The most effective e-Infrastructure will be possible only when developers, providers and users all see its development and support as tackling their own problems, on their timescales, and in their interest. Exploiting the available e-Infrastructure service components, will prevent users from ‘reinventing the wheel’, and service-orientation will greatly facilitate this endeavour. Service-orientation is, however, not only acting as a facilitator for users, but also as an enabler for e-Infrastructure innovation, and sustained funding acts as a safeguard for continued innovation.

Next steps

The next steps in this process are to continue to encourage and directly support engagement across user, developer and provider communities. Barriers to this collaboration should be identified and addressed. Tackling these barriers can offer a very good return on investment. In parallel, with the preparation of this paper, the EEF have carried out an extensive survey of technical requirements of the ESFRI RI with respect to the current e-Infrastructure provision. This later work identifies some clear short-term actions. Finally, the generic issues listed in Section 3 of this report would **merit further consideration.**

e-IRG "Blue Paper" 2010

Appendices

APPENDIX A: RI Requirements and Common Practices

A.1 Introduction

The following presents a short description of a selection of the ESFRI RI projects with an emphasis on e-Infrastructure aspects. The projects described below were chosen, following discussion with the chairs of the ESFRI Thematic Working Groups, as representative of many of the issues and opportunities in their respective thematic area[17]. In a few specific cases, additional ESFRI projects were selected by e-IRG. Except where noted otherwise, the text has been iterated with representatives of the RI, however, ultimately this represents the view from e-IRG rather than the RI. The following focuses on organisational issues and does not represent a detailed analysis of the technical requirements of the RI. As such it should be viewed as independent of, and complimentary to, the requirements analysis recently published by the EEF.

There is significant repetition across the different sections, reflecting a number of common issues and priorities. Common themes emerging across multiple RI include:

- Authentication and identity management including federated identity management across Europe being pursued by Europe’s NRENs.
- Providing and managing secure access to data.
- Federating and integrating multiple data centres (including databases and web-based data services).
- Persistent and multidisciplinary European data infrastructure.
- Standards for metadata and interoperability at multiple levels.
- Integration with national e-Infrastructure developments such as NGIs.
- Legal, ethical and funding/financing issues, particularly where these cross national boundaries.

17 A full list of projects and their implementation status is available at http://ec.europa.eu/research/infrastructures/pdf/esfri/home/implementation_report_2009_en.pdf

A.2 BBMRI: Biobanking and Biomolecular Resources Research Infrastructure

BBMRI (www.bbmri.eu) aims to prepare and operate a pan-European infrastructure for biomedical and biological research, providing and managing access to biological resources required for health-related research in Europe.

BBMRI will build on existing sample collections, resources, technologies and expertise, which will be complemented with innovative components. Key components of the infrastructure will be comprehensive collections of biological samples (such as blood, DNA and tissue) from different (sub-) populations of Europe, linked with continuously updated data on the health status, lifestyle and environmental exposure of the sample donors. The infrastructure will also include molecular genomic resources, and bio-computational tools to optimally exploit this resource for global biomedical research.

In practice, BBMRI will be composed of a federated network of centres established in European Member States, and organised in a distributed hub structure in which the hubs act as key actors for the coordination of activities, including collection, management, distribution and analysis of samples and data. Data sharing will be enabled using service-oriented architecture based on web service and grid technologies.

Users of the infrastructure can be grouped into three distinct types:

- Biobanks, biomolecular resources and technology centres representing the key providers of resources and technologies.
- A variety of associated public or private partners providing biological samples, data, technologies or services (e.g., universities, hospitals, companies, governments, research councils, funding agencies).
- Users from different fields of academia and industry.

For many biomedical studies, obtaining a sufficient number of cases is crucial, regardless if the study concerns a complex or rare disease. BBMRI will enable researchers to search across more than 250 biobanks in Europe using a wide range of queries. Major synergies, gain of statistical power and economies of scale are expected to be achieved by interlinking, standardizing and harmonizing a large variety of existing and *de novo* national resources.

The challenges related to the set-up of the infrastructure are both organizational and technical, stemming from federating and integrating multiple autonomous biobanks and database systems, to providing and managing secure access to data. They also include ethical, legal, and funding and financing issues that are crucial to the sustainability of the infrastructure.

One of the key technical challenges is how to best integrate existing and future biobanks into a single network. This will entail the development of harmonised standards for the collection, storage and analysis of both the biological sample and its associated data. Security and ELSI issues are also important concerns and must be tackled to enable the use of data. A robust, easy-to-use global authentication and authorization framework will be required.

Globally, the biggest challenge for BBMRI, in terms of ICT requirements, is likely to lie on the network and software sides, more than on the hardware.

It is possible to identify a number of specific e-Infrastructure-related actions required over the next few years, including:

- Coordinating with GÉANT and TERENA (NRENs) to identify and address issues around connectivity and federated user identity management across Europe.

- Working with EGI to investigate whether the national infrastructures in BBRMI can be connected to the NGIs and some of the grid-related work coordinated by EGI.
- Collaborating with other ESFRI projects in the areas of biomedical science (such as EATRIS, ECRIN, ELIXIR) and other standards-oriented projects (such as GEN2PHEN, epSOS, BioSHaRE) to facilitate the development of a global biobanking network.
- Strengthening links with European projects focusing on data management and preservation, including new initiatives working on a strategy for a persistent and multidisciplinary European data infrastructure.

A.3 CLARIN: Common Language Resources and Technology Infrastructure

CLARIN (www.clarin.eu) represents a large-scale pan-European collaborative effort to create and coordinate language resources and technology, and to make them available and useful to scholars of all disciplines, in particular the humanities and social sciences. It aims to offer a comprehensive service with tools and resources interoperable across languages and domains, and to provide a persistent and stable infrastructure on which researchers can rely for the next few decades, thus contributing to preserving and supporting multilingual and multicultural European heritage.

The volume of written, recorded and multimodal texts is growing exponentially, which makes the use of computer-aided methods indispensable for many scholars working in the humanities and in neighbouring areas and who are concerned with language material. An ever-increasing amount of texts and recordings has been digitized or is directly created in digital form, which increases the need for and readiness to resort to computer-aided methods. At the same time highly specialized technical skills and expertise are required to make the application of computers effective and efficient.

The role of CLARIN is to overcome the present fragmented situation by coordinating and integrating the current language resources and making them available via web services. In practice, CLARIN will build on the existing national infrastructures, and will rely on a network of powerful centres that can offer a wide range of stable and highly available services, ranging from archiving services allowing others to store data resources with a guarantee of long-term accessibility, to advanced ontology services offering widely accepted and well-defined domain concepts. Every country will be responsible for its own content, and no central funding for content creation is foreseen. Therefore, CLARIN is not about technology development or content creation, but aims rather at integrating what is available and making it accessible. Its ultimate goal is the construction and operation of a shared distributed e-Infrastructure.

The technical challenges of the project are mostly related to interconnecting existing archives that may use very different ways to encode and describe data; and ensuring that existing language technology tools made for material in Archive A will also work for material in Archive B, and will work together. Developing common standards and interoperability is one of the key issues. Grid-type functionalities must ensure that communication occurs between trusted servers; that all resources in the domain have unique and persistent identifiers; that authentication and authorization are working seamlessly in distributed scenarios; and that users of organizations participating, for example, in national identity federations, are accepted with a single identity. Ultimately, CLARIN will rely on a high capacity network layer, provided by GÉANT.

It is possible to identify a number of specific e-Infrastructure-related actions required over the next few years, including:

- Coordinating with GÉANT and TERENA (NRENs) to identify and address issues around connectivity and federated user identity management across Europe.
- Strengthening links with European projects focusing on data management and preservation, including new initiatives working on a strategy for a persistent and multidisciplinary European data infrastructure.

A.4 CESSDA - Council of European Social Science Data Archives

The CESSDA (www.cessda.org) organisation promotes the acquisition, archiving and distribution of electronic data for the European Social Science and Humanities (SSH) research community. It encourages the exchange of data and technology. It associates and cooperates with other international organisations sharing similar objectives. Archives that preserve and disseminate social data ensure that these culturally significant materials are accessible in perpetuity. The CESSDA RI provides long-term access to social science data archives across Europe.

The CESSDA Preparatory Phase Project, which started on January 1, 2008, has launched a major upgrade of the CESSDA RI to make it more comprehensive, efficient, effective and integrated. The project aims at resolving strategic, financial and legal issues including:

- Transforming CESSDA into an ERIC legal entity.
- Extending the coverage of the CESSDA RI by fostering the development of national data archiving initiatives in countries which are not currently part of CESSDA, and by extending the network to organisations outside of CESSDA which host important data collections.
- Strengthening the CESSDA RI through developing the skills, knowledge and abilities of less-developed and less-resourced CESSDA organisations.
- Promote member organisations as places of deposit for publicly funded data collections, thus extending their collections for the benefit of the ERA.
- Provide a one-stop-shop for data discovery, access, analysis and delivery across the SSH community of Europe.

The technical challenges for CESSDA are related to increasing the quality of available data, creating a more dynamic knowledge-management-oriented web, developing the data portal to allow seamless access to data holdings across Europe, developing common authentication and access middleware tools, developing metadata standards, creating thesauri management tools, investigating the potential of grid technologies, and improving data harmonisation tools.

The CESSDA RI upgrade will develop a system of networked data collections that is open, extensible, and evolvable, and will support development of a new generation of tools and services for data discovery, integration, visualization, analysis and preservation.

It is possible to identify a number of specific e-Infrastructure-related actions required over the next few years, including:

- Coordinating with GÉANT and TERENA (NRENs) to identify and address issues around connectivity and federated user identity management across Europe.
- Strengthening links with European projects focusing on data management and preservation, including new initiatives working on a strategy for a persistent and multidisciplinary European data infrastructure.
- Work with EGI to investigate if and how grid technologies can be incorporated in the future CESSDA-based SSH e-Infrastructure.

A.5 DARIAH - Digital Research Infrastructure for the Arts and Humanities

Research practices in the Arts and Humanities community increasingly require the availability of an advanced pan-European digital infrastructure to enable researchers to access and enrich the information and the knowledge that is embedded in digital content. DARIAH (www.dariah.eu) aims to develop the technical framework needed for this, and to offer this RI to researchers across Europe.

The goal of DARIAH is to enhance and support digitally-enabled research through long-term access to digital information on arts and humanities. This means that distributed digital information resources of archives, libraries, museums and other 'repository agencies' will have to be linked to make the wealth of data compiled by researchers and institutes across Europe widely accessible. In view of the average technical skills of the end-user community, it is mandatory that users can easily access the shared data from their desktop. This will require the development of innovative ICT tools for finding, interpreting and processing digital content.

Another goal is to further promote the sharing of data, ideas, methodologies and expertise and make different collections of information accessible and compatible. Therefore DARIAH is collaborating with communities of practice to explore and apply ICT-based methods and tools to facilitate research and to exchange knowledge and methodologies across domains and disciplines.

The organisational challenges for DARIAH are related to:

- Bringing together the best efforts at national, organisational and individual level in order to launch enhanced Europe-wide actions, initiatives and services.
- Developing national services and digitisation programs, particularly in countries where these are non-existent at present.
- Bringing together research, education, cultural heritage and 'memory' institutions and organisations in the commercial sector.
- Enhancing digital scholarship in the humanities and arts across Europe.

The technical challenges for DARIAH are related to linking the existing information sources, developing the ICT tools and ensuring the long-term preservation of data. The key issues to be solved include technical interoperability (of data, metadata and tools), and data access, curation, preservation and management. The development of common standards and the use of best practices is key to addressing these challenges.

DARIAH has already joined forces with other Humanities and Arts Infrastructures and Networks in the CHAIN coalition, which aims to support and promote the use of digital technologies in research in the arts and humanities. CHAIN brings together members with experience in creating and operating digital infrastructure with the aim to create a shared environment where technology services can interoperate and be sustained, thus enabling new forms of research.

It is possible to identify a number of specific e-Infrastructure-related actions required over the next few years, including:

- Coordinating with GÉANT and TERENA (NRENs) to identify and address issues around connectivity and federated user identity management across Europe.
- Strengthening links with European projects focusing on data access, curation, preservation and management, including new initiatives working on a strategy for a persistent and multidisciplinary European or Global data infrastructure, such as the ones initiated by the CLARIN and CESSDA consortia.

A.6 ECRIN: European Clinical Research Infrastructures Network

The ECRIN project (www.ecrin.org) is intended to develop and provide a European infrastructure for clinical research. By integrating capacities and capabilities of national networks in Europe, ECRIN helps to overcome the fragmentation previously seen in this field. By making it possible to leverage pan-European populations and competences, ECRIN will support multinational clinical trials.

The value of this RI is in making Europe more attractive for clinical research and thus increasing the competitiveness of European biomedical research for academic institutions, small- and medium-sized enterprises, and the health industry. These groups will make use of ECRIN through services designed to support investigators and sponsors in multinational studies.

Fragmentation of European knowledge, approach and legislation has a negative impact on clinical research, as clinical trials require access to large populations in order to effectively develop new approaches and therapies. Recognition of this at the national level has lead to the creation of networks in some member states to coordinate and connect clinical research centres and clinical trial units.

The emergence and success of national networks showed the need for ECRIN, a Europe-wide network to support the practices required for quality and credibility of data and successful performance in clinical trials. ECRIN began as a series of FP6 projects in 2004 and is currently in its third phase (ECRIN-PPI). Earlier phases identified bottlenecks (2004-5), and designed suitable infrastructure (2006-2008). The current preparatory phase under FP7 (2008-2011) creates the legal basis for the infrastructure, develops a sustainable business plan and evaluates the proposed infrastructure through pilot clinical research projects. This will be followed by an implementation phase (2011-2014) that will construct the infrastructure itself.

ECRIN uses a multi-angle approach to support all areas of clinical research for all diseases. This includes:

- Supporting patient enrolment in clinical trials, data management, quality assurance, monitoring, ethics and regulatory affairs.
- Promoting further development of the national networks that contribute to ECRIN, and extension of the network to all the EU member and associated states, and developing partnerships with other global regions.
- Ensuring unbiased methodology, protection of trial participants, high ethical standards, transparency of data use, and reporting at all levels.
- Promoting participation of patients in clinical research and ensuring communication of the challenges raised by clinical research with all stakeholders.
- Harmonising national legislation to provide a framework that protects trial participants, while at the same time minimising obstacles to research institutions and facilitating the connection of disease-orientated networks across borders.
- Promoting European-level integration of funding for clinical research and facilitating education, training and mobility programs for all stakeholders.
- Work with other infrastructures in the biomedical field to pursue common strategies, ensure compatibility of data and procedures with other parts of the pipeline from basic research to patient therapy.

ECRIN is a successful network but does not appear to be a technical infrastructure in the way the e-Infrastructure domain normally describes them. Rather, it is focussed on preparatory work, consultancy, information sharing and brokering services. While this may change in the construction phase (2011-2014), it appears the major technical requirements will be for data management rather than a broader set of technical services. However, its work does have some clear links to work in the e-Infrastructure sphere, where cooperation could benefit both groups. As such, the following actions should be considered.

- Work closely with groups investigating legal issues related to e-Infrastructure. The partners should include both policy bodies (such as e-IRG) and projects (ongoing projects and consortia of concluded projects) dealing with patient-related biomedical information to avoid potential legal barriers to the procedural, scientific and technical advances ECRIN provides. Help to mobilise this knowledge in order to provide national and European lawmakers with concrete and united recommendations for future legislation in the area.
- Connect the national networks on clinical research to their respective NGIs in order to provide local lateral as well as central links between ECRIN and the emerging European e-Infrastructure. In areas without ECRIN partner networks, use NGIs as a framework to support extension of ECRIN in line with its stated desire to operate in all member and associated states.
- Work with EGI to help both EGI and ECRIN understand the differences between ‘pure research’ and clinical infrastructures. This will make ECRIN better aware of the existing e-Infrastructure available to them, and help EGI engage with the clinical community, who have a unique combination of requirements for privacy (due to patient data), security (due to the commercial value of results) and legal issues (due to the close oversight of clinical work by governments) not seen in other EGI user communities.
- Clarify the status of actual production use of clinical trial infrastructure (including the involvement of commercial entities) with regards to the acceptable use policies of GÉANT and NREN networks.
- Use ECRIN’s connection with commercial sponsors and participants in clinical trials and research to act as a pipeline for integrating existing, de-facto or emerging standards for data, processing and storage into the traditionally more academic EGI sphere.
- Ensure that the data management challenges (e.g. ones related to long-term curation of clinical data and long-term maintenance of the confidentiality of patient data) related to ECRIN are included prominently in the planned blueprint for enabling data-intensive research.

A.7 ELIXIR – European Life Sciences Infrastructure for Biological Information

The ELIXIR project (www.elixir-europe.org) aims to construct and operate a sustainable infrastructure for biological information in Europe, to support life science research and its translation to medicine and the environment, the bio-industries and society. In practice, this means the creation of a pan-European network of national and international centres (nodes) to store and curate biomedical data and to serve this data to a large, varied and widely distributed range of users. The nodes, coordinated through EMBL-EBI, will operate as a single logical entity from the perspective of users. The project thus spans a number of technical and socio-political issues ranging from semantic technologies, through to coping with the rapidly increasing volumes of digital biomedical data, to organizational issues addressing long-term funding models for the data and the nodes themselves.

It is recognized that the current exponential growth in data cannot equate to exponential growth in funding. Therefore there is a need to make more effective use of distributed resources, including international activities outside Europe (e.g. USA and Japan). This is in addition to more effective linking of budgets for data processing and data generation and taking steps to control staff growth through automation. To achieve the required efficient integration, ELIXIR nodes have to be closely coordinated and work as a single unit. This must be achieved within the context of national funding and priorities. In parallel, it will be important to develop policies for data retention.

Potential user communities in the medical and life sciences are very large (millions of potential users), relatively technically naïve, and have a expectation that applications and data are freely available on-demand through web-based services. The need for computation is traditionally seen as limited, with data storage as a key driver. However, the computational requirements to support the large volumes of data involved are increasingly demanding. Users can be grouped into three distinct types:

- Data generators (hundreds to thousands): data acquisition, reduction and publication
- Computational scientists and bioinformaticians (thousands): large computations, some with large memory, some HPC but mostly large capacity.
- Users (millions): academic life science researchers, clinicians and (increasingly) members of the public. Use *via* web without authentication, mainly access to reference data collections with some large computation and rather more small computations. Need training.

ELIXIR is a pan European e-Infrastructure anticipated to provide services for the other national and international biology projects. As such, it provides an opportunity and a forum for influencing and interacting with other e-Infrastructure providers focused on efficient and effective horizontal delivery. Triggered in part by interactions with the current e-Infrastructure projects, ELIXIR has already taken steps to develop a more detailed understanding of the requirements and opportunities for using coordinated/generic e-Infrastructure across the biological and medical sciences through a workshop on BMS RI Co-operation with e-Infrastructure (29-30 March 2010 Hinxton, Cambridge). This workshop reiterated the importance of capacity computing, networking and storage across BMS projects, with only very limited requirements for high end HPC.

It is possible to identify a number of specific e-Infrastructure-related actions required over the next few years, including

- Coordination with GÉANT and TERENA (NRENs) to identify and address issues around connectivity for current and potential ELIXIR nodes, and possibly for federated user identity management across Europe.

- Work with GÉANT and GLIF to ensure that the required inter-continental connectivity is provided to partner biomedical data infrastructures in USA and Japan, etc.
- Work with EGI to investigate whether the emerging EGI structures can be used to co-ordinate physical infrastructure provision across Europe. The national partners in EGI (National Grid Initiatives) are generally the providers of physical ICT infrastructure for research in small to medium EU countries (and will need to work closely with domain specific providers in larger countries). EGI is focused on integrated international support and data driven computing. Hence EGI provides a possible single point of contact for interactions with these providers. This in turn could help to influence the national investment strategies to better align with ELIXIR requirements.
- Raising awareness of the INSPIRE directive (http://inspire.jrc.ec.europa.eu/) and its potential impact on biomedical data.
- While work in areas around bio-medical metadata is likely to remain domain-specific. Experience operating pan-European (or broader) catalogues and data movement tools already exists within EGI and GÉANT; this experience could be leveraged through inclusion in relevant ELIXIR work packages (e.g. WP 12), possibly leading to joint initiatives in the future.

Each of these interactions will likely require significant engagement from ELIXIR personnel. However, this must be matched by a commitment on the part of the e-Infrastructure providers to deliver sustainable solutions.

A.8 e-VLBI: Very Long Base Interferometry in Europe

Radio astronomy is a subfield of astronomy that studies celestial objects at radio frequencies. Radio astronomers use different techniques to observe objects in the radio spectrum. The most common and well-known device used in radio astronomy is a single radio telescope pointed at an energetic radio source to analyse its type of emission.

Since the 1970s, telescopes from all over the world have been combined to perform Very Long Baseline Interferometry – VLBI. VLBI is a technique invented by astronomers in which physically independent and widely separated radio telescopes observe the same region of sky simultaneously in order to generate very high-resolution continuum and spectral line images of cosmic radio sources. Since VLBI telescopes are usually separated by many thousands of kilometres, data from each telescope is digitally sampled and stored locally, using magnetic disk-array systems. The disks are physically transported to a central data processor (a purpose-built supercomputer) where the data from each telescope is decoded, accurately aligned, and then correlated in an exhaustive pair-wise fashion for every possible telescope combination.

The total flow of data into the central processor is approximately 10 to 100 Terabytes per observation; after processing, the data can be reduced by two to three orders of magnitude. The output from the data processor may be pipelined – a process in which the data is automatically calibrated and preliminary images are made. The data (including the raw processed data) is presented to the astronomer for further detailed and careful analysis.

In Europe, VLBI observations are organised and conducted by the European VLBI Network (EVN), which was formed in 1980 by a consortium of five major radio astronomy institutes in Europe. EVN is the most sensitive VLBI array, thanks to the collection of extremely large telescopes from all over the world. The central EVN data processor was developed and is now operated by Joint Institute for VLBI in Europe (JIVE), located in Dwingeloo, the Netherlands. As the EVN telescopes observe the cosmic radio sources simultaneously, they produce an extremely large amount of data. Originally, the data was recorded on high capacity magnetic tapes which were transported to JIVE after experiment end. Thus additional time was required for shipment of the tapes and finally data processing and correlation: experiment results were available for astronomers after a significant delay. Later on, the EVN consortium introduced disk packs, which replaced magnetic tapes. This slightly improved the situation. However, the main problem still has not been solved: disk packs have to be transported to JIVE via regular mail.

A decade ago, a 20,000 node PC cluster would have been required to match the processing power of the EVN MkIV Data Processor at JIVE. However, this number has been falling rapidly as processor capabilities continue to develop. The networking infrastructure is also increasingly capable of dynamically handling very large data transfers over long periods of time. This has led to the introduction of e-VLBI system, where data is sent via the network to the central processing facility at JIVE. This has drastically shortened the time required to obtain final results from actual observation. However, a new bottleneck has emerged in the e-VLBI approach: the single central data processor at JIVE.

To solve the problem of the single correlation point, a concept of software-based distributed correlation embedded in the grid computing environment has been introduced. The bottleneck in the previous approach – the central data processor at JIVE – has been replaced with a distributed, software-based correlator spread all over the grid environment. The grid approach to e-VLBI required the creation of some form of user interface as a typical approach based on creating a text-based single control file for an experiment was not sufficient. In this case, a user should specify a set of grid-specific parameters and define required resource types together with their specific connections, creating an e-VLBI workflow.

Telescopes are set up and combined together to observe the same region of sky. Each radio telescope is assigned to a *File Server* by the VLBI Operator during the observation workflow design phase. There can be many File Servers defined. However, each radio telescope has to have one File Server assigned. Data from each telescope is sent to the attached File servers. In order to correlate the stream from telescopes, the exact data ranges from all corresponding telescopes are required and they have to be present at one location. To speed up the correlation process and make use of grid resources, the software correlation modules have been deployed within the grid environment. The data stream from the telescopes is divided into small data chunks. The data chunks are prepared by File Servers and they are sent to Correlation Nodes. The process is controlled by the correlation algorithm, which controls the way data chunks are delivered to the Correlation Nodes. Whenever corresponding data chunks are correlated, they are combined with the previous results and stored in the central data archive.

The first tests of VLBI experiments proved that distributed software correlation with grid resources is possible. The tests also showed that improvements and enhancements are required in order to make it a production service. The telescopes produce an extremely large amount of data. Even though the data is split into small chunks, the network is still the bottleneck. The infrastructure needs to support high bandwidth networking on demand. The paths need to be established dynamically on demand so multiple requests can be served at the same time for different radio telescopes. Moreover, the infrastructure should also provide a high-speed storage or buffering functionality for storing data chunks and processed data in case of network failure. These improvements are necessary if we think of e-VLBI as a service able to execute many VLBI experiments at the same time.

A.9 ESRF: European Synchrotron Radiation Facility

ESRF (http://www.esfr.eu) operates one of the most powerful synchrotron radiation sources in the world, serving the thousands of researchers who use high-intensity X-ray beams to study the structure and behaviour of matter at the atomic level. The Upgrade Programme initiated in 2009 will renew one third of the beamlines available for experiments, and provide the most advanced instrumentation and innovative solutions for carrying out the most challenging experiments.

ESRF is structured as a company under French law, responsible for running the laboratory's infrastructure, which consists of the synchrotron facility itself, 31 public and 11 collaborative research beamlines, specialised support laboratories and a modern computing environment. Most of the experiments are conducted by visiting research groups belonging to numerous disciplines, such as materials sciences, life sciences and chemistry. In 2009, 1731 experimental sessions were carried out by 6927 visiting European scientists under the ESRF user programme.

Scientists interested in visiting ESRF submit proposals to one of the peer review committees, which meet twice a year. In 2009 these committees reviewed 2047 applications for beamtime, and selected 929 experiments.

As a single experimental apparatus, supporting a very multidisciplinary user community, ESRF is a natural hub for inter- and cross-disciplinary activities. This has led to the formation of partnerships such as the Partnership for Structural Biology (PSB, with ILL, EMBL Grenoble and IBS) and hosting the Unit of Virus Host Interactions (UVHCI) on the ESRF campus. These partnerships and collaborations support ESRF's goal of offering an integrated platform for the numerous applications using synchrotron radiation.

The basic data management model is based on the users storing their data on the ESRF onsite computing infrastructure for short periods of time and shipping it to their host institutions for in-depth processing. ESRF will store user data for a maximum period of six months. Thus, long-term data curation and analysis issues are currently the responsibility of the visitors rather than ESRF. However, ESRF needs to maintain a sizeable IT-infrastructure to serve the immediate needs of the users: the computing centre currently has 750TB of disk storage capacity, a 600 CPU compute cluster, and large tape storage systems.

The volume of data produced by the scientific experiments is expected to grow to 500 – 1000 TB/day within the next decade. Despite the planned upgrade of the onsite IT infrastructure, handling of this increased data flow will likely approach the limits imposed by the available cooling capacity and electric power. This means that ESRF sees computing as an issue that is on the critical path for the organisation's continued success. ESRF has already made considerable efforts to explore grid computing solutions to support its scientific mission by evaluating the suitability of the gLite middleware for future data-intensive synchrotron research activities. While the results of this particular evaluation indicated that the current solutions were not yet suitable for data-intensive ESRF use cases, the organisation wants to play a leading role in establishing virtual organisations for synchrotron radiation research and offering a virtual computer for online and offline computing for its users.

These ambitious e-Infrastructure-related goals open up several areas of potential collaboration between the different e-Infrastructure service providers, ESRF, and its user communities:

- Optimisation of the movement of large data sets (for exporting user data, both for processing and long-term storage) and HPC computing capacity could be an interesting case for both advanced networking services offered by the NREN networks and the HPC computing services (backed up by an efficient global filesystem solution) brought together by the PRACE initiative.

- ESRF, together with the other synchrotrons and neutron sources (such as ILL, which shares the EPN Science Campus with ESRF) in Europe, has very demanding requirements related to the management of the credentials used to access various e-Infrastructure resources. A thorough requirement-gathering exercise done with the e-Infrastructure service providers would produce valuable data that could allow ESRF to serve its visitors better through new (possibly more fine-grained) security models supported by the e-Infrastructure providers.
- Matching the end-user application to the optimal solution from the e-Infrastructure service palette is a major challenge. For example, selecting the wrong type of computing resource for a task will result either in higher than optimal processing costs or poor performance or both. ESRF could be one of the ideal locations for the e-Infrastructure initiatives to study this matching problem.
- The multidisciplinary partnership organisations of ESRF are likely to have interesting data management related issues that could form a major contribution to the blueprint for enabling data intensive research mentioned in the 2010 e-IRG roadmap. The results of the PaN-Data and VEDAC initiatives, which aim at structuring data management between ESRF and ILL, are of special interest with regards to data management issues.

A.10 EuroFEL: Free Electron Lasers of Europe

Fact-checking by project representatives pending- Request for comments sent

EuroFEL (http://www.eurofel.eu) will integrate the national activities in Europe to deliver a unique, distributed European RI of Free Electron Laser light sources and other advanced accelerator-based short-pulse and coherent radiation sources demanded by a large, multi-disciplinary user community.

Coordinated research and development of relevant technologies and joint efforts in education and training will make optimum use of resources and know-how and will maintain the EU's technological infrastructure at a world-class level.

Scientific experiments at the EuroFEL facilities dramatically benefit from the opportunities provided by modern, mostly two-dimensional detectors. However, the high data rates generated by these devices pose considerable challenges for the facilities and their users. Therefore, a coordinated and common approach to cope with these high data rates is urgently required. This includes the development and availability of appropriate hardware and software tools, comprising online data evaluation to enable feedback information for fast quality assurance during the course of an experiment, and hence for the most efficient use of valuable experimental time; a standard data format for an easy exchange of data and compatibility with evaluation software; a data lifetime policy for remote user access to large amounts of data; long-term archiving; the development of simulation, modelling, and visualization tools enabling experimentalists to obtain scientific results from their samples more quickly, easily and reliably. A close collaboration with the user community and with other RI dealing with FELs, synchrotron radiation, lasers, neutrons and ions is mandatory.

EuroFEL as a consortium of facilities will be able to make more effective use of Europe's e-Infrastructure. Part of the ICT requirements can be satisfied sharing resources between partners (e.g. online processing and short to medium term storage) and part by using other resources of the e-Infrastructure (e.g. off-line processing, long-term storage).

It is possible to identify a number of specific e-Infrastructure related actions required over the next few years, including:

- Coordinating with GÉANT and TERENA (NRENs) to identify and address issues around connectivity and federated user identity management across Europe.
- Working with EGI to investigate whether the national infrastructures in EuroFEL can be connected to the national NGIs and some of the grid-related work coordinated with EGI.
- Strengthening links with European projects focusing on data management and preservation, including new initiatives working on a strategy for a persistent and multidisciplinary European data infrastructure.

A.11 KM3NET - The Cubic Kilometre Neutrino Telescope

KM3NeT (www.km3net.org) is a future deep-sea RI hosting a neutrino telescope with a volume of at least one cubic kilometre to be constructed in the Mediterranean Sea. The Design Study for the infrastructure, funded by the EU FP6 framework, started in February 2006 with the primary objective of developing a cost-effective design for the neutrino telescope infrastructure. The Preparatory Phase of the infrastructure, funded by the EU FP7 framework, started in March 2008 and is planned to continue into 2011, possibly extended to 2012.

The Mediterranean Sea appears to be an ideal place for this installation: it provides water of excellent optical properties at the right depth and excellent shore-based infrastructure for marine operations and on-shore data processing. With its foreseen technical characteristics, the KM3NeT neutrino telescope will be unique in its physics sensitivity and will provide access to scientific data that will propel research in different fields, including astronomy, dark matter searches, cosmic ray and high-energy physics.

Over the past decade, the three pilot projects ANTARES (http://antares.in2p3.fr/ in the south of France), NEMO (http://nemoweb.lns.infn.it/ south east of Sicily) and NESTOR (http://www.nestor.noa.gr/ on the west coast of the Peloponnese) have been exploring the technologies, building and deploying smaller scale prototype telescopes designed to operate at depths ranging from 2500 to 4500 metres. Construction of the ANTARES telescope was completed on May 30, 2008. ANTARES is now the largest neutrino telescope at the Northern Hemisphere. The design, construction and operation of the KM3NeT neutrino telescope will be pursued by a consortium formed around the institutes currently involved in the ANTARES, NESTOR and NEMO pilot projects. Based on the leading expertise of these research groups, the development of the KM3NeT telescope is envisaged to be achieved within about four years for preparatory R&D work, plus four years for construction and deployment.

An important problem to be resolved, which is related to wider RI issues, is the handling of the data that will be produced by the telescope. Specifically, the following are topics related to the interaction between KM3NeT and the e-Infrastructure community:

- Data arriving on the shore will be filtered and then transferred to various computing centres to be further processed and stored on a permanent medium, and for permanent and final archiving. Therefore, a high bandwidth link of typically 100 MB/s to the various European computing centres is required. Remote operation of the infrastructure is foreseen. Issues arising should be resolved in consultation with GÉANT and local NRENs.
- For processing of the data by distributed computing centres, it would be useful to investigate progress in the European grid infrastructure offered by EGI and the status of grid middleware as pursued by EMI.
- To deal with data management issues (like metadata and quality of data, interoperability, retention policy, etc.), KM3NeT might benefit from the work of the e-IRG Data Management Task Force and its follow-on efforts.

A.12 LIFEWATCH – e-Science and technology infrastructure for biodiversity data and observatories

LifeWatch (http://www.lifewatch.eu) defines itself as focused on 'e-Science and Technology Infrastructure for Biodiversity Data and Ecosystem Research.' This ambitious project brings together numerous academic partners, representatives from nineteen countries and eight existing scientific networks. It also collaborates with other international and global efforts such as the Global Biodiversity Information Facility (GBIF), and forms a key component of the European contribution to the Global Earth Observation System of Systems (GEOSS).

There is a considerable body of biodiversity-related data within Europe and globally, and there have been some creditable attempts to bring it together, such as GBIF. However, there remain gaps in the data (both spatial and temporal), unincorporated legacy data sources, and concerns about erosion of the observational infrastructure and archiving methods used for this data. These deficits impede large-scale analysis and modelling based on biodiversity and related data, especially in areas such as climate change, where long-term data is required for meaningful analysis.

The LifeWatch initiative answers these needs through a series of phases that began in 1999. From 2005-2011 the project is in a preparatory phase, with EC support via the Seventh Framework Programme since early 2008. This will be followed by a construction phase (2011-2016) and a test and operational phase (2014-2036).

The overall aims of the initiative are to 'secure the essential infrastructure and information systems necessary to collate existing data on biodiversity, and distribute this information with analytical and modelling capabilities to the scientific community and to other users in the public, commerce and policy sectors.'

LifeWatch mentions four ambitions in its plans: Opening up new opportunities for research, providing services to society, providing value-added networking and carrying out data integration from different scientific domains. Of these, the last is most relevant to the e-Infrastructure community, as LifeWatch proposes using a grid environment to allow for 'large-scale distributed data storage and high-performance computing'.

Overall, LifeWatch makes frequent reference to technical aspects of e-Infrastructure, and does indeed seem to be an RI in the way that the e-Infrastructure community describes them. However, unlike some other RI, LifeWatch explicitly mentions running an end-to-end infrastructure, from the resource layer, through the e-Infrastructure layer, up to the service and user layers.

LifeWatch plans to use a grid-based approach, and proposes an architecture comprising grid-enabled data resources, extension of these with various grid services, and finally, a biodiversity e-Science interoperability layer. The project has also constructed a reference model based on one produced by the FP6 ORCHESTRA project. The ORCHESTRA reference model (RM-OA) specifies a service-orientated spatial data infrastructure that takes into account the EU INSPIRE Directive and relevant ISO specifications, as well as being supported as a best practice by the Open Geospatial Consortium.

The project has already had contact with the EGEE project to ensure its requirements are taken into account by the common e-Infrastructure, but it is somewhat unclear whether LifeWatch plans to create its own internal solution or transit to being a user of a larger common system that can fulfil its needs. There may also be some mismatches between the maturity of solutions available from the e-Infrastructure community and the needs of LifeWatch, and these present a danger if LifeWatch were to adopt a solution not mature enough or otherwise inappropriate for them, or alternatively choose to create a new system without benefitting from the investment in common e-Infrastructure made by the EC.

The following are actions or topics for investigation related to interaction between Life-Watch and the e-Infrastructure community.

- As LifeWatch appears to be at the forefront of the data management challenge they should liaise closely with EGI and PRACE to understand one another's approaches and let the infrastructures formulate strategies that will support LifeWatch. Additionally it would be useful for LifeWatch to investigate work from e-Infrastructure projects working on generic data management issues to understand whether the results of such efforts could be applicable to LifeWatch.
- Related to the above, LifeWatch could play an important role in passing emerging standards for data management and access from the biodiversity field (created through efforts such as GBIF that LifeWatch works with) to the e-Infrastructure community.
- Using the results from the above recommendations, clarify plans as to whether LifeWatch will be an e-Infrastructure provider or will be a consumer of services from the current e-Infrastructure. Ensure that LifeWatch benefits in some form from the work in the e-Infrastructure sphere, both as a user and by using experience of successes and difficulties in e-Infrastructure development to inform its own technical work.
- Interact with EGI and EMI to ensure LifeWatch is aware of the new developments in the common middleware and infrastructure (that build on the work of the EGEE project that LifeWatch investigated previously). Especially, investigate how the ORCHESTRA reference model used by LifeWatch maps to implementations in the e-Infrastructure sphere.
- Look into how development of other virtual laboratories (such as the Virtual Physiological Human or other virtual laboratories in the bioinformatics sphere), such that both communities can learn from one another. Investigate whether there are environments or tools that can be shared by these communities (perhaps related to more generic modelling or simulation tasks) to avoid repetition of effort.

A.13 *Preparing for SKA – Square Kilometre Array*

The Square Kilometre Array (SKA) http://www.jb.man.ac.uk/prepska/ will be a revolutionary radio telescope made of thousands of receptors linked together across an area the size of a continent. The total collecting area of all the receptors combined will be approximately one square kilometre, making the SKA the largest and most sensitive radio telescope ever built. SKA will be built in either South Africa or Australia – New Zealand; the final selection will be made in 2012. The SKA will address five fundamental questions[18] :

- *How do galaxies evolve and what is dark energy?* The expansion of the Universe has been attributed to a mysterious dark energy. The SKA will investigate the expansion of the Universe after the Big Bang by mapping the cosmic distribution of hydrogen. The map will track young galaxies and help identify the nature of dark energy.
- *Are we alone?* The SKA will be able to detect very weak extraterrestrial signals and will search for complex molecules, the building blocks of life, in space.
- *How were the first black holes and stars formed?* The SKA will look back to the Dark Ages, a time before the Universe lit up, to discover how the earliest black holes and stars were formed.
- *What generates the giant magnetic fields in space?* The SKA will create three-dimensional maps of cosmic magnets, to understand how they stabilise galaxies, influence the formation of stars and planets, and regulate solar and stellar activity.
- *Was Einstein right?* The SKA will investigate the nature of gravity and challenge the theory of general relativity.

There are several issues that need to be addressed before construction of the SKA can begin: What is the design for the SKA? Where will it be located? What is the legal framework and governance structure under which it will operate? What is the most cost-effective mechanism for the procurement of the various components? How will the SKA be funded?

The purpose of the PrepSKA project is to address all of these points. PrepSKA will integrate R&D work from around the globe to develop a fully costed design for the first phase of the SKA, and a deployment plan for the full instrument. With active collaboration between funding agencies and scientists, all of the options for the policy-related questions will be investigated. The principal deliverable will be an implementation plan that will form the basis of a funding proposal to governments to start the construction of the SKA.

The transfer and handling of the data captured by SKA, as well as work towards its governance and legal framework, are closely related to e-Infrastructure:

- The processing of the data captured by SKA will require vast computational power, and proposed solutions include new technology developed in partnership with the industry. It could be useful to identify possible interactions with the infrastructure that will be built by the PRACE petaflop computing initiative.
- Data processing will certainly require both high-performance computing resources and techniques, including the use of grid computing methods dominant in high-end computing for the physical sciences. In this context, it would be useful to look into progress in the European Grid infrastructure offered by EGI and the status of grid middleware as pursued by EMI.

18 http://www.skatelescope.org/PDF/100510_Journalist_Factsheet.pdf

- In the effort to derive a plausible governance and legal framework for SKA, PrepSKA could benefit by the experience gained from the efforts that led to the establishment of EGI, where governance models for multinational collaborations were investigated and different legal frameworks were analyzed.
- Coordination of the e-VLBI EC FP7 project NEXPReS with other FP7 radio astronomy initiatives such as RadioNET and PrepSKA could be exploited. The commissioning and operation of international LOFAR stations will build up a large numbers of long baselines to the LOFAR core relying on GÉANT and other e-Infrastructures. To maximize science return, further development of e-Infrastructures are needed, as are research activities focused on distributed methods for stream processing and data storage.
- Fostering of cooperative relationships between radio astronomers and other physicists via relevant EC actions. The demands for energy-efficient real-time processing are common to astronomy and astro-particle projects like SKA, CTA and KM3NET, but also to a wide range of other physics projects on the ESFRI Roadmap. These projects also face common problems in areas such as data handling and industrial engagement, to which common approaches may prove fruitful.
- Development of links between radio astronomers and the widest possible range of research areas featured in the ESFRI Roadmap. Radio telescopes are attached to sensor networks, often deployed in challenging environments such as deserts, that can be shared with other sensors informing on research into the environment (as has been demonstrated by LOFAR), as well as life and social sciences. Direct sharing of networks and cooperative research can exploit synergies in data processing, management and analysis. EGI structures may be useful in coordinating this activity.

A.14 European XFEL - European X-ray Free Electron Laser

European XFEL aims to construct and operate an X-ray facility allowing scientists to map the atomic details of viruses, decipher the molecular composition of cells, take three-dimensional images of the nanoworld, film chemical reactions and study the processes in the interior of planets. The XFEL project started in 2008 and the infrastructure is expected to be operational in 2015. The research facility is currently under construction in the Hamburg area, Germany. From 2014 on, it will generate extremely intense X-ray flashes to be used by researchers from all over the world. The constructed facility will be more than tenfold bigger than the currently utilized FLASH[19] infrastructure, the biggest X-ray facility in the world.

The infrastructure will address following research activities:

- The European XFEL will help to better understand the structure of biological cells and create new materials with optimized properties. Deciphering such tiny structures becomes possible thanks to the short wavelengths (of the order of atoms) and the laser-like properties of the X-ray flashes.
- Using the X-ray flashes, scientists can film ultrafast processes such as the formation of molecules. Research at the European XFEL will thus enable better understanding of chemical processes, which will aid development of more efficient industrial production processes, for instance. These studies also provide an essential basis for the development of new medicines.
- The X-ray flashes enable researchers to make films of ultrafast processes because each single flash is less than 100 trillionth of a second long and brilliant enough for the snapshot.
- Using the X-ray flashes, scientists can study matter under extreme conditions. The flashes are so intense that they can be used to create pressures and temperatures similar to those in the interior of planets. At the European XFEL, these states of matter can thus be investigated in detail. In addition, studies of how single atoms behave in the intense light flashes will provide new methods for X-ray physics.

XFEL is anticipated to attract scientists from a broad spectrum of sciences including biology, medicine, chemistry, materials science, astrophysics, energy research, electronics and nanotechnology.

It is possible to identify a number of specific e-Infrastructure-related issues that have to be addressed in the next few years, including:

- The XFEL project shares technological challenges similar to those addressed by EGEE/EGI. Especially important are: data storage and management, as the anticipated data stream will be counted in GB per second, and distributed computing to process the data. Work with EGI or NGIs representing partner countries can be crucial to investigate whether the solutions developed by the EGEE project can be implemented in XFEL project, as well as avoiding repetition of research and implementation efforts. There are also other projects striving (e.g. EXPRESS II) to solve similar problems, therefore establishing liaison might be beneficial for all partners.
- The data handling issues mentioned earlier might result in implementation of a distributed model for storage. In such a case, it is desirable to coordinate the efforts with TERENA/GÉANT (who work on similar issues and task forces) to ensure proper connections between partners.
- As the XFEL will have unique capabilities, it is wise to investigate the possibility of sharing resources by means of virtual laboratory environments. Currently ongoing research in this area should be investigated along with the possibility of employing experience gained in other areas.

19 http://zms.desy.de/research/photon_science/free_electron_laser_flash/index_eng.html

APPENDIX B: List of contacts

e-IRG chair

Dr. Leif Laaksonen

CSC - IT Center for Science Ltd.
P.O. Box 405
FI-02101 Espoo
Finland

Phone: +358-9-4572378
Fax: +358-9-4572302
E-mail: chair [.at.] e-irg.eu
Leif.Laaksonen [.at.] csc.fi

e-IRG secretariat[20]

c/o CSC - IT Center for Science Ltd.
(Keilaranta 14)
P.O. Box 405
FI-02101 Espoo
Finland

Phone: +358 9 457 2281
Fax: +358 9 457 2302
secretariat [.at.] e-irg.eu
http://www.e-irg.eu/contact.html

European E-Infrastructure Forum (EEF)

http://www.einfrastructure-forum.eu/
E-mail: members@einfrastructure-forum.eu

20 The e-IRG secretariat will be moving to the Netherlands in December 2010. Web-site and e-mail addresses will be still valid after this migration.

APPENDIX C: Glossary

AAI – Authentication and Authorisation Infrastructure 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 networks refer 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 are one the main authentication mechanisms used for 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.

DCI – Distributed Computing Infrastructure such as EGEE, EGI, DEISA, PRACE, etc.

DEISA - Distributed European Infrastructure for Supercomputing Applications is a series of FP7 EC-co-funded projects interconnecting major high-performance computers around Europe.

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).

EEF - The European e-Infrastructure Forum is a forum for the discussion of principles and practices to create synergies for distributed Infrastructures. Its current membership includes GÉANT, TERENA (research networking), EGEE, EGI (grid computing), DEISA and PRACE (high performance computing).

EGEE - Enabling Grids for E-sciencE is a series of FP7 EC-co-funded projects interconnecting more than 100,000 computers in Europe and beyond. EGEE serves e-Science. When EGEE-III ended in April 2010, EGI took over the current infrastructure (supported by the EC-co-funded EGI-InSPIRE project).

EGI - European Grid Initiative is the next phase in the implementation of capacity computing in Europe. EGI unites the resources of the NGIs, guaranteeing trans-national access to data and services.

EGI.eu is the legal body that hosts the EGI headquarters. It includes personnel with central responsibility, as well as the management structure. EGI-eu is located in Amsterdam (after a bidding process) and the EGI-eu team is currently being recruited.

e-Infrastructure or electronic infrastructure covers ICT-related infrastructure, encompassing, among others, 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).

EMBL - European Molecular Biology Laboratory is a major research centre coordinating molecular biology research.

EMI - The European Middleware Initiative is a close collaboration of the three major middleware providers in Europe -- ARC, gLite and UNICORE -- with other software providers. It will deliver a consolidated set of middleware components for deployment in EGI (as part of the Unified Middleware Distribution - UMD), PRACE and other distributed computing infrastructures, as well as extend interoperability and improve integration with emerging computing models, strengthen the reliability and manageability of the services, and establish a sustainable model to support, harmonise and evolve the middleware, ensuring it responds effectively to the requirements of the scientific communities relying on it.

ESA - The European Space Agency is Europe's gateway to space. Its mission is to shape the development of Europe's space capability and ensure that investment in space continues to deliver benefits to citizens of Europe and the world.

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 and digital communications.

ESO - the European Organisation for Astronomical Research in the Southern Hemisphere is an inter-governmental organisation operating three unique world-class observing sites in Chile.

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

eVLBI - Electronic Very Long Baseline Interferometry is the use of high speed networks to connect radio telescopes separated by large distances (100-1000s of km) instead of the traditional method of recording onto magnetic tape and shipping the recorded data to a central correlator. eVLBI is part of a collaboration between major radio astronomical institutes in Europe, Asia and South Africa, performing high-angular resolution observations of cosmic radio sources.

A **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 data 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 organization 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 centralized 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.

ICT is the standard abbreviation for Information and Communication Technologies.

IEEE - Institute of Electrical and Electronic Engineers is the world's largest professional association dedicated to advancing innovation and technological excellence through its highly cited publications, conferences, technology standards, and professional and educational activities.

IETF - Internet Engineering Task Force is a large, open, international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet.

IGE - Initiative for Globus in Europe 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).

IPR - Intellectual Property Rights 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, each of which apply to different created items: copyright, designs, patents, trademarks, protection from passing off and protection of confidential information.

International Thermonuclear Experimental Reactor (ITER) is a joint international research and development project that aims to demonstrate the scientific and technical feasibility of fusion power. Fusion is the energy source of the sun and the stars. Fusion research aims to demonstrate that this energy source can be used to produce electricity on Earth in a safe and environmentally benign way, providing abundant fuel resources to meet the needs of a growing world population.

ITU – The International Telecommunication Union is the major standardisation body for ICT issues.

The **LHC – Large Hadron Collider** is a major RI facility located at CERN in Geneva, Switzerland.

A **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 wavelength division multiplexing. 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.

Middleware has different meanings in different contexts, but normally denotes specialised, high-level software that sits between applications or between the application layer and lower layers in a layered software architecture.

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.

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.

OGF - Open Grid Forum is an open community committed to driving the rapid evolution and adoption of applied distributed computing.

OpenID is an open, decentralized 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 log-in 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 an initiative aiming to implement 3-5 petaflop supercomputing systems in Europe. PRACE manages extreme computing power and a select set of highly specialized services.

A **repository** is 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 one type of repository.

RFID - Radio-frequency identification refers to the use of an object (typically referred to as an RFID tag) in a product, animal, or person for the purpose of identification and tracking using radio waves. Some tags can be read from several meters away and beyond the line of sight of the reader.

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 to collaborate, innovate and share knowledge that fosters the development of Internet technology, infrastructure and services to be used by the research and education community.

VOs - Virtual Organisations are communities facilitated on-line and spanning multiple administrative and geographical boundaries. Often used in grid computing. Also referred to as Virtual Research Communities (VRCs).

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.

Web2.0 is a term commonly associated with web applications that facilitate interactive information sharing, interoperability, user-centreed 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 e-Infrastructure Reflection Group

e-IRG secretariat
P.O. Box 405, FI-02101 Espoo, Finland
Phone: +358(0)9 457 2281, Fax: +358 (0)9 457 2302
secretariat@e-irg.eu

Visiting address
Life Science Center,
Keilaranta 14, Keilaniemi
Espoo, Finland