



EGI-InSPIRE

SEEKING NEW HORIZONS: EGI'S ROLE IN 2020

EU DELIVERABLE: D2.30

Document identifier:	EGI-1098-D230-V2
Date:	27/04/2012
Activity:	NA1
Lead Partner:	EGI.eu
Document Status:	FINAL
Dissemination Level:	PUBLIC
Document Link:	https://documents.egi.eu/document/1080

Abstract

The EGI Strategic Plan – Seeing New Horizons: EGI's Role in 2020 – describes how EGI will evolve into a universal federated platform for supporting compute and data intensive Research and Education communities. EGI will evolve to provide a framework that will host a range of high-throughput solutions, including both grid and cloud approaches. The intention is to separate the higher level software solutions (the virtual research environments) from the operational infrastructure so that a complete range of approaches can be hosted depending on the needs of each Research and Education community. The infrastructure platform will provide a managed environment capable of providing appropriate services to a wide variety of software stacks. These services will be defined and progressively implemented in the coming 24 months so that by the end of EGI-InSPIRE, EGI will be a universal e-infrastructure capable of hosting the complete range of distributed computing approaches required by European Research and Education.

The plan is targeted at European and national policy makers, senior managers in resource providers, virtual research communities and other stakeholders within the EGI Ecosystem. Additional information can be found in: the EGI Technical Roadmap¹ which details work taking place within the EGI-InSPIRE project, and the EGI Business Model² which describes the value generation activities of different components in the EGI Ecosystem.

¹ D2.31 – <https://documents.egi.eu/document/1094>

² D2.18 – <https://documents.egi.eu/document/1040>



Copyright notice

Copyright © Members of the EGI-InSPIRE Collaboration, 2010-2014. See www.egi.eu for details of the EGI-InSPIRE project and the collaboration. EGI-InSPIRE (“European Grid Initiative: Integrated Sustainable Pan-European Infrastructure for Researchers in Europe”) is a project co-funded by the European Commission as an Integrated Infrastructure Initiative within the 7th Framework Programme. EGI-InSPIRE began in May 2010 and will run for 4 years. This work is licensed under the Creative Commons Attribution-Noncommercial 3.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/3.0/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, and USA. The work must be attributed by attaching the following reference to the copied elements: “Copyright © Members of the EGI-InSPIRE Collaboration, 2010-2014. See www.egi.eu for details of the EGI-InSPIRE project and the collaboration”. Using this document in a way and/or for purposes not foreseen in the license, requires the prior written permission of the copyright holders. The information contained in this document represents the views of the copyright holders as of the date such views are published.

I. DELIVERY SLIP

	Name	Partner/Activity	Date
From	Steven Newhouse	EGI.eu	10/2/12
Reviewed by	Moderator: Reviewers:	PMB	27/4/12
Approved by	AMB & PMB		27/4/12

II. DOCUMENT LOG

Issue	Date	Comment	Author/Partner
1	10/2/12	Initial draft for internal EGI.eu discussion.	Steven Newhouse/EGI.eu
2	16/2/12	Revised following internal feedback for wider review.	Steven Newhouse/ EGI.eu
3	17/2/12	Revised following internal review for EGI-InSPIRE PMB discussion.	Steven Newhouse/ EGI.eu
	7/3/12	Revised following comment from EGI-InSPIRE PMB for public comment.	Steven Newhouse/ EGI.eu
	3/4/12	Revised following comment from the EGI Council, EGI Council Workshop, EGI-InSPIRE PMB and individual community contributions.	Steven Newhouse/ EGI.eu
	19/4/12	Revised following comment from the UK, PL, NL & CERN	Steven Newhouse/ EGI.eu

III. APPLICATION AREA

This document is a formal deliverable for the European Commission, applicable to all members of the EGI-InSPIRE project, beneficiaries and Joint Research Unit members, as well as its collaborating projects.

IV. DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the authors. The procedures documented in the EGI-InSPIRE “Document Management Procedure” will be followed:

<https://wiki.egi.eu/wiki/Procedures>



V. TERMINOLOGY

A complete project glossary is provided at the following page: <http://www.egi.eu/about/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.



VI. PROJECT SUMMARY

To support science and innovation, a lasting operational model for e-Science is needed – both for coordinating the infrastructure and for delivering integrated services that cross national borders.

The EGI-InSPIRE project will support the transition from a project-based system to a sustainable pan-European e-Infrastructure, by supporting ‘grids’ of high-performance computing (HPC) and high-throughput computing (HTC) resources. EGI-InSPIRE will also be ideally placed to integrate new Distributed Computing Infrastructures (DCIs) such as clouds, supercomputing networks and desktop grids, to benefit user communities within the European Research Area.

EGI-InSPIRE will collect user requirements and provide support for the current and potential new user communities, for example within the ESFRI projects. Additional support will also be given to the current heavy users of the infrastructure, such as high energy physics, computational chemistry and life sciences, as they move their critical services and tools from a centralised support model to one driven by their own individual communities.

The objectives of the project are:

1. The continued operation and expansion of today’s production infrastructure by transitioning to a governance model and operational infrastructure that can be increasingly sustained outside of specific project funding.
2. The continued support of researchers within Europe and their international collaborators that are using the current production infrastructure.
3. The support for current heavy users of the infrastructure in earth science, astronomy and astrophysics, fusion, computational chemistry and materials science technology, life sciences and high energy physics as they move to sustainable support models for their own communities.
4. Interfaces that expand access to new user communities including new potential heavy users of the infrastructure from the ESFRI projects.
5. Mechanisms to integrate existing infrastructure providers in Europe and around the world into the production infrastructure, so as to provide transparent access to all authorised users.
6. Establish processes and procedures to allow the integration of new DCI technologies (e.g. clouds, volunteer desktop grids) and heterogeneous resources (e.g. HTC and HPC) into a seamless production infrastructure as they mature and demonstrate value to the EGI community.

The EGI community is a federation of independent national and community resource providers, whose resources support specific research communities and international collaborators both within Europe and worldwide. EGI.eu, coordinator of EGI-InSPIRE, brings together partner institutions established within the community to provide a set of essential human and technical services that enable secure integrated access to distributed resources on behalf of the community.

The production infrastructure supports Virtual Research Communities (VRCs) – structured international user communities – that are grouped into specific research domains. VRCs are formally represented within EGI at both a technical and strategic level.

VII. 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, these capabilities need to be deployable on demand in order to provide a foundation for the digital 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) was established in 2010 as a result of pioneering work that has, over the preceding decade, built a collaborative production infrastructure of uniform services through the federation of national computing and storage resource providers that now supports multi-disciplinary science across Europe and around the world. An ecosystem of national and European funding agencies, research communities, technology providers, technology integrators, resource providers, operations centres, over 350 resource centres, coordinating bodies and other functions has now emerged to serve over 21,000 researchers in their intensive data analysis research by running over 1.2 million jobs a day. A selection of case studies relating to the work done by some of these researchers is available³. 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 digital ERA can be brought online by building upon the strengths that have been developed over the last decade of deploying, operating, monitoring, accounting and supporting virtual research environments through a federation of national resource providers for multiple research communities at a European scale by:

- **By building an ecosystem that promotes competitive cooperation, collaboration and interaction at local, national and European level.** 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.
- **Extending a European scale operational infrastructure that ensures open uniform access to locally deployed, domain specific services.** While continuing to support the current services for the currently supported research communities, the operational infrastructure needs to be developed to allow other European scale research communities to monitor and manage their own services operating at their distributed facilities. In addition, the operational infrastructure needs to support the federation of resources (such as institutional private clouds) in the public sector and public clouds from the commercial sector to support uniform standards-based transnational cloud access as a new capability to attract new research communities.
- **Reducing the technical barriers to accessing EGIs resources by investing in the virtual research environments needed by the research community.** Research communities need virtual research environments (that span from the low-level e-infrastructure services to the simple easy to use user interfaces on desktop and mobile devices used to access them) personalised to their particular needs and available to them across Europe. This software needs to be composed of open extensible software solutions that can be reused across the different virtual research environments needed by different research communities

The **sustainability of the EGI community** as a whole can only be achieved by ensuring that the diverse providers (whether projects or organisations) and the capabilities that they contribute to make

³ <http://www.egi.eu/case-studies/>



up the rich open loosely coupled ecosystem are themselves sustainable. The EGI Council has an important role to play by ensuring that the policies set by and for the community are open and inclusive across both public and private sector providers. Having multiple providers for each capability, which can evolve and possibly be replaced over time, is essential in ensuring long-term sustainability of the ecosystem .

Funding is certainly a key element of sustainability; but the funding for innovation leading to establishing or improving the operational efficiency and technical capabilities of the ecosystem needs to be decoupled from the funding that is needed to operate the ecosystem. Many of the providers in the ecosystem 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.

Continued investment is needed from European and national funding bodies to support the development of the human capital that makes up the EGI community through networks of excellence, local EGI champions and forums that continue to develop the interactions between different groups. Funding is required to support new innovations in the development of the operational tools needed by the operations community. Likewise, innovations that improve the capability and usability of the virtual research environments needed to support research communities should be driven by the researchers that use these environments.

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

Vision: To support the digital European Research Area through a pan-European research infrastructure based on an open federation of reliable services that provide uniform access to computational, 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 and world-spanning research.

Core Values:

- **Leadership:** EGI is a leading pan-European infrastructure, integrating worldwide computing, storage and data resources to support an ecosystem built on innovation and knowledge transfer.
- **Openness:** EGI operates collaboratively 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.

This strategic plan defines some of the initiatives that will take place within the EGI community over the next two years, supported through the FP7 EGI-InSPIRE. Areas for future coordinated investment in innovation from national and European funding bodies are identified that will help to establish the technical foundation of the digital ERA through the continued support, development and interchange of best-practice within individual national grid infrastructures. This allow the rapid evolution of EGI's activities to be a key enabling foundation of the digital ERA⁴ by supporting the transnational deployment and uniform operation of virtual research environments for simulation, data sharing and data analysis activities customised for the needs of individual multi-disciplinary research collaborations of all sizes and at a European scale.

⁴ part of the European Commission's Innovation Union initiative - http://ec.europa.eu/research/innovation-union/index_en.cfm?pg=home



TABLE OF CONTENTS

1	THE EUROPEAN GRID INFRASTRUCTURE	8
1.1	Background	8
1.2	Ecosystem	8
1.3	Stable Operation with Innovative Technology	10
1.4	Analysis	11
1.5	The Future	12
2	EUROPE 2020	13
2.1	Strategic Impact on EGI	13
2.2	What is the digital European Research Area?	13
2.3	EGI Personalised for the Researcher	14
2.4	Technology in EGI	15
3	STRATEGIC INVESTMENT AREAS	16
3.1	Overview	16
3.2	Community & Coordination	16
3.2.1	Strengths	16
3.2.2	Opportunities	17
3.2.3	Short-term Activities	18
3.2.4	Long-term Goals	18
3.3	Operational Infrastructure	20
3.3.1	Strengths	20
3.3.2	Opportunities	20
3.3.3	Short-term Activities	22
3.3.4	Long-term Goals	22
3.4	Virtual Research Environments	23
3.4.1	Strengths	23
3.4.2	Opportunities	24
3.4.3	Short-term Activities	25
3.4.4	Long-term Goals	25
4	FUNDING EGI'S NEW HORIZONS	26
4.1	Funding Sources	26
4.1.1	Community Funding	26
4.1.2	Horizon 2020 or National Innovation Funding	27
4.1.3	Horizon 2020 Support Funding	27
4.2	Planning Scenarios	28
4.2.1	Scenario 1: Zero EC Funding	28
4.2.2	Scenario 2: Minimal EC Funding	28
4.2.3	Scenario 3: Expanded EC Funding	28
5	CONCLUSIONS	30
5.1	EGI in 2012	30
5.2	Vision, Mission and Core Values	30
5.3	EGI during Horizon 2020	31
6	ANNEX 1: INDIVIDUAL RESEARCHERS IN THE EUROPEAN RESEARCH AREA	32



1 THE EUROPEAN GRID INFRASTRUCTURE

1.1 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.

Globally, e-Infrastructures are now an essential tool available to researchers and collaborative research communities across all digital sciences providing the means to rapidly move data and perform calculations at a previously unparalleled scale. Driven by the needs of their individual user communities, national or regional resource providers' work together to ensure that their services are interoperable and interoperate to support the needs of the collaborative multi-disciplinary research communities that they support.

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.eu⁵), 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 21,000 users submitting over 1.2 million jobs a day over more than 15 different disciplines. A selection of case studies relating to the work done by some of these researchers is available on the EGI website⁶.

1.2 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

⁵ <http://www.egi.eu>

⁶ <http://www.egi.eu/case-studies/>

software development activities or the platform operations role) but their roots can be found within EGI's current portfolio of activities.

A healthy sustainable ecosystem, i.e. with the capacity to endure, implies having a structure where some providers can fail and be replaced by others without damaging the ecosystem as a whole. The different individual providers 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 providers 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 provider 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, where a platform is the virtual research environment that needs to be integrated, deployed and operated to meet the needs of different research communities:

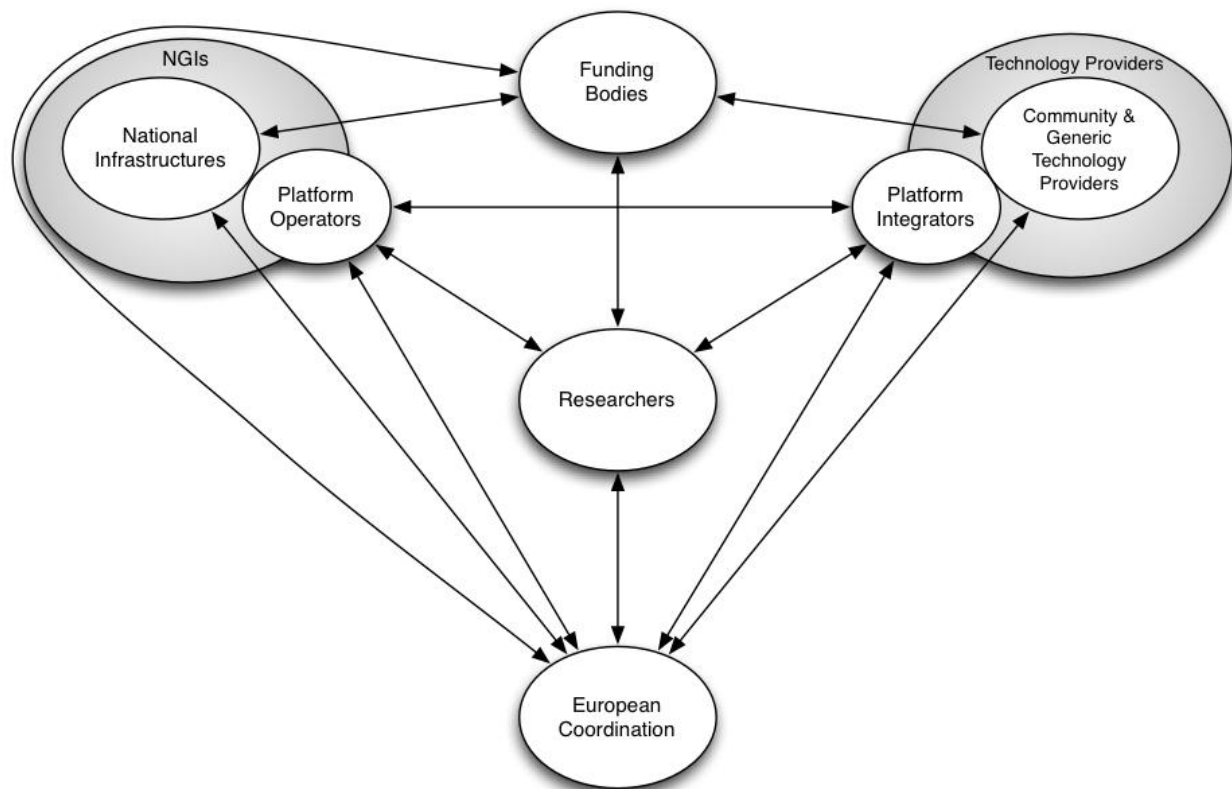


Figure 1: Overview of the EGI Ecosystem

- 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 researchers that potentially span different disciplines in different organisations



across different countries that have structured themselves to tackle a ‘grand challenge’ within their own scientific community (e.g. WeNMR).

- **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 project⁷ 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 the product teams within 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. Coordination also needs to take place between other European e-Infrastructure providers (e.g. DANTE for networking, PRACE for high performance computing, commercial cloud providers and projects such as EU-DAT for data) and e-Infrastructures around the world in order to provide an integrated platform to support the digital European Research Area and its international collaborations.
- **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.

1.3 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 IGE⁸). 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

⁷ <http://www.eu-emi.eu/>

⁸ <http://www.ige-project.eu/>



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 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.

1.4 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 21,000 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? How can the needs of private sector researchers be met with private sector resources using the innovative data analysis technologies generated in the public sector?
- **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 bringing the digital European Research Area online?

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 or much harder 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.
- A culture of open collaboration facilitated by technical services (application database, training marketplace, etc.) and human interaction (forums, workshops, blogs, newsletters, etc.).

Weaknesses:

- No clear standalone 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.



- For researchers or research communities without resources there is 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 that are unable to influence individual resources centres or bring their own physical resources to the infrastructure.
- Technical barriers are perceived to be high, leading to a slow uptake with research communities not already using the infrastructure due to the lack of a clear standards based architecture to build upon.
- No current business model to directly charge user communities for the utilisation they make of core central components appropriate to community size and if necessary community facing services that belong to other communities. No inter community economy with real exchange of currency and hence real value.

1.5 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 transnational e-Infrastructure across Europe through the uniform deployment of services to access compute and storage resources that will be the foundation of the digital ERA.
- Attracting further research communities by exploring the use of a uniform interface to cloud compute and storage resources that researchers can use to establish the virtual research environments needed to support their 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.
- Ensuring that Europe retains its global leadership in e-Infrastructure by working with other established and emerging national and regional e-Infrastructure providers around the world to provide service interoperability and interoperation to meet the needs of European researchers and the global collaborations.

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.



2 EUROPE 2020

2.1 Strategic Impact on EGI

EGI's strategy is 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 Europe⁹ 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 of the digital 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 digital ERA, EGI is capable of generating a substantial socio-economic impact for European society¹⁰.

2.2 What is the digital 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 digital 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 foundation of this ICT integration in the ERA – the digital 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.

⁹ http://ec.europa.eu/information_society/digital-agenda/index_en.htm

¹⁰ https://wiki.egi.eu/wiki/Europe_2020_actions



- **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 is one where its individual components deliver value to each other and are kept in balance by appropriate processes that provide feedback and governance.

2.3 EGI Personalised for the Researcher

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 digital 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' (the majority of the 2.8 million researchers in Europe) made up of smaller individual researchers and research collaborations many of which are based exclusively within a single country. 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.

The support for these different researchers is built around cooperation of independent NGIs who focus primarily on meeting the needs of their local researchers locally. However, the NGIs and domain specific resource providers (e.g. EIROs and ESFRIs) recognise that an increasing number of these researchers will need to increasingly access or share electronic resources with others as part of international collaborations. To simplify such transnational cooperation, the NGIs have established EGI (and EGI.eu) to provide the coordination needed between European NGIs to ensure uniform access to services across Europe.

It is the NGIs that contribute the bulk of the resources to EGI and are the foundation upon which it is built. They are the primary contact point for many of the researchers using the infrastructure, and through the customisation of the virtual research environments made available to the researchers, the NGIs are the main route to accessing the 'long-tail' of independent researchers. The virtual research environments will be composed of local services provided by the NGI, services provided by the researcher's own community, and services provided centrally across the whole of EGI. For established research communities (the 1%) EGI offer the means to have their software deployed and operated at European scale using NGI resources, or to have the services in the research communities own resource centres across Europe be integrated with EGI's operational infrastructure, or some hybrid combination of these options.

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.

2.4 Technology in EGI

EGI has the proven technological and operational experience to integrate both local and national resources into an integrated European environment. Establishing a Core Infrastructure Platform that can be used by researchers (or those acting on their behalf) to deploy their own virtual research environments as part of a niche research community or by individual researchers empowers the researchers using EGI to use the software that best meets their needs. This flexibility and choice is essential in order to provide an e-Infrastructure for the long-tail of smaller research communities needing to undertake their multi-disciplinary collaborative research.

The technology 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.

Exploring the use of cloud services in EGI's production infrastructure alongside its existing services is central to having a Core Infrastructure Platform. Federating the private clouds being established by its resource centres, would enable EGI to provide Infrastructure as a Service (IaaS) solutions to meet the needs of the research community which are not being met from commercial IaaS offerings. These would leverage existing national investments in networking and hardware infrastructures and EGI's expertise in using grid technologies to provide uniform access to IaaS capability across Europe. 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.

3 STRATEGIC INVESTMENT AREAS

3.1 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 digital 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 and the ecosystem it supports through communication, outreach, support and marketing events.
- **Operational Infrastructure:** EGI federates a European-wide operational infrastructure structured along geographical regions (normally counties) or research communities comprising over 350 resource centres that will be used to:
 - Support the currently deployed commonly used functional services and their evolution in demand to the use from current and new research communities.
 - Help pan-European research infrastructures scale out the uniform operation of their functional services across Europe.
 - Integrate the institutional private clouds emerging in the public sector with public clouds coming from the commercial sector to provide a uniform capability to new research communities to deploy and operate the virtual research environments they need across Europe.
- **Virtual Research Environments:** A key requirement to the wider scale adoption of e-Infrastructures is the ability for individual researcher and research collaborations to personalise the virtual research environments (spanning the low-level platform services to the user interface used by the researcher) needed by a particular research community 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.

3.2 Community & Coordination

3.2.1 Strengths

The EGI has developed from a domain-specific activity focused on a handful of resource centres to an infrastructure that supports multi-disciplinary research communities with usage across many different disciplines. This has been achieved by developing a broad operations 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 partner e-Infrastructures around the world. This uniform service access benefits the researcher by providing a single market for services regardless of the provider 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 nationally by the NGIs and coordinated across Europe by EGI.eu.

The EGI Council with representation from the national (NGI) and domain specific (EIRO) resource providers 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 data providers) and with comparable providers around the world (e.g. Open Science Grid, XSEDE) and initiatives in sub-Saharan Africa, Latin America and Asia Pacific. 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.

3.2.2 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. Key to this process is the support that can be offered through the NGIs. Forums that allow these independent service providers to meet each other, meet representatives of national and European 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 and rationalising the technical requirements across the operations and user communities; managing the internal and external policy development activity through representation at EC, e-IRG and other policy related meetings; and establishing and supporting a

governance model across the different aspects of the community through briefings on Horizon 2020 and the ERA 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 digital ERA and is aligned with the policy initiatives of the EC. **Example:** Providing briefing and planning documents to stakeholders on the ERA. A network of NGI International Liaisons to bridge European and local human networks. Acting as a broker between research communities by aggregating demand for peak or bespoke resources from public sector or private sector resource providers.

- **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.

3.2.3 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 quickly tackle specific community issues within a short time scale. Virtual teams are different to the Virtual Research Communities established within EGI to provide representation to researchers within similar virtual organisations (whether by discipline or computational technique) in their use of EGI.

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 brought together at a European level and duplicates eliminated to provide a European wide analysis of activity.

3.2.4 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 between the ICT specialists in the individual research communities and national resource providers that needs to continue to strengthen the EGI ecosystem. 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. These open forums encourage participation from outside Europe and allows collaborations with initiatives similar to EGI throughout the world. EGI.eu is in a position to lead on behalf of the EGI community international collaborations with other e-infrastructures in order to build a global infrastructure to meet the needs of worldwide research collaborations.
- **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, link them to the area of EGI that they need, and support the improvement of the human capital in local research communities through the introduction of training and education on accessing NGI (and then EGI) resources into the local curriculum.
- **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 improve the communication between the different European e-Infrastructure providers (PRACE, TERENA, DANTE and now the data community represented by the EU-DAT



project) 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. Closer alignment and integration of the service delivery function across these organisations 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.

3.3 Operational Infrastructure

3.3.1 Strengths

Currently, at the heart of EGI is the federation of institutional compute-clusters into a single integrated infrastructure to provide high throughput data analysis, data storage and simulation. EGI has built on the experience and technologies of the projects that preceded it to produce an operational infrastructure that provides secure data movement and access to compute and storage resources in over 350 resource centres, and over 300,000 cores distributed across more than 50 countries by federating national resources that now supports over 1.2 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 centres 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.

3.3.2 Opportunities

The current operational infrastructure and expertise will continue to be used to deploy and operate the current functional services used by EGI's currently supported research communities. While maintaining the current capabilities, additional functionality that will allow EGI's operational infrastructure to integrate services from other communities running in their own data centres and to support the integration of private institutional clouds and public commercial clouds will be developed in order to give research communities the opportunity to select the provider that best meets their needs. While maintaining the current capabilities these developments will support 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. Re-using EGI’s monitoring, accounting and support infrastructure to support EU-DAT’s deployment of its data services.

- **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.eu working with the community’s appointed technology provider and with the support of the relevant NGIs would coordinate the validation, deployment and operation a community specific set of services across Europe on behalf of the WLCG collaboration.

Private institutional clouds are beginning to be established across Europe to meet the needs of local researchers for more flexible access to compute and storage capability. A decade ago it was such work around commodity cluster computing that established the predecessors to EGI, so it is only natural that research communities now want to see the integration of such private institutional clouds at a European level. Established grid technologies can be re-used where needed to provide consistent uniform access to cloud based storage and compute capability. While maintaining our current service offering establishing uniform access to cloud resources into the operational infrastructure across Europe will provide a significant new capability for communities to integrate and run the services that they want to run wherever and whenever they could gain access to 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. Integrated cloud resources 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:** A social scientist deploying a virtual machine imaged generated by their local IT staff into a remote cloud resource to analyse a data set stored locally to the remote cloud.
- **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 can be easily accessed by its researchers.
- **Integration of commercial public cloud resources into EGI alongside public sector resources.** Commercial public cloud resources operate at economies of scale that provide cost-effective resources for some application workloads with defined service level agreements. Having such



resources integrated alongside public sector resources gives research communities the opportunity to select the best option to meet their unique needs. Integrating these resources through the emerging cloud standards and their associated open source reference implementations provides the best route towards an open ecosystem. Interoperability through agreed open standards allows different individual sites to select the cloud management system that suits their individual local needs. **Example:** Through the HelixNebula project EGI.eu will work with public and private sector resource providers to establish relevant policies for technical integration.

3.3.3 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. The EGI Federated Cloud Task Force is exploring how the established technologies and techniques already established in EGI can be re-used to integrate the private clouds emerging all over Europe. 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 alongside the currently offered services.

3.3.4 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. Continued investment will be needed to develop the scalability of the operational tools across increasing number of resource centres, increasing number of supported research communities, and increasing number of services (from the different virtual research environments) that form part of the operational infrastructure. These services will include the current core functional services, the emerging cloud services around storage and compute, and the services from the virtual research environments coming from other research communities that are integrated into EGI. 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 through funds coming from the countries and research communities benefiting from the integration, will continue to drive further operational efficiency gains. The funds needed to support the necessary innovations and integration activity needed to achieve this would be funded through specific projects. 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 improve their human capital that will lead to an 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.
- **Investment in open-source core infrastructure platform:** As the production infrastructure continues to grow in scale and scope additional innovations will be needed in the federating infrastructure – the Core Infrastructure Platform. These include maintenance and development of the current services needed by all research communities while exploring the integration of cloud related services for compute and storage being deployed by many of EGI’s resource centres. Domain specific infrastructure services should be seen as being included as being part of that research community’s virtual research environment.
- **Investment in open-source technologies for cloud provisioning:** Higher-level services that enable the optimal use of a federated standards-based cloud infrastructure by providing to the platform operator provisioning and autonomic management tools will become critical in exploiting the resources within EGI. Many of these services are already in advanced development within other EC projects (e.g. Conrail project¹¹) and need to be brought to a production quality through collaboration with the original development teams.
- **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 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.

3.4 Virtual Research Environments

3.4.1 Strengths

Virtual Research Environments (VREs) 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.

¹¹ <http://conrail-project.eu/>



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 XSEDE¹²) through their Science gateway programme and the Globus Online¹³ 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.

3.4.2 Opportunities

The use of the web and dedicated applications 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 easily 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 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 USA¹⁴.
- **Using mobile devices to access EGI services:** 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 has been 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. Developing a toolkit designed around accessing EGI's collaborative services from the mobile web and supporting its uptake 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.
- **Open Software Development Ecosystem** Software development activities over the last decade have shown the difficulty in providing a single software environment that meets the needs of all

¹² <https://www.xsede.org/>

¹³ <https://www.globusonline.org/>

¹⁴ http://www.collab-ogce.org/ogce/index.php/Main_Page



research communities. Having a core infrastructure platform (see Section 3.3.4) that other developers can build upon is an essential part of building an open ecosystem and at the heart of the virtual research environments initiative. While providing structures such as the Applications Database and the proposed Platform Store that promoting software re-use to eliminate duplication and reduce the time to satisfying new use cases by adapting existing solutions, the ultimate choice as to what software to use lies with the researcher and their community. EGI's role is to facilitate operation at a European scale so that it does not impact the work of others. **Example:** EMI's ScienceSoft initiative provides an approach to bringing software development teams and research communities closer together through the use of social and reputation networks.

3.4.3 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. A balance needs to be struck between tightly customised solutions that work for a single research community or researcher, versus generic solutions that allows a single investment to bring a return across multiple research communities.

3.4.4 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.

- **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.
- **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 sharing¹⁵.

¹⁵ MS510: EGI Platform Roadmap <https://documents.egi.eu/document/970>



4 FUNDING EGI'S NEW HORIZONS

4.1 Funding Sources

Funding, regardless of the source, provides a means of driving change towards the priorities of that funding source: 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 elsewhere¹⁶.

4.1.1 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 as a result of the ratification process required of an ERIC at a nationally could bring additional national recognition and funding commitment from the member states.
- A **'heavyweight' EGI ERIC** would build on a 'lightweight' ERIC and provide an additional service to the EGI community through the purchasing and management by NGIs of ICT facilities (potentially funded by the EC and by member states taking advantage of preferential ERIC taxation options) to meet the needs of the research communities it supports.

Both of these models remain under discussion at this time.

However, community funding models relating to the direct use of the infrastructure must be considered in the short-term to be a limited 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 are through a legal entity able to join EGI.eu as an associated partner to influence the governance of EGI, or collaborate with EGI in a way that does not directly bring cash into the ecosystem, but the partnership at a European level can translate into

¹⁶ D2.18 Evolving the EGI Business Model. <https://documents.egi.eu/document/1040>.



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.

4.1.2 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.

4.1.3 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 the exchange of best practices between NGIs; by raising awareness of the different activities taking place within the ecosystem, 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.



4.2 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.

4.2.1 Scenario 1: Zero EC Funding

Scenario: No further EC funding for the coordination, maintenance and operation of the infrastructure is forthcoming 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 would no longer remain a dependable resource then research communities would return to using their own solutions resulting in fragmentation and eventual collapse of a European wide e-Infrastructure for computing and storage.

4.2.2 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.

4.2.3 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. This could use a model where EGI.eu, on behalf of the EGI Community, is asked to run an open call for proposals to develop innovative solutions that build on existing work. Such a model could be in addition to the current call model made by the EC.

Long-term Consequence: Investing in research led collaborations across strategic societal areas will help build engagement within the research community and help establish the digital 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 digital ERA. This scenario would broaden access to e-Infrastructures and start bringing some of the potential 2.8M researchers in Europe online.



5 CONCLUSIONS

5.1 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 is a vital asset to the research communities that have adopted it for their routine research activities and needs to be maintained and incrementally developed to support their evolving requirements. However, the production infrastructure in its current form 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 (EMI and IGE) 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.

A complimentary approach would be to give research communities the capability to deploy across Europe their own personalised virtual research environments by being able to access storage and compute resources through cloud interfaces. Many of the resource centres within EGI are already beginning to deploy private clouds to meet the needs of their local research communities. Ensuring uniform access to these cloud resources across Europe as part of EGI's production infrastructure would give individual research communities great flexibility as to the what, when and where they deploy their software to meet their community's needs. Research communities would have the use of virtual machine images prepared and operated to meet their specific requirements. By allowing different platforms to be operated and deployed on demand within the production infrastructure many potential new research communities who already have their own software frameworks could 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.

5.2 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 digital 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 and world-spanning research.

Core Values

- **Leadership:** EGI is a leading pan-European infrastructure, integrating worldwide computing, storage and data resources to support an ecosystem built on innovation and knowledge transfer.
- **Openness:** EGI operates collaboratively 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.

5.3 EGI during Horizon 2020

For the digital European Research Area to come online by the end of Horizon 2020, an investment programme needs to be established immediately that extends EGI's core activities of deploying, operating, monitoring, accounting and supporting virtual research environments through a federation of national resource providers for multiple research communities at a European scale. 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 foundations that have come from previous investments. The ability for a project to provide a reusable legacy needs to be a key enforced assessment criteria. Alongside reusability, support needs to be provided to maintain the output from projects which are seen to have value and use within the community. The continued maintenance and operation of the digital 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.

6 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) ¹⁷	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 of 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

¹⁷ 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.