Seeking new horizons: EGI's role for 2020

Abstract

The European Grid Infrastructure (EGI) was established in 2010 as a European-wide federation of national computing and storage resources for multiple research communities as a result of over a decade of investment by national governments and the European Commission. EGI now supports over 18,000 researchers across over 15 research disciplines through a uniform set of services in over 350 resource centres in all countries across Europe that support over 1 million jobs a day and the related data access and movement. EGI is centred on a persistent coordinating legal entity, EGI.eu, which is governed by the national stakeholders and early-adopting international research communities.

This strategic plan outlines the initiatives that can take place within the EGI community over the next two years, supported through the FP7 EGI-InSPIRE in the first instance and subsequent related projects that will, in the future with further investment, develop EGI's strengths in:

- European-wide coordination and interaction with research communities and national resource infrastructure providers
- Coordination, maintenance, operation and delivery of an open uniform European-wide federated production infrastructure
- Developing and promoting technologies for federating new resources
- Supporting the integration and operation of scalable interdisciplinary Virtual Research Environments personalised to each research community

This strategy will with continued investment from national and European funding bodies, evolve EGI's activities to be a key enabling foundation of the online European Research Area (ERA) which is part of the European Commission's Innovation Union[1] initiative. Its contribution will be to provide the transnational multi-disciplinary research collaborations within the ERA with a world class e-Infrastructure able to support innovative collaborative virtual laboratories for simulation, data sharing and data analysis activities that is sustainable for decades to come.

This document (D2.30) will be updated annually and provides EGI's Strategic Plan which is targeted at European and national policy makers and senior managers in resource providers, virtual research communities and other stakeholders within the EGI Ecosystem. Additional information can be found in: the EGI Technical Roadmap[2] which details work taking place within the EGI-InSPIRE project, and the EGI Business Model[3] which describes the value generation activities of different components in the EGI Ecosystem.

[1] http://ec.europa.eu/research/innovation-union/index_en.cfm?pg=home
[2] D2.31 – PM24 Deliverable Reference to be added when available
[3] D2.18 – PM21 Deliverable Reference to be added when available

1

2 **Document History**

Version	Date	Author	Comment
0.1	10/2/12	Steven Newhouse	Initial draft for internal EGI.eu discussion.
0.2	16/2/12	Steven Newhouse	Revised following internal feedback for wider review.
1.0	17/2/12	Steven Newhouse	Revised following internal review for EGI-InSPIRE PMB discussion.
1.1	7/3/12	Steven Newhouse	Revised following comment from EGI-InSPIRE PMB for public comment.

3

Table of Contents

Glossary

EGI - European Grid Infrastructure: A federation of shared computing, storage and data resources from national and intergovernmental resource providers that delivers sustainable, integrated and secure distributed computing services to European researchers and their international partners.

NGI - National Grid Infrastructure: The national federation resources which is coordinated through a single point of contact that has an exclusive mandate to represent its national grid community in all matters falling within the scope of EGI.

VRC - Virtual Research Community: A group of large-scale research collaborations, or a number of separate Virtual Organisations (VOs) grouped according to research domain or computational technique. The group shares information and experience in achieving their goals through the usage of an e-Infrastructure (e.g., best practices, applications, training material).

ERA - European Research Area: The area that brings together all of the European Union's (EU) resources to better coordinate research and innovation activities at the level of both the Member States and the Union. The area also aims to achieve a major ambition of the EU: to arrive at a truly common research policy.

DAE - Digital Agenda for Europe: One of the seven flagship initiatives of the Europe 2020 Strategy set to deliver sustainable economic and social benefits from a digital single market based on fast and ultra fast Internet and interoperable applications by 2020.

EMI - European Middleware Initiative: An FP7 project that provides a software platform for high performance distributed computing, namely grid middleware distributions that are used by scientific research communities and distributed computing infrastructures worldwide.

IGE - Initiative for Globus in Europe: An FP& project that provides a distribution and customisation of the Globus Toolkit in Europe for European e-Infrastructures.

EC - European Commission: The executive body of the European Union responsible for proposing legislation, implementing decisions, upholding the Union's treaties and the general day-to-day running of the Union. The EC allocates part of the EU budget to companies and organisations in the form of calls for tender, grants or funds and other financing programmes.

EIRO - European Intergovernmental Research Organisation: A legal organisation and member of the EIROForum that has extensive expertise in the areas of basic research and the management of large, international infrastructures, facilities and research programmes.

For further terms and definitions see https://wiki.egi.eu/wiki/Glossary_V1.

1 Executive Summary

E-Infrastructures are electronic services which integrate physical computing, storage, networking and other hardware to connect researchers from all disciplines with the reliable and innovative ICT services for uniform access to commodity computing, archiving, and management of distributed data that they need to undertake their collaborative world-class research. By 2020, this capabilities need to be deployable on demand in order to provide a foundation for the online European Research Area (ERA). The physically distributed hardware deployed within national infrastructures is connected through high-speed networking to support the activities of its research communities of up to 1.8 million publicly funded and 1.0 million privately funded researchers spread across Europe.

The **grand challenges in science and in society** that need to be solved in 2020 and beyond will increasingly require both geographical and intellectual collaboration across multiple disciplines. The European Grid Infrastructure (EGI) is the result of pioneering work that has, over the last decade, built a collaborative production infrastructure of uniform services through the federation of national resource providers that supports multi-disciplinary science across Europe and around the world. Through this initiative an ecosystem of national and European funding agencies, research communities, technology providers, technology integrators, resource providers, operations centres, resource centres, coordinating bodies and other functions has emerged to serve over 18,000 researchers in their intensive data analysis. Further targeted investment in the infrastructure platform now provided by EGI will develop this ecosystem as a foundation for other research communities to build upon.

The **EGI ecosystem** provides a common foundation upon which the online ERA can be established by building upon the strengths that have been developed over the last decade:

- **An ecosystem that promotes competitive cooperation, collaboration and interaction at local, national and European levels.** Engagement with technical users and researchers can be enhanced through the support of local 'community champions', national and European events and workshops that promote EGI and its activities within the ERA.
- **A monitoring and support infrastructure for locally deployed, domain specific services** that has been proven to scale to over 350 resource centres across Europe and around the world. This would allow a research community to monitor and manage their own the services operating at its distributed facilities.
- **Consistent uniform access across Europe to institutional clouds by** through an open extensible solution that builds upon technologies that have been developed to federate access to institutional computing clusters to provide a European-wide *Infrastructure as a Service* cloud model that spans different technologies and administrative domains. Researchers will be able to use public or commercial resources to run services dynamically where their data is being generated or stored through the same programmatic interfaces.
- **Reducing the technical barriers to accessing EGIs resources** by investing in virtual research environments composed of open extensible and reusable software solutions that simplify access across desktop and mobile portals and applications to services deployed across the EGI that can be rapidly and consistently customised by individuals or communities themselves for their researchers' applications.

The **sustainability of the EGI community** as a whole can only be achieved by ensuring that the diverse components that make up the rich open loosely coupled ecosystem are themselves sustainable within the governing context set by the EGI Council. This means that the individual components in the ecosystem (whether projects or organisations) delivering these activities may have multiple providers, which can evolve and possibly be replaced over time as they establish long-lived processes with the components they are dependent on. Funding is certainly a key element of sustainability; but the funding for innovation leading to improvements in the operational efficiency and technical capabilities needs to be decoupled from the funding that is needed to establish and operate the persistent coordinating structures and physical infrastructure within the ecosystem. Many of the ecosystems components can be foreseen to have a service delivery activity – sustained frequently by the community that drives the most value from that activity – but where the innovation in this service delivery is funded, when required, through short-term focused projects driven by community need and supported by European and national funding bodies.

As a result of this analysis, EGI will build its future on a vision, mission and core values of:

Vision: To support the online European Research Area through a pan-European research infrastructure based on an open federation of reliable services that provide uniform access to national computing, storage and data resources.

Mission: To connect researchers from all disciplines with the reliable and innovative ICT services they need to undertake their collaborative world-class research.

Core Values:

- **Leadership**: EGI is a leading pan-European infrastructure, integrating worldwide computing, storage and data resources to support an economy built on innovation and knowledge transfer.
- **Openness**: EGI operates with a transparent governance structure that integrates the views and the requirements of all stakeholders, from research communities to resource providers.
- **Reliability**: EGI provides a reliable infrastructure that research communities can depend on to collaborate with their peers and deliver innovation.
- **Innovation**: EGI will continue to meet the needs of research communities operating at unparalleled geographic and technical scale by partnering to bringing new technologies into production.

Such an ambitious but coordinated investment in innovation will enable the rapid evolution of the EGI to support deployment and operation of services customised for the needs of individual research communities of all sizes and at a European scale.

2 The European Grid Infrastructure

2.1 Background

The European Grid Infrastructure (EGI) has its origins in the pioneering distributed computing research that took place in the early 2000's as the opportunities presented by federating organisational resources to compute and data oriented applied researchers emerged. Through the European Data Grid (EDG: 2001-2004) and three successive Enabling Grids for E-sciencE (EGEE: 2004-2010) projects, Europe has established a world-leading infrastructure that integrated resources across different administrative domains to support a multi-disciplinary research community. Alongside these core technology development and operational deployment projects, focused projects grew the community in the Baltics, South East Europe, and built collaborations with areas outside Europe in North America, Latin America and Asia Pacific.

The successful move from research activity to production infrastructure supporting thousands of applied researchers that integrated such technology into their daily research activities triggered discussions about how to move away from a short-term project structure. As a result of a community driven EC funded design study project – the European Grid Initiative Design Study (EGI-DS) – a transition from a regional operational model to a national operational model was established. This work started during the final phase of EGEE and has been completed within the EGI-InSPIRE project (2010-2014) that now provides continued community building and operational support to EGI and its regional partners.

The monolithic infrastructure that was established during the EDG and EGEE research and development projects has been evolving. The operation of the infrastructure, the technology providers to the infrastructure and the use of the infrastructure by applied researchers now take place through separate independent but loosely-coupled activities. These activities are coordinated through a dedicated coordinating organisation (EGI.eu[4]), which is governed by the National Grid Infrastructures (NGIs), established to coordinate and manage their national resources and some of the European Intergovernmental Research Organisations (EIROs) that directly use the infrastructure.

EGI currently comprises over 350 resource centres, in over 35 countries supporting 18,000 users submitting over 1 million jobs a day over more than 15 different disciplines.

2.2 Ecosystem

The EGI ecosystem continues its evolution from a monolithic entity established in EDG and EGEE a decade ago, to the independent technology providers, resource infrastructure providers, research communities and the national (NGIs) and European (EGI.eu) coordination bodies that have established themselves over the last few years (see Figure 1). Other roles have yet to formally emerge from within the community's current activities (e.g. the clear definition of the platform integration role within the software development activities or the platform operations role) but their roots can be found within EGI's current portfolio of activities.

[4] http://www.egi.eu

A healthy sustainable ecosystem, i.e. with the capacity to endure, implies having a structure where some components can fail and be replaced by others without damaging the ecosystem as a whole. The different individual components of the ecosystem must therefore be able to be individually developed, maintained, supported and allowed to establish effective processes with each other that enable them to scale as required - how and to what extent will depend on their value. Opening up the processes between the components in the ecosystem will enable greater scalability as there are then fewer interactions that need to be controlled centrally. The role of the EGI Council (and EGI.eu) in such an ecosystem would be to identify the components and coordinating support actions needed to establish such an open ecosystem that can sustain itself and the community it supports.

The EGI ecosystem at a high-level comprises:

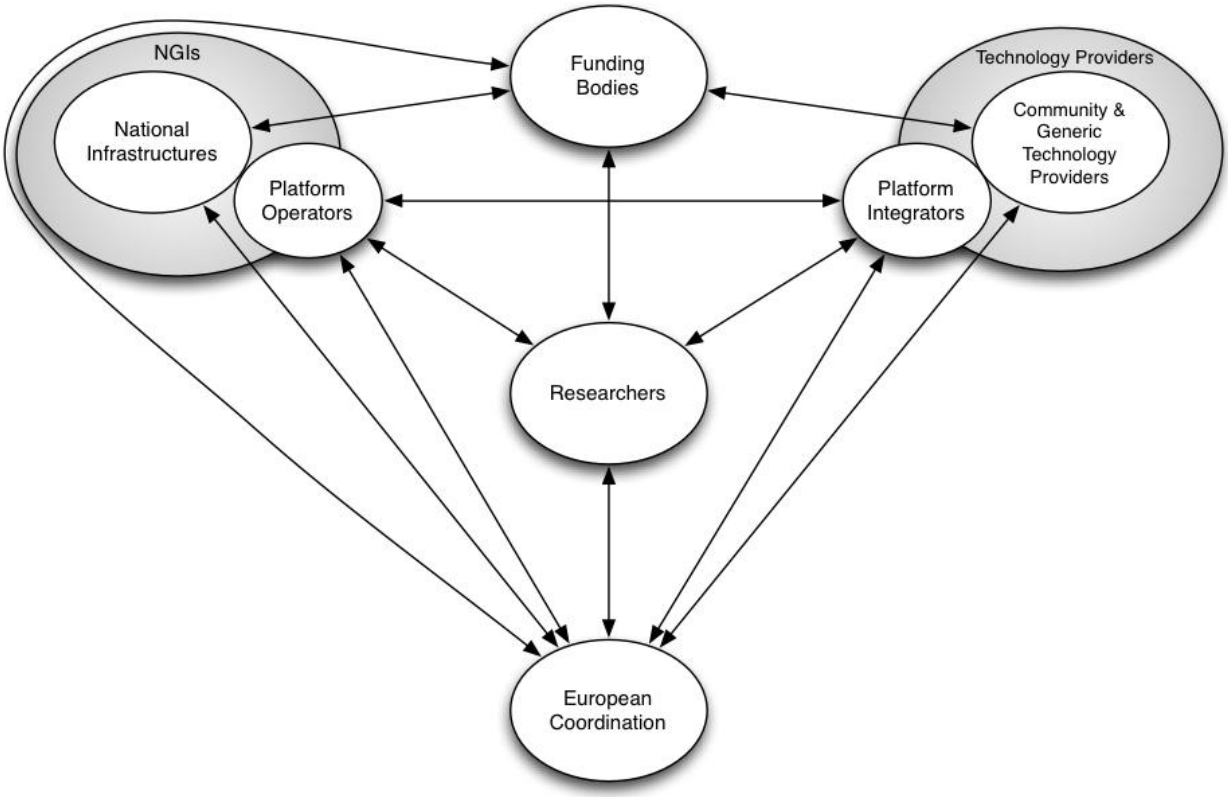

Figure 1: Overview of the EGI Ecosystem

- **Researchers:** The *individual researcher*, who is interested in exploiting the available e-Infrastructure for their research - 'me-science' – by using whatever they can rapidly adapt or integrate to meet their needs in order to publish first and gain the recognition of their peers. At a larger scale, the *research collaboration*, while composed of individual 'me-science' researchers, and inheriting many of their self-centred goals, may have sufficient critical mass and coordination that they are able to contribute to and to a limited extent sustain their own community around shared resource goals needed to tackle societal challenges. *Virtual Research Communities* (VRCs) are composed of research groups that span different disciplines in many organisations across different countries that have structured themselves to tackle a 'grand challenge' within their own scientific community.

- **Technology Providers**: The technology area within the EGI ecosystem is built upon open-source or commercial software coming from *community* and *generic technology providers* that is put together by *platform integrators* to meet the needs of particular user groups. For instance, the EMI project[5] integrates a platform for high-throughput computing from software that is developed within the project primarily for the EGI community (i.e. community technology providers such as EMI) with software developed outside the EGI community (i.e. generic technology providers such as Apache) to meet particular use cases coming from their target research community (e.g. WLCG).
- **National Infrastructures**: These include the *NGIs* that represent national activities within EGI and undertake national coordination duties through the *resource infrastructure provider* role. They have the responsibility to manage and deliver the operational infrastructure coming from the individual *resource centres* within the country. They may also have the responsibility to act as *platform operators* ensuring that any community specific services provided by the resource centres are operating effectively or this responsibility may be undertaken centrally by that community.
- **European Coordination**: Within the EGI ecosystem, the community has recently established an independent legal entity (EGI.eu) to provide *European Coordination* through a defined governance structure and to coordinate on behalf of the community the community's activities.
- **Funding**: The primary source of EGI funding has come from the public sector through *national funding bodies* with additional investment from *European funding bodies*, such as the EC, to support European level integration and structuring. *Commercial organisations* remain a largely untapped source of funding due to legal and logistical concerns and are more likely to be delivering services to the EGI ecosystem in the future rather than purchasing services from the EGI ecosystem. Within EGI a *community funding* scheme has been established for EGI.eu where organisations with the community that benefit from EGI contribute to the coordination costs.

2.3 Stable Operation with Innovative Technology

The ICT used to build EGI has changed radically over the last ten years and will continue to do so in the future. One of the most significant achievements of the EGI community over the decade is to establish a community ethos of bringing new innovative technologies into operation on a European, and through its international collaborations, on a global scale to ensure an interoperable worldwide infrastructure. This work has brought worldwide recognition of European activity. The challenges presented by new requirements from both the research communities and the resource infrastructure providers will drive further innovative research, development and deployment in the future.

The focus within EGI-InSPIRE on the operational infrastructure has moved much of the innovation into other related projects (e.g. EMI and IGE[6]). As a result, over 90% of the funding provided for the purchasing and operation of the physical infrastructure within EGI comes from national sources. This national support is also applicable to the coordinating functions provided by EGI.eu where over half the running costs come from the community. The sustainable management of the infrastructure requires the separation of funding for operation, from the higher-risk investment needed for innovation and development. The operational costs need to be supported by those that directly benefit from the

[5] http://www.eu-emi.eu/
[6] http://www.ige-project.eu/

infrastructure – the national research communities represented within EGI by the NGIs, the international research collaborations by the EIROs and in the future the ESFRI organisations.

2.4 Analysis

EGI faces a number of distinct challenges as it prepares for its next decade of operation:

- **Sustainability**: A key aspect of the EGI community's sustainability strategy is now evident in EGI.eu, the coordinating organisation established to provide a persistent governance function across its many activities. However, EGI.eu's sustainability is dependent on the sustainability of its participants – the European NGIs and EIROs. How can the adoption of EGI's resources within more research communities be achieved so that NGIs can play a pivotal national role in providing services around platform integration, platform operation and technology provision to new and current research communities?
- **Adoption**: Currently, EGI supports over 18,000 users but there are 1.8 million publicly funded and 1.0 million privately funded researchers in Europe (See Annex 1), a significant proportion of which are facing intensive data analysis challenges similar to the research communities currently supported by EGI. How can EGI's operational infrastructure and expertise enable the technologies that these researchers need to deploy at a European scale to support their intensive data analysis?
- **Technology**: EGI's experience in providing controlled secure access to different resources is unparalleled. How can new technologies and operating models provided by virtualisation and cloud computing be merged with the federation technologies developed over the last decade and the private clouds appearing within the NGIs to help accelerate the establishment of the online European Research Area?

In order to successfully address these challenges it is vital to establish the strengths, weaknesses, opportunities and threats around the individual components (See Annex 2) and of the EGI ecosystem as a whole.

Strengths:

- Enables researchers able to use computation and data services to perform data analysis that would otherwise be impossible on local resources or to undertake the analysis in less time.
- Provides a European-wide infrastructure that is built from a secure and integrated federation of national grids that provides secure and seamless access to services for researchers in Europe and internationally.
- Strong community-oriented governance and management functions.
- Collaborative community spirit facilitated by technical services (application database, training marketplace, etc.) and human interaction (forums, workshops, blogs, newsletters, etc.).

Weaknesses:

- No clear stand alone service and product portfolio that can be offered in isolation to individual research communities.
- Current functional services appear tailored to only a few specific research communities.

- No clear top-down resource allocation model or process comparable to commercial cloud services due to the bottom-up ownership and allocation of resources directly to specific local research communities.
- Not oriented around the needs of individual researchers or small research collaborations.
- Technical barriers are perceived to be high, leading to a slow uptake with research communities not already using the infrastructure.

2.5 The Future

The transition that has taken place in EDG, through EGEE to EGI of a small experimental activity to a production-quality infrastructure that enables research is complete. The move away from a project-based structure to a sustainable coordinating organisation (EGI.eu) has provided a clearer managerial and operational focus, but it is clear in 2012 that the engagement of research communities with e-Infrastructure providers and the individual technology providers still needs to develop.

The next decade presents opportunities that build on EGI's existing strengths by:

- Providing a key e-Infrastructure that will be the foundation of the online ERA.
- Attracting further research communities by moving to a more generic, generally applicable set of service offerings that individual researchers can customise for their research activities or those of their community.
- Demonstrating the economy of scale and customisation of ICT services to the research community compared to commercial solutions as utilisation can be optimised at a local, national and European perspective.

The strategy that follows details how these opportunities can be exploited and provides mitigation around the key threats to the EGI community around:

- An inability to respond promptly to the needs of new research communities who switch to other publicly-funded or commercial service or their own local resources which may then be underutilised.
- A lack of a shared vision and continued funding across participants may lead to community fragmentation and an e-Infrastructure divide across Europe.

3 Europe 2020

3.1 Strategic Impact on EGI

EGI's strategy needs to be aligned with the EC's vision for Europe in 2020. For EGI, the two important flagship initiatives defined in the Europe 2020 strategy are the Digital Agenda for Europe (DAE) and the Innovation Union (IU).

The Digital Agenda for Europe[7] offers opportunities for EGI to:

- Provide a single uniform market for accessing distributed computing resources and connected data in Europe through EGI's federation of national resource providers.
- Promote competitiveness and interoperability through open standards within the European Interoperability Framework (EIF).
- Reduce inefficient research spending and stimulate innovation across Europe by maximising the utilisation of federated national resources and knowledge within a common infrastructure.
- Offer large scale ICT facilities that enable the exploration of new computing and data processing models that address scientific grand challenges facing society.

In the context of the Innovation Union, EGI can:

- Position itself as a key enabler for the online ERA for the free circulation of researchers, knowledge and technology.
- Promote excellence in education and skills development by simplifying multi-disciplinary cooperation.
- Bridge geographical boundaries beyond Europe thanks to the many collaborations and integration with worldwide e-infrastructures.

Overall, by contributing to the Europe 2020 strategy, primarily through the establishment of the online ERA, EGI is capable of generating a substantial socio-economic impact for European society[8].

3.2 What is the online European Research Area?

The ERA is the mechanism adopted by the European Commission to provide the driving force for innovation within Europe by 2020 through the Innovation Union initiative. Bringing the ERA online, through the increasing adoption of ICT, is an essential component of this initiative in response to the increasing digitisation of research and the data-deluge being generated from large-scale European Research Infrastructures down to individual researchers. The integration of the different public and private stakeholders in the innovation process through ICT can reduce the 'time to market' of new products and services, as well as producing new innovations by reducing the barriers to collaboration between different communities.

[7] http://ec.europa.eu/information_society/digital-agenda/index_en.htm
[8] https://wiki.egi.eu/wiki/Europe_2020_actions

The foundation of this ICT integration in the ERA – the online ERA – needs to build on the DAE and IU principles and be based on an open, collaborative, sustainable e-Infrastructure ecosystem that supports distributed data analysis:

- **Open**: Open processes and an open architecture enabled through standards that allow the integration and flexible composition of the best services to meet the needs of each individual research communities regardless of the supplier is essential in delivering excellence.
- **Collaborative**: Researchers in different locations and disciplines connect with each other in order to access and share data and knowledge, and will need to access the electronic resources distributed across Europe to undertake their research.
- **Sustainable**: Researchers will become increasingly dependent on ICT services to collaborate across all areas of their research lives, and will need to be assured that these services will be present for as long as their research continues.
- **E-Infrastructure**: The electronic services that can be flexibly and easily deployed to integrate the physical computing, storage, networking and other hardware that supports the intensive data analysis needed by Europe's research community.
- **Ecosystem**: A successful ecosystem in one where its individual components deliver value to each other and are kept in balance by appropriate processes that provide feedback and governance.

3.3 A Personalised EGI

Allowing individual researchers and research collaborations to customise and therefore personalise the services they have access to when using EGI's resources will be critical in broadening uptake across the diverse research communities that comprise the ERA.

Research computing strategies that the online ERA is building upon are split into two broad categories: approaches that serve the niche requirements of high-end science (~1%) of which there are perhaps a small number of research communities (<50) across Europe, and approaches that serve the 'long-tail of science'. This long tail includes a large number of smaller individual researchers and research collaborations and comprises the majority of the 2.8 million researchers in Europe. Both the niche and long-tail research communities can increase their productivity by being able to easily share their own physical, electronic and intellectual resources between different countries and disciplines using the expertise built up by the EGI community to process the large data sets generated by instruments or simulations.

Researchers need many ICT services to support the whole research lifecycle regardless of whether they work as individuals or in small or large research collaborations. However, the type of services that they require will vary depending on their research field and the scale of their collaborative activities. These services may range from the non-technical (e.g. bibliographic services, repository services, publishing services) to the technical (e.g. authentication services, data analysis services, workflow services, information services, data movement services) and the social (e.g. collaboration services, reputation services). These services need to scale either as individual instances or through interoperation with other instances across research communities of different sizes. EGI cannot expect to successfully scale its activities across all these areas. Therefore, it must establish an ecosystem that allows the researcher (or those acting on their behalf) to provide a personalised e-Infrastructure for their use.

EGI has the proven technological and operational experience to integrate both local and national resources into an integrated European environment. A federating Infrastructure Platform can be used by researchers to deploy their own services as part of a niche research community or by individual researchers as part of the long-tail of smaller research communities needing to undertake their multi-disciplinary collaborative research.

The technical revolution that has taken place around the delivery of ICT services in private enterprises though the adoption of cloud computing now needs to be reflected in the delivery of ICT services coming from the public sector such as EGI. This does not mean that EGI should automatically attempt to replicate all the services coming from the private sector, instead it should focus on the areas that are not being served on the basis of functionality or cost of commercial solutions. One consequence of these technological changes is that the direct users of ICT resources in the research community are expecting to see the flexible on-demand services coming from commercial providers available from the public sector but at reduced or zero cost and tuned to their specific workloads.

Refactoring EGI's production infrastructure around a Infrastructure as a Service (IaaS) offering built by federating the private clouds coming from its resource centres, would enable EGI to provide solutions to meet the needs of the research community which are not being met from commercial IaaS offerings by leveraging existing national investments in networking and hardware infrastructures. Such a platform would enable collaboration between different research communities and allow individual researchers, research groups and virtual research communities to integrate, deploy and operate their own domain specific Platforms as a Service (PaaS) for their individual research communities on a public sector IaaS.

4 Strategic Investment Areas

4.1 Overview

The analysis of the EGI ecosystem demonstrated a number of strengths that need to be reinforced and grown in the years to come to provide a basis for the online ERA. These are primarily:

- **Community & Coordination**: The network of national interfaces (the NGIs) into a European coordination body (EGI.eu) that provides governance to the community and the continued development of that community through communication, outreach, support and marketing events.
- **Operational Infrastructure**: EGI federates an operational infrastructure comprising over 350 resource centres that has been proven to work at a large-scale. As other research communities seek to build European-wide operational infrastructures this national network of services and people provides a key asset for EGI to offer for them to build upon.
- **Technology**: For the last decade, EGI has been growing its supported research communities and developing its operational infrastructure around a single domain specific technology stack. EGI needs to move to a model where research communities and individuals are easily able to personalise their e-Infrastructure environments and a federation of private clouds is being prototyped to deliver this capability.
- **Virtual Research Environments**: A key aspect to wider scale adoption of e-Infrastructures is the ability for individual researcher and research collaborations to personalise their environments. Virtual Research Environments require development in two main areas: the web-based user interface layer which will encompass desktop and mobile access, and investment in the domain specific platforms operated as a service for particular communities to undertake their research.

The following sections identify the individual strengths that EGI has in each area, present the opportunities that exist for development and give an overview of the plans during the timescale of EGI-InSPIRE (with more details in the EGI Technical Roadmap) and the funding-dependent plans that could continue after EGI-InSPIRE.

4.2 Community & Coordination

4.2.1 Strengths

The EGI has developed from a domain-specific activity focused on a handful of resource centres to a multi-disciplinary infrastructure with usage across many different disciplines. This has been achieved by developing a broad multi-disciplinary community across many institutions and roles at a national level that has been federated and integrated across Europe. This community continues to evolve as the infrastructure scales out from being an internally focused research activity to offering professional services to other research communities. This is a process that must be supported to ensure that potential technical and political fragmentation is eliminated and uniform access to national e-Infrastructure is maintained across Europe.

EGI's federated network of national resource providers, the NGIs, represent and integrate their resource centres into a single uniform infrastructure that spans Europe and around the world. This uniform service access benefits the researcher by providing a single market for services nationally and through the EGI

wide federation extends this uniform access to services across Europe. Individual researchers in one institution can collaborate nationally and internationally expecting the same uniform access to reliable services regardless of where they are located. The value of such widespread uniform services has been demonstrated during the initial data collection runs of the Large Hadron Collider in 2010-12 where scientists were able to rapidly analyse data and publish their results using services provided by EGI. The EGI Council is therefore in a position to define an ecosystem that promotes open competition, allows for failure and the replacement of individual components in the ecosystem, and yet retains a cooperative spirit that is inherent in the publicly funded research sector. The coordinated implementation of these policies is left to EGI.eu, which through the participation of NGIs and EIROs, provides a formal link to the national or domain-specific resource infrastructure provider and resource centres. EGI.eu also provides a focal point for collaboration with the other European e-Infrastructure providers (i.e. PRACE, DANTE, TERENA) and with comparable providers around the world. The whole EGI community comes together through bi-annual meetings that build collaborations and trust within the ecosystem and provide a venue for interaction between technology and resource providers, platform operators, and platform integrators in order to raise the profile of the research activities EGI supports.

4.2.2 Opportunities

EGI needs to build an ecosystem that can scale across the different research communities by:

- **Developing the technical expertise within the newer and smaller NGIs to support a borderless ERA**: Previous capacity building efforts within the community through BalticGrid and SEE-GRID projects helped develop the countries that were new to the EGI community. As the development of ERA progresses, it is essential that the e-Infrastructure that underpins the ERA within Europe spans all EU28 countries by investing in the development of the NGIs where needed to ensure that no country is left behind as EGI develops. **Example**: Exchange programme that embed experienced staff with less experienced staff in newer NGIs
- **Building and promoting an open EGI ecosystem**: The EGI ecosystem will need to rely on many independent service providers (i.e. technology providers, platform integrators, platform operators) if it is to successfully scale out to further research communities. Conferences that allow these independent service providers to meet each other, meet e-Infrastructure providers, provide feedback on the development of the infrastructure platform, pass on requirements from their research customers will be critical to future expansion and growth. **Example**: The EGI Technical and Community Forums provide 'professional developer' tracks (comparable to Apple's Worldwide Developer Conference, Java ONE, Microsoft's Professional Developer Conference) that talk about how the EGI services can be used.
- **Communication and Marketing within the EGI ecosystem***:* Growing the EGI ecosystem and its usage by the ERA is dependent on having a clearly defined service portfolio that enables a researcher's professional objectives and then ensuring that the research communities that could benefit from it know about it. Communication activities that report on the successes of all researchers in using the different e-Infrastructure and a marketing campaign targeting specific disciplines are specific to growing awareness. **Example**: The iSGTW weekly publication has over 8000 subscribers around the world and provides a platform for researchers using EGI and other e-Infrastructures to publicise their achievements within their own and to other communities such as the general public.
- **Supporting the governance and coordinated planning**: As the EGI ecosystem evolves understanding the technical requirements, managing the internal and external policy development

activity, and establishing and supporting a governance model across the different aspects of the community takes considerable effort. Part of this activity directly benefits the internal stakeholders represented within EGI through their participation in EGI.eu foundation. However, much of the focus is also around establishing an open ecosystem that supports the e-Infrastructure needs of the ERA and is aligned with the policy initiatives of the EC. **Example**: Providing briefing and planning documents to stakeholders. A network of international liaisons to bridge European and local human networks.

- **Building a grass-roots infrastructure for technical dissemination and support**: For EGI to achieve pervasive adoption across all countries and all research areas, concerted investment is needed to develop a grass roots human infrastructure for technical dissemination and support that expands the role, visibility and reach of an NGI within its country. Local 'champions', volunteers that are embedded in research communities or resource centres and are able to provide a human face to EGI. Building up the expertise in these individuals within an organised national and European wide structure based around the NGI's existing network can help promote EGI and the local NGI to new communities. **Example**: Campus champions schemes that have been established within Open Science Grid and XSEDE, and adopted in NGIs such as the UK National Grid Service.

4.2.3 Short-term Activities

The non-operational coordination that takes place within EGI has been restructured since the start of EGI-InSPIRE around NGI International Liaisons - also known as NILs. These individuals within the NGIs link activities in other NGIs or at a European level with work taking place within an NGI relating to, e.g. outreach, marketing, communication, training, new community engagement. They may draw on other resources within the NGI to form virtual teams designed to tackle specific community issues within a short-term scale.

During the remainder of the EGI-InSPIRE project the non-operational coordination structures need to be developed and scaled out to cover the resource centres (frequently located in research organisations) and the researcher that are using them. Some NGIs (picking up on experiences from the Open Science Grid and TeraGrid in the USA) have adopted a 'Campus Champions' model that provides local outreach contacts. This is a model that can be developed during EGI-InSPIRE and, if successful, can then be implemented with dedicated support to help EGI expand fully across the geographical and scientific scope of the ERA.

Demonstrating the scientific impact coming from the use of EGI's resources by different research communities needs to be improved. The current Virtual Organisation model means the strongest relationship lies between the individual researcher and their local resource centre. While individual resource centres may collect the publications resulting from the use of their facilities, this information is only just beginning to be collected nationally. Nationally collected data will need to be normalised and brought together for a European wide analysis of activity.

4.2.4 Long-term Goals

A successful open EGI ecosystem needs balanced investment between non-operational and operational activities. Such investment will provide the supporting framework to build a transnational borderless federated infrastructure by:

- **Supporting the governance functions through planning and policy development:** Maintaining cohesion across a federated infrastructure with over 35 national participants takes concerted effort in

governance, policy and planning. Such cohesion is critical if EGI is to continue as a supplier of integrated compute and storage capability to the researchers within the ERA. Legal entities such as the European Research Infrastructure Consortium (ERIC) may provide an effective mechanism that solidifies national support. This requires central effort that coordinates complementary national activities across all non-operational areas. Having a clear organisational focus within the community provides a clear point for collaboration within Europe and with bodies outside Europe.

- **Community building and collaboration through workshops and forums**: Continuing the bi-annual Forums is seen as an essential aspect of community building and human networking across disciplines and countries that needs to continue. In between these larger events subject-specific workshops provide the opportunity to discuss in detail many of the technical issues that can then be disseminated during the Forums. Although the meetings are financially self-supporting, central funding to provide secretariat, design and marketing support increases the professionalism and the quality of the meeting for the attendees, reduces registration fees and reinforces cost-effective integration across borders.
- **Both internal and external communication around EGI across all types of readers**: The successful activities of the e-ScienceTalk and related projects need to continue. Providing a professionally produced weekly online magazine and professionally produced communication materials is essential for raising the profile of EGI's activities across Europe and internationally to e-Infrastructure providers and the research communities and policy-makers across Europe.
- **Providing outreach to research communities new to EGI**: Significant adoption of EGI by the long-tail of individual researchers within the ERA will not take place by accident. While investment can be made in technological innovations that lower the barriers to adoption, researchers new to EGI still need to discover, understand and adopt these techniques. Complimentary approaches are needed: strategic outreach to research communities by engaging their thought leaders, providing exemplars that demonstrate reduced time to publication, and promotion at their meetings; and a grass-roots outreach by having a local 'EGI Champion' who can connect with local individual researchers and link them to the area of EGI that they need.
- **Technical support to research communities new to EGI**: Legitimate technical barriers to the adoption of EGI by researchers are to be expected – it would be very surprising if every use case from every researcher was satisfied by the existing e-Infrastructure. Some of these use cases will require targeted innovation that fits within a multi-year project structure. Other use cases can be tackled through short-term consultancy coming from EGI's already established network of national support structures within the NGIs to provide training for researchers, the support of the platform integrator. Building this capacity within EGI is critical to wider adoption of distributed computing within the ERA. New requirements from researchers already using EGI are handled directly by the operational infrastructure.
- **Integrated European e-Infrastructures**: Initiatives such as the European E-Infrastructure Forum (EEF) have helped the communication between the different European e-Infrastructure providers (PRACE, TERENA and DANTE) but closer organisational and service delivery alignment remains an area for further discussion. Research communities wishing to use these services currently have to seek out support and services from each individual provider. The integration of both governance and operations to research communities across all European e-Infrastructure within providers would provide researchers with a single front door for all their e-Infrastructure needs. This would greatly simplify access from a researcher's perspective, removing one of the barriers to entry and therefore providing more usage of the individual resource centres.

4.3 Operational Infrastructure

4.3.1 Strengths

The European Grid Infrastructure has built on the experience of the projects that preceded it to enable the operational infrastructure to operate over 350 resource centres across more than 50 countries by federating national resources that now supports over 1 million jobs a day. This includes the infrastructure-wide monitoring of services provided by individual resource centres; the collection of accounting records nationally and across the whole infrastructure; a support structure that allows issues to be raised, tracked and resolved across different organisational structures; the generation of resource centre or community specific statistics relating to the availability and reliability of services and a centralised registry of services and contact points to enable the operation of the infrastructure as a whole. There is no other grid infrastructure operating with such geographical distribution and performance anywhere in the world.

This federation of European wide resource centres provides international and national research communities with unprecedented economies of scale. Domain specific resource centre do not need to be built as existing local resource centres can be expanded to their maximum capacity and these resources integrated together and accessed through uniform services regardless of their geographical location and administrative domain. This has only been achieved by the knowledge and expertise built up and shared by individual resource centres within their NGIs and brought together within EGI.

For many years, this work was focused around a single set of domain specific services. During EGI-InSPIRE the operational infrastructure has been generalised to integrate services from ARC, UNICORE and Globus alongside the existing gLite integration. Full integration is expected to be completed during 2012.

4.3.2 Opportunities

This operational infrastructure and expertise is available to integrate services from other communities running in their own data centres. These services are currently being decoupled from each other to be offered as individual distinct products, which could then support some key scenarios including:

- **A VRC ensuring their distributed services are being reliably delivered by their community**: Once the deployed services have been integrated into the monitoring service by a simple script that probes their operation, statistics relating to the services availability and reliability can be collected and generated for the VRC management. NGIs can monitor the services and raise 'alarms' with the service provider if their services are not operating correctly for the service provider to resolve. **Example**: Providing consistent monitoring of the 600+ community data services across Europe within ELIXIR.
- **A VRC accounting for the usage of their distributed services by their community**: The usage records from each individual service (e.g. web logs, service invocations, or other metrics) are collected locally and injected as either individual or aggregated usage records into the accounting framework. The usage records are collected centrally and can be made available to the VRC management grouped by service provider, the group using the service or the action performed if this can be exported.

Alternatively, a fully managed service could be provided with EGI by to support scenarios such as:

- **A VRC specifying the services and storage that they need across Europe**: The European-wide network of NGIs allows the VRC to have the resources they have access to through their research community across Europe to be integrated for their community's exclusive use into EGI's monitoring infrastructure. EGI would work with the VRC to ensure that their services can be reliably deployed to meet their community's needs. **Example**: EGI validates, deploys and operates a community specific set of services across Europe on behalf of the WLCG collaboration.

4.3.3 Short-term Activities

As mentioned previously, full integration of UNICORE and Globus will be completed by the end of 2012 with the cooperation of the individual technology providers to deliver the necessary software components – primarily around accounting functionality – and for these to be integrated with the current operational infrastructure. A proof of concept integration with EBI and the ELIXIR project will be demonstrated by the end of 2012 by extending EGI's operational infrastructure to include the monitoring of some of their services..

4.3.4 Long-term Goals

The investment made during EGEE-III and in EGI-InSPIRE has established a decentralised scalable monitoring and operation of a nationally distributed production infrastructure. Future activity includes:

- **Innovation in the operational tools**: As the size of the infrastructure and the diversity of the services it monitors for different communities increases, there will be a continued need to innovate the operational tools to deal with these challenges. One foreseen need would be the increased volume of monitoring and accounting records being generated due to the greater scale of the production infrastructure.
- **Customisation of the community facing operational tools**: Many of the operational tools have web-based interfaces that are accessible by research communities in order to understand their use of EGI's resources. This may include accounting and support portals, or dashboards showing site availability. As more research communities use these tools, enabling community specific customisation and the user interfaces that work across different devices will be needed.
- **Sustained infrastructure operation**: Critical for many communities is the guarantee that the operational integration of the production infrastructure will be assured at a European level. Reliance on national funds to provide the local hardware and its operational staff has shown to be sustainable over the years for campus or national research computing centres. Ensuring the routine operation and coordination of these services at a European level (with the innovation and integration activity funded through specific projects) needs to rely on contributions coming from the countries and communities benefiting from the integration. With such transnational integration excess capacity in one resource centre can be made temporarily available to a wider research community.
- **Capacity building in small NGIs**: As the operational infrastructure continues to evolve some of the smaller and newer NGIs will need more support than the larger and more mature NGIs. Investing in these NGIs will ensure the increase in the skill base needed to sustain their inclusion in EGI during Horizon 2020.
- **Provision of computing and storage resources for European Research Communities:** Many international research communities new to EGI will not have established dedicated national or domain specific resource allocation agreements. Resources provisioned centrally by the EC could be provided to allow these newer communities to fully explore how EGI could benefit their research.

Existing research communities could temporarily 'burst out' to these resources when their existing resources become fully utilised.

4.4 Federated Cloud Infrastructure

4.4.1 Strengths

Currently, at the heart of the EGI is the federation of individual compute-clusters into a single infrastructure to provide high throughput data analysis, data storage and simulation. The technology that has developed within the EGI community provides secure data movement and access to these compute resources and supports over 1M jobs a day over 300,000 cores spread across over 350 resource centres.

As technologies evolve, so do the expectations of those using and providing the infrastructure. Virtualisation technologies are already being adopted by many resource centres through private clouds (virtualised resources accessed only to internal users) that improve their internal management of their resources and to give greater flexibility to their local users. However, no clear model has yet emerged as to how such private cloud infrastructures can be federated for consistent external use through uniform interfaces to access compute and storage capability provided by different resource providers. Cloud standards and open source reference implementations are emerging within the European DCI community and elsewhere that need to be validated for the unique federated environment within which Europe operates, allowing researchers to flexibly deploy the services they need when they need to do so.

EGI provides a perfect environment for the deployment of technologies built around open standards. Interoperability through agreed standards between different resource centres is vital as in an open ecosystem where each resource centre needs to be able to deploy the private cloud management system that suits their particular needs. The transition to a European wide federated cloud infrastructure can be increment. Resource centres are free to host the software environments required by their researchers either within their local private cloud by allowing federated access, or by deployment directly onto the local hardware resources.

4.4.2 Opportunities

A federated cloud that encompasses national resources across Europe would provide a significant new capability for communities to integrate and run the services that they wanted to run wherever they could gain access to their resources. This would enable:

- **Individual researchers to access and analyse remote data**: Researchers frequently wish to analyse large data sets stored remotely with their own algorithms. A federated cloud model would enable researchers to gain consistent access to computing capacity near the data sets that they need to access, onto which they can launch their own virtual machine environment. Exploiting the integrated networking between the data store and the virtual machine would allow rapid analysis of the data. Re-analysis with a new algorithm is simple as the researcher just needs to provide new virtual machine. **Example**: Social scientist starting a locally prepared virtual machine to deploy into a remote cloud analyse a remotely stored data set.
- **Research collaborations can deploy an environment to meet their needs**: Large multi-national research collaborations will be able to access resources across many resource centres, but they need these resources to have the latest software environments that they require for their collaborative research. Given the critical mass of a larger collaboration they will probably have the expertise to

integrate their own platforms for deployment and management on the e-Infrastructure that they have access to. **Example**: The WeNMR project works with research centres across the world to integrate their resources into a web portal that provides easy access to resources that support their research activities.

4.4.3 Short-term Activities

Many of the technologies developed to federate compute clusters can be reused to federate private clouds. The EGI Federated Cloud Task Force is exploring how technologies coming from within our community and from other communities can be deployed and integrated together. First results from this initiative will emerge in 2012 and initial production deployments could be achieved by the end of the EGI-InSPIRE project in 2014. Deployment and adoption can be incremental within EGI as resource centres can choose to continue deploying software directly onto their local hardware instead of allowing external operators to deploy software appliances directly onto their cloud infrastructures. Such a deployment approach has no impact on the services currently used by the research community.

This activity will allow an initial production deployment across early adopting sites by the end of the EGI-InSPIRE project. The experience gained in this initiative over the next two years with the early adopting resource centres and research communities will enable technical issues relating to functional capability, scalability and reliability to be identified and resolved through innovation projects that follow.

4.4.4 Long-term Goals

Although the effort from EGI-InSPIRE, and the proven technical and operational expertise within the EGI community, will enable a basic federated cloud infrastructure prototype to be put in place further investment will be needed to develop the ecosystem around it. These will need to include:

- **Investment in open-source infrastructure technology providers**: As the production infrastructure continues to grow in scale and evolve its service offering additional innovations will be needed in the federating infrastructure – the Infrastructure Platform. For instance, higher-level services that enable the optimal use of a federated cloud infrastructure by providing to the platform operator provisioning and autonomic management tools will become critical in exploiting the resources within EGI. The services used to federate the private clouds in EGI will help other public service sector groups protect themselves against cloud vendor lock in.
- **Investment in domain specific open-source technology providers**: Fragmented development efforts between different research communities can lead to duplicate competing efforts. This may be desirable in order to introduce new technical innovations but sources of reliable, documented and proven software within which new technological innovations can be delivered would reduce fragmentation. Many of these technologies will be incorporated into the domain-specific Community Platforms made available to research communities, or incorporated into a general Collaboration Platform (e.g. services such as Federated Identity, Persistent Data Identifiers) that are used across different research communities to promote collaboration and data sharing[9].
- **Supporting the platform integrators**: Platform integrators emerge in the EGI ecosystem as the key enabler for researchers to access the production infrastructure. Supporting these teams in the integration of bespoke software environments needed by the individual researchers would be

[9] MS510: EGI Platform Roadmap https://documents.egi.eu/document/970

accelerated by software catalogues containing high-quality community-rated software components that can be assembled into a virtual machine image, and a platform store (equivalent to the Apple AppStore) where the virtual machine images can be published as Community Platforms for direct use by any research community.

4.5 Virtual Research Environments

4.5.1 Strengths

Virtual Research Environments (VRE) are a combination of environments that provide the researcher with easy access to the services deployed in EGI to enable their data analysis activities. Initially, the VRE consisted of a command line interface for the researcher to access the services deployed across the infrastructure. Over the years, higher-level generic tools and domain specific environments have been developed by many research communities to simplify the data analysis process.

Science gateways (or portals) have provided an approach to reducing the technical barriers to accessing remote computing resources for many years. EGI supports this mode of access through a dedicated portal policy that relates the levels of authentication to the ability of the individual user to customise the portal activity. The growing support in Certificate Authorities for robot certificates enables a user's portal authentication to be decoupled from the identity they use to access EGI's resources.

This is however a rich area of activity in terms of EC and national investment alongside an active open-source community and standards activity around generic portal frameworks. Although, the level of integrated activity provided by EGI is below that coming from TeraGrid (and now XSEDE[10]) through their Science gateway programme and the Globus Online[11] model to support data transfer, there is a rich activity of community specific portals in Europe that demonstrate an easier alternative route to accessing e-Infrastructures than directly through the middleware.

EGI.eu is helping to coordinate activity and requirements within the EGI ecosystem and its related projects at domain specific workshops (e.g. 4th International Workshop on Science Gateways for Life Sciences, WeNMR), sessions at the Technical and Community Forums, and with leaders in the community.

4.5.2 Opportunities

The use of the web as a route to accessing e-Infrastructure (and here the web ranges from desktop browsers to tablet applications to mobile browsers) provides an amazing opportunity to meet the demand coming from young researchers to access e-Infrastructures with the tools available to them during their daily activities. Specifically,

- **Providing a source of maintained customisable portlets to access EGI services**: With no clearly supported portal framework or repository of portlets that have been designed and maintained to access EGI services, it is difficult for platform integrators to reuse and contribute portlets to the community. Having such a resource of documented and proven software will make it easier to access

[10] https://www.xsede.org/
[11] https://www.globusonline.org/

capabilities within EGI such as file management, job management, application execution, workflows, etc. **Example**: The long-running Open Grid Computing Environment (OGCE) activity in the USA[12].

- **Using mobile devices to access EGI servi*ces:*** Mobile devices (e.g. smart phones, tablets) are becoming commonplace in the research community for accessing web services. Providing customisable access to EGI services by using industry standard environments (e.g. HTML5) will promote the uptake of EGI services through usable web interfaces. **Example**: Many of EGI's current portals are accessible through mobile devices but are not currently optimised for these smaller screen sizes and restricted input options.
- **Provision of centrally managed EGI services**: The Software as a Service (SaaS) model has been shown in industry to lower the barriers to adoption, broaden the uptake of services and shown to be a cost effective service delivery mechanism. Such an approach presents many attractive possibilities for EGI (i.e. by eliminating effort relating to distributing the software - deployment testing, installation support, porting to other environments) to deliver scalable and reliable centrally managed collaboration services to its supported research community. **Example**: The data movement service developed by Globus Online.
- **Training the software developers**: By adopting usability best practices during the development process the need to train the end-user of a mobile application can be considerably reduced or even eliminated. Supporting the uptake of toolkit designed around accessing EGI's collaborative services from the mobile web through the training of software developers will ensure that the next generation web based applications will be easier to use. **Example**: Training courses provided by EMI and IGE focus more on API level integration rather than usability training. Nor do any of the current software developers undertake usability studies of their graphical interfaces

4.5.3 Short-term Activities

EGI-InSPIRE does not have any technical effort dedicated to portal frameworks. Therefore its focus will be on developing synergies and cooperation within the EGI community. This has already started with workshops at the EGI Technical Forum in Lyon and separate workshops held in conjunction with other community activities. These activities will help foster cooperation between current portal activities such as SCI-BUS and with the EGI community at national and European levels.

This period will allow a synthesis of requirements and the assessment for activities that could be supported in 2014 and beyond through renewed funding from the EC. The balance that needs to be struck between tightly customised solutions that work for a single research community or researcher versus a generic solution that allows investment to bring a return across multiple communities.

4.5.4 Long-term Goals

Much of the investment in software innovation effort over the last decade has been invested in low-level middleware activities with minimal investment being made into considering the best and easiest way for the researcher to access the underlying e-Infrastructure. Basic portal environments have reduced the barrier to entry and domain specific solutions built around particular workloads (e.g. GANGA and DIANE) have been generalised for use by other research communities.

[12] http://www.collab-ogce.org/ogce/index.php/Main_Page

- **Web based frameworks to access EGI services**: Use the exponential growth and adoption of connected mobile and desktop devices to access EGI services by providing the EGI research communities with documented, tested and customisable frameworks ready for platform integrators to assemble with the applications and workflows available and ready to use that researchers' need.
- **Engaging the long-tail of researchers in the ERA**: For EGI to have impact across the whole ERA the barriers to accessing resources need to be reduced. Web-based access to consumer services is becoming ubiquitous and such a paradigm needs to be extended to EGI services. By using the previously developed frameworks, platform integrators working on behalf of researchers should be able to rapidly integrate and customise an environment and populate it with the applications and workflows needed for them to undertake their data analysis tasks.

5 Funding EGI's New Horizons

5.1 Funding Sources

Funding, regardless of the source, provides a means of driving change towards the priorities of the funding agency. This holds irrespective as to if the funding source is from within the community (such as EGI.eu participation fees), from the European Commission (as part of Horizon 2020), or national research or infrastructure funds. EGI's future strategy is to provide a clearer alignment of the funding it receives to the activities it supports. For instance the distinction between the routine operation of the ecosystem (which should become supported solely by national infrastructure or community funds), the innovation needed to change the operation or technology in the ecosystem (funded from national research councils or the European Commission), and the structuring and support needed to build a borderless transnational e-Infrastructure (funded by the European Commission).

The future activities identified previously are grouped under the most appropriate funding sources. An analysis of how these funding sources could be exploited by the different components of the EGI Ecosystem is available elsewhere[13].

5.1.1 Community Funding

Community based funding models, either through membership fees or direct service charges, provides a mechanism to drive consolidation and efficiency in the routine baseline operation of the infrastructure. Community funding can assure the continued operation of the core infrastructure to meet the needs of the research communities that directly benefiting from it. The majority of the infrastructure's operating costs (i.e. hardware, staff, support, buildings, electricity, etc.) are already funded directly from national funding sources. Moving the long-term steady state support of the infrastructure completely to national funding schemes would send a strong message as to its sustainability to the research communities that they are depending upon it.

A community based funding model is currently used to support EGI.eu (a Dutch foundation) through participation fees set by the EGI Council and levied annually. A new legal instrument has recently been adopted by the EC to facilitate the establishment of pan-European organisations dedicated to supporting research communities – the European Research Infrastructure Consortium (ERIC). Two models have been explored by the EGI community around the adoption of the ERIC model:

- **A 'lightweight' EGI ERIC** would absorb the work currently undertaken by EGI.eu but through the ratification process required of an ERIC could bring additional national recognition and funding commitment from the member states.
- **A 'heavyweight' EGI ERIC** would absorb the work currently undertaken by EGI.eu and provide an additional service to the EGI community through the purchasing and management of ICT facilities (potentially funded by the EC and by member states taking advantage of preferential taxation option) to meet the needs of the research communities it supports.

Both of these models remain under discussion at this time.

[13] D2.18 Evolving the EGI Business Model. Add in reference when available.

However, community funding models relating to the direct use of the infrastructure must be considered in the short-term to be a finite source of funds from individual researchers and those in small research collaborations. Researchers are not accustomed to a 'pay for use' model for the research services they consume, nor are they set up to do so from an institutional or funding grant perspective. Virtual research communities may enter into a partnership with EGI which does not directly bring cash into the ecosystem, but the partnership at a European level can translate into partnerships at a national level through the joint development and integration of software to build community platforms, operation of community specific platforms on national resources, which could allow local funds to be leveraged to support national deployment costs.

The introduction of technological innovation into the infrastructure with its associated technology stabilisation and operational integration costs would need to be supported exclusively by short-term focused projects funded from other sources.

5.1.2 Horizon 2020 or National Innovation Funding

Further technological innovation is needed within the operational infrastructure and crucially in the mechanisms that are available for individual researchers to easily access the infrastructure. These areas include:

- Improved web based access for individual researchers and research collaborations to easily access services tuned to their needs through virtual research environments.
- The supported development of centrally managed services using the Software as a Service model to eliminate the need for individuals to distribute, install, operate and maintain software services.
- Development of easily reusable open source technologies that provide common functionality that can be reused by platform integrators to serve the needs of multiple communities.
- Consolidating the roles of the platform integrator (who works closely with a research community to assemble the software environment needed for their research activities) and the platform operator (who operates the established software environment across distributed resources for the research community) to provide technical outreach to researchers new to EGI by facilitating interactions between these groups.
- Enhancement of the operational tools to meet the technical challenges of operating a larger and more diverse infrastructure for a greater number of research communities.

5.1.3 Horizon 2020 Support Funding

While the EGI community has been in existence in various forms for over a decade it continues to evolve. The pioneering experimentation phase undertaken between resource providers and a few early adopting research communities is over. The challenge for the next decade is to scale out the delivery of services to meet the needs of the whole ERA. This will require further investment in areas such as:

- Development of policies and processes to scale an ecosystem designed for a few large research communities to one that can manage many large communities, small communities and even support countless individual researchers.
- Continue to build the international ecosystem and the community within it through regular Forums that promote collaboration and awareness raising of the different activities, and topical workshops to help develop a community wide approach in specific areas.

- Communications within EGI's own technical community (resource providers technology providers, platform integrators and platform operators), and communications within the research community that uses e-Infrastructures (such as publications like iSGTW) and marketing to research communities not already using e-Infrastructures such as EGI.
- Training and education of the technology providers and operations staff to promote community wide best practice to ensure that the smaller NGIs and communities new to EGI have the required skills for the effective and secure use and operation of EGI's services.

5.2 Planning Scenarios

The strategic investment areas both directly in the EGI community and in the research communities that surround it are dependent on funding. The following sections explore the impact of different funding levels on the plans presented here and their consequence for the EGI community and the research communities using its federated resources.

5.2.1 Scenario 1: Zero EC Funding

Scenario

No further EC funding for the coordination, maintenance and operation of the infrastructure is forth coming and there is continued national funding to support the integration of national resources and support of national research communities.

Short-term Impact

With European wide coordination being supported solely by community funding (the participants of EGI.eu) the focus would be on the steady state operation of the infrastructure. There would be no funded activities such as community building, marketing, communication, policy development, organised events, workshops, etc. All of this activity would be delegated to best effort activities coming from the EGI.eu participants (the NGIs) with any spare national resource they may have. With the EC demonstrating its lack of commitment so clearly it is unlikely that national governments would see this as a priority area for their funding.

Long-term Consequence

Within no funding to develop the operational infrastructure, or the community that surrounds it, then the operational infrastructure would begin to decay. Research communities would become frustrated at the responsiveness of what they would consider to be a centrally funded resource leading to dissatisfaction. As EGI no longer remained a dependable resource then research communities would return to their own bespoke solutions resulting in fragmentation and eventual collapse of a European wide e-Infrastructure for computing and storage.

5.2.2 Scenario 2: Minimal EC Funding

Scenario

EC funding continues at levels comparable to EGI-InSPIRE and this is complemented by continued national funding to the NGIs.

Short-term Impact

Funding would be diverted into priority actions with Community & Coordination and Operational Infrastructure and to establish a federated European cloud built from nationally supported private clouds. This would assure the continued development of the community and the core aspects of the operational infrastructure enabling innovations to be introduced in response to new requirements emerging from increased scale and new cloud functionality that would allow researcher's to personalise the software environments available to them.

Long-term Consequence

The result would be for EGI's future to be secured but with no concentrated effort to increase its uptake across the whole ERA through the collaborative development of virtual research environments targeted around their needs. Resources would only permit a general purpose framework to be provided for European wide deployment of virtualised appliances. As a result, EGI would remain a resource for those research communities that had sufficient organisational and technological critical mass to provide their own integrated platforms and to operate them (approx. 50 in Europe). EGI would have sufficient resources to work with these researchers' and to support their integration – but little beyond that.

5.2.3 Scenario 3: Expanded EC Funding

Scenario

EC funding levels increase significantly above the current EGI-InSPIRE level but are invested VRE related projects engaging directly within the EGI ecosystem. This increase is matched by national funding within the NGIs to support the increased usage coming from the broader adoption across different research communities.

Short-term Impact

Additional funding beyond Scenario 2 must be invested in a projects that help build out the virtual research environments needed by particular research communities. This may involve organisations active within EGI partnering with research led consortiums to act as technology providers, platform integrators or platform operators and even to help build the web-based environments used directly by the researchers. These projects must be critically reviewed during the proposal and execution phase to ensure that they re-use existing solutions where available and contribute fully developed software components back into the community for others to build upon.

Long-term Consequence

Investing in research led collaborations across strategic societal areas will help build engagement within the research community and help establish the online ERA built upon the established e-Infrastructures. To avoid excessive duplicated development activity and projects with generic software outputs unable to be reused by others, this investment needs to be proactively coordinated either by the European Commission or their delegate. A curated legacy of high-quality software components within the ecosystem would lower the barriers to undertaking data analysis and help other research communities engage in becoming users of the online ERA. This scenario would broaden access to e-Infrastructures and start bringing some of the potential 2.8M researchers in Europe online.

6 Conclusions

6.1 EGI in 2012

The establishment of NGIs across Europe has been achieved. Over 35 NGIs are participating in the EGI.eu organisation and its governing body the EGI Council. However, behind this political integration lies a broader spectrum of technical and operational activities. Some NGIs are able to deliver in one national structure a number of tightly coupled national roles, including resource centres, resource infrastructure providers, platform integration, platform operation and even technology provision. Other NGIs have chosen to focus on being a resource infrastructure provider collaborating with independent resource centres to ensure their services are accessible and available to their own national research community and their international collaborators if capacity permits. The diversity of NGI activity and scope strongly indicates that further immediate investment is needed to reduce the variance between NGIs by improving their capability where needed. NGIs will never all be the same but it is essential that they have all achieved a critical sustainable mass for their local environment.

The technology within the production infrastructure still provides the greatest barrier to further adoption by other research communities. The integration of the different technology solutions coming from two distinct platform integrators has been achieved, but this has not sufficiently changed the functional capabilities offered by EGI or clearly opened up the infrastructure as a platform for other new research communities. Federating the private clouds already beginning to be deployed by many of the resource centres within EGI would lower the barriers to entry for new research communities. A federated European cloud would offer research communities the use of virtual machine images prepared and operated to meet their specific requirements, by allowing distinct workloads to be operated within a single European wide resource. This would enable many potential new research communities who already have their own software frameworks to be approached and supported by the EGI community, who could provide the skills and supporting e-Infrastructure needed to enable them to scale out across Europe.

Some of these issues are being addressed with EGI-InSPIRE by:

- Travel support for smaller NGIs that need to improve their technical skills and contacts.
- Prototyping a federated private cloud infrastructure through a dedicated task force.
- Generalising the operational infrastructure to support different domain specific services.

6.2 Vision, Mission and Core Values

In developing this strategic plan for EGI it has been recognised that EGI has 'come of age'. The *ad hoc* infrastructure operations that started in EDG is over. EGEE developed the operational and technical structures for professional service delivery on a European wide scale. EGI-InSPIRE has consolidated these operational and technical structures and is now undertaking the organisational changes, starting with establishment of EGI.eu, that will help build an open ecosystem that will sustain the development of EGI in the decades to come. These principles are encapsulated in the following Vision, Mission and Core Values:

Vision

To support the online European Research Area through a pan-European research infrastructure based on an open federation of reliable services that provide uniform access to national computing, storage and data resources.

Mission

To connect researchers from all disciplines with the reliable and innovative ICT services they need to undertake their collaborative world-class research.

Core Values

- **Leadership**: EGI is a leading pan-European infrastructure, integrating worldwide computing, storage and data resources to support an economy built on innovation and knowledge transfer.
- **Openness**: EGI operates with a transparent governance structure that integrates the views and the requirements of all stakeholders, from research communities to resource providers.
- **Reliability**: EGI provides a reliable infrastructure that research communities can depend on to collaborate with their peers and deliver innovation.
- **Innovation**: EGI will continue to meet the needs of research communities operating at unparalleled geographic and technical scale by partnering to bringing new technologies into production.

6.3 EGI during Horizon 2020

For the online European Research Area to come to fruition by the end of Horizon 2020 an investment programme needs to be established immediately that targets innovation and service delivery activity in EGI and in the Virtual Research Environments needed across the different research communities. Such a programme needs to be pro-actively orchestrated to invest in areas which will provide the greatest impact, but must also build on the existing foundation that has come from previous investments. The ability for a project to provide a reusable legacy needs to be a key enforced assessment criteria alongside the support of organisations to maintain the output from these projects which are seen to have value and use within the community. The continued maintenance and operation of the online ERA needs to be assured through the member states for all researchers ranging from the individual to the global virtual research community to be confident in its adoption.

7 Annex 1: Individual Researchers in the European Research Area

There were 1.75 million researchers employed in the European public sector in 2008 which marked an increase of almost 40% when compared with 2000.

Sector	**Head Count (2008)**
All sectors (higher education, government, business enterprise and private non-profit sector)[14]	2.837.347
EU Government	230.557
EU Higher Education	1.185.247
EU Public Sector (Higher Education and Government)	**1.416.000**
EU Public Sector plus Russia, Norway, Switzerland, Turkey and Croatia	**1.715.000**

Given the average growth rate of 1.5% the number or researchers active in the public sector in 2011 is estimated to be around 1.83M.

Public Sector Field	**Head Count (2011)**
Social Sciences And Humanities	525,000
Natural Sciences	420,000
Engineering and Technology	370,000
Medical and Health Sciences	355,000
Agricultural Sciences	85,000
TOTAL	**1,755,000**

[14] According to Eurostat, researchers are divided into four R&D-performing sectors (OECD classification): 1) Business enterprise sector 2) Government sector 3) Higher education sector 4) Private non-profit sector. **The government sector is combined with higher education sector in order to create one integral category relevant for EGI- public sector category. The government sector** is composed of:

- All departments, offices and other bodies which furnish, but normally do not sell to the community, those common services, other than higher education, which cannot otherwise be conveniently and economically provided, as well as those that administer the state and the economic and social policy of the community. (Public enterprises are included in the business enterprise sector.)

- Non-profit institutions (NPIs) controlled and mainly financed by government, but not administered by the higher education sector.

The higher education sector is composed of:

- All universities, colleges of technology and other institutions of postsecondary education, whatever their source of finance or legal status.

- It also includes all research institutes, experimental stations and clinics operating under the direct control of or administered by or associated with higher education institutions.

8 Annex 2: EGI Ecosystem SWOT

The following section provides a SWOT (strengths, weaknesses, opportunities & threats) analysis of the components within the EGI ecosystem and should be read in conjunction with the overview of the EGI ecosystem given in Section 2.2.

8.1 Researchers

	Strengths	Weaknesses	Opportunities	Threats
Individual Researcher	- Source of innovative research ideas. - Generator of scientific data, information, knowledge and expertise.	- Lack of easy access to large-scale data management and processing facilities. - Unlikely to have internal resources for significant software development so dependent on available external solutions. - Lack of influence in the specification of external e-Infrastructures services.	- Greater benefit and use of e-Infrastructures through lowered technical barriers & increased flexibility for faster & better results. - Focusing on research rather than managing their local e-Infrastructures needs.	- Unable to analyse data to extract knowledge and produce innovations. - International & local publicly funded e-Infrastructures do not meet their needs. - Commercial offerings drive up the cost of their research and may not meet technical requirements.
Research Collaboration	- Collaborative generator of ideas, data and innovation. - Able to access ICT expertise to breach technical barriers within their local organisation or through the collaboration.	-Distributed community makes it harder to achieve critical mass. - Organisational borders may provide barriers to data access, analysis and resulting innovation. - Ad-hoc solutions can lead to inefficiencies or inability to conduct research and a longer time to achieve results.	- Easier collaboration across organisational boundaries through e-infrastructures. - Focusing on research rather than managing their local e-Infrastructures needs.	
Virtual Research Community	- Collaborative generator of ideas, data and innovation. - Able to access ICT expertise to breach technical barriers local within their local organisation or through the collaboration. - International critical mass as recognised science mission. - Structured governance (formal or informal).	- Distributed community makes it harder to achieve critical mass. - Organisational borders may provide barriers to data access, analysis and resulting innovation. - Ad-hoc solutions can lead to inefficiencies or inability to conduct research and longer time to achieve results. - The need to share and manage increasing amount of digital data is mission critical.	- Easier collaboration across organisational boundaries through e-Infrastructures. - Focusing on research rather than managing their local e-Infrastructures needs.	- International & local publicly funded e-Infrastructures do not appear to meet their needs. - With no suitable public e-Infrastructure, they will need to develop their own solution as commercial solutions unlikely to support extreme requirements.

8.2 Technology

	Strengths	Weaknesses	Opportunities	Threats
Community Technology Providers	- Development of community-specific open source software components. - SLA for software and support services supplements close relationship with consumers. - Technical expertise from within the community.	- Lack of critical developer mass for true open-source model. - Relatively small community bears maintenance costs. - Immature support around software components.	- Able to respond to and engage with specific community needs. - Expand market shares and improve reuse by improving quality & functionality.	- Specific community needs may evolve faster than available development effort. - Technical failures can endanger adoption or retention. - Commoditisation of other software components leading to competition
Generic Technology Providers	- Offer of free/inexpensive generic software components. - SLA for software and support services on a professional basis with mature community or paid support. - Strong and trusted brand name across multiple user groups.	- Inability or unwillingness to react to the needs of small user groups.	- Provides a technology source with maintenance and development shared across many communities.	- Technical failures can endanger adoption or retention. - Pressure to maintain or expand features may lead to reduced quality.
Platform Integrators	- Integration of independent software components into a coherent software stack to enable fully-functional services based on user needs. - Offer an SLA for integrated software and support services. - Technical expertise across an integrated solution.	- Inability to provide consistent and up-to-date documentation due to external software sources. - Dependency on external sources for the quality of software components.	- Able to respond to specific needs by adapting or sourcing required software components.	- Inability to find suitable software components for integration. - Incompatibility of independent software components.

8.3 European Coordination

	Strengths	Weaknesses	Opportunities	Threats
EGI.eu	- International coordination, visibility, branding and management of services. - Strategy, policies and planning for European resource federation. - Negotiation on behalf of research and resource provider stakeholders. - Interaction with international research communities at European level.	- Dependent on external partners for providing technical services. - Demonstrating added value of coordination activity.	- Coordinate service delivery for individual researchers, research collaborations and VRCs. - Coordinating EGI's role in DAE and ERA. - Drive the evolution of EGI to attract new research communities. - Explore collaborations with commercial sector.	- Fragmentation between European e-Infrastructure organisations. - Failure of national grid infrastructures. - Unable to open the ecosystem and infrastructure to enable wider adoption. - Inability to prove European value-add. - Strategy and business models of the NGIs become unsustainable.

8.4 National Structures

	Strengths	Weaknesses	Opportunities	Threats
National Coordination (the NGI)	- National single point of contact for government, research communities and resource centres as regards ICT services for e-science.	- Short-term funding. - Lack of effective strategies to outreach new research communities at the national level. - No on-demand model for resource allocation. - Low visibility within research communities.	- Become an authoritative voice for influencing scientific computing activities, DAE & ERA at the national level. - Increase impact of scientific computing by broadening uptake in research communities.	- Limited relevance and role may endanger sustainability. - Weak engagement with government and resource centres.
Resource Infrastructure Provider (the National Operations Centre)	- Coordination of operations across their federated resource centres. - Provision of general ICT services for their federated resource centres.	- Reluctance in changing the current operation model and opening up to new technologies from different communities. - Inability to move quickly due to the complexity of the infrastructure.	- Consolidate activities to achieve economy of scale. - Source of expertise for consultancy to other communities.	- Inability to prove added value to resource centres. - Resources centres not delivering services to agreed quality.
Platform Operators	- Technical expertise to operate domain-specific platforms on behalf of research communities. - Source of platform-related support and consultancy.	- Dependent on resource centres and resource infrastructure provider delivering the required platform infrastructure.	- Able to deploy and operate products and services to meet the needs of particular researchers.	- Inability to maintain reliable services due to lack of documentation or reliable software components. - The reliability of the underlying infrastructure that is out of their direct control.
Resource Centres	- Operate services to access local physical resources. - Source of local technical expertise and consultancy.	- Unreliable service offering due to unreliable software or hardware. - Limited available technical effort for software deployment and operation.	- Expand and optimise usage of resources across different research groups. - Ability to provide potential high-value customisation for specific research groups.	- Increase of low cost commercial providers offering more flexible cost-effective resources.

1 # 8.5 Funding

	Strengths	Weaknesses	Opportunities	Threats
National Funding	- Funding for national interests and effort.	- Short-term funding. - Difficulty to invest in non-national activity.	- Facilitate research and development aligned with national strategies.	- Funding results do not meet local success criteria.
European Funders	- Funding for innovation done within international teams. - Long-term vision for e-Infrastructures in society.	- EC project model imposes administrative constraints.	- Provide e-Infrastructure to support ERA, DAE and other EC initiatives.	- Disconnect between vision and needs. - Unable to deliver within constraints of EC model.
Commercial Funds	- Provides direct links to market. - Business experiences and efficient processes.	- Potential lack of transparency and openness.	- Enable development and operation costs to be shared. - Enable access to new markets and customers. - Diversify income sources.	- Low return on Investment would jeopardise funding.
Community Funds	- Can directly support community needs.	- Subject to the available of funds.	- Direct link of value received to value delivered.	- Limited source of funds so must target key functions.

2