EGI: an Open e-Infrastructure Ecosystem for the Digital European Research Area

Steven NEWHOUSE, Sergio ANDREOZZI, Sy HOLSINGER, Damir MARINOVIC[1]

[1]*EGI.eu, Science Park 140, 1098 XG, Amsterdam, The Netherlands*

Tel: +31 (0)20 89 32 007

Email: {steven.newhouse, sergio.andreozzi, sy.holsinger, damir.marinovic}@egi.eu

Abstract: Bringing the digital European Research Area (ERA) online means modernising Europe's research infrastructure by promoting open science through the availability, accessibility and reuse of scientific data and results, the use of web-based tools that facilitate scientific collaboration and ensuring public access to research. As the European Grid Infrastructure (EGI) is the largest European distributed computing infrastructure providing 24/7 access to large scale computing, storage and data resources through a federation of national resource providers, it allows scientists from all disciplines to make the most out of the latest computing technologies for the benefit of their research. This paper describes the methodology and approach for defining EGI's role in bringing this digital ERA online. The work presented defines the roles and functions of EGI as an open ICT ecosystem, required service redesign, the added value of EGI for the European research communities and demonstrates the role that EGI plays in contributing to the Europe 2020 strategy for social-economic impact.

1. Introduction

Today, science is no longer exclusively produced in a single research lab or within national boundaries. In the coming years, modern scientific grand challenges will call for integrated solutions, cross-country and multi-disciplinary collaborations and above all, computing power to analyse vast amounts of digital data being generated from instruments and simulations. This trend for data-intensive science and collaborative discovery will continue to develop and will form the basis of future innovations in science, society and business in Europe.

In the wider context, the European Commission has unveiled a growth strategy for the EU to become a smart, sustainable and inclusive economy within the next decade. The 'Digital Agenda for Europe' (DAE) and 'Innovation Union' (IU) are two of the seven flagship initiatives that will bring crosscutting benefits to European research and society as a whole.

One of the driving forces behind this strategy is the digital European Research Area (ERA) - a transnational effort combining research centres with on-going programmes and projects aimed at building multidisciplinary collaborations and enabling rapid knowledge transfer across borders. Bringing the digital ERA online means modernising Europe's research infrastructure by improving the capacity of the human capital, digitising workflows and providing the physical infrastructure needed to deal with a digital research future. E-Infrastructures play a pivotal role in this context, as they allow scientists and researchers to share information securely, collaborate with colleagues worldwide and manipulate and analyse complex data faster and more efficiently in ways otherwise not possible.

The European Grid Infrastructure (EGI), a distributed computing e-Infrastructure for e-science, traces its origin back to the early pioneering days of distributed computing research

that took place in the early 2000's as local services started to emerge for compute- and data-oriented applied researchers who wanted to share these services with their collaborators (e.g. high-energy physics). Over the years, thousands of applied researchers have integrated innovative technology from Europe and beyond into their daily research activities supporting a multi-disciplinary research communities (e.g. life sciences, biology, chemistry, astronomy), developing technology and innovation and increasing in geographical scope through collaborations with areas outside Europe covering North America, Latin America and Asia Pacific.

Currently, EGI is running as a 24/7 production infrastructure, which comprises an ecosystem of researchers and research collaborations, technology providers, technology integrators, resource providers, over 350 resource centres and managed by defined coordinating bodies and other functions in almost 40 countries. EGI is used by over 21,000 researchers from 15 different disciplines who run in excess of 1.4 million jobs a day for their intensive data analysis research. Across the infrastructure, more than 270,000 CPUs and 150PB of disk and 140PB of tape storage are available with roughly a 95% availability and reliability rate.

2. Objectives

The full implementation and potential of the digital ERA online will only be achieved through the acceleration of e-Infrastructure development, which will enable the accomplishment of the 'fifth freedom' - the free circulation of knowledge and technology across Europe. EGI is enabling this vision by offering (and continuously innovating) an open e-Infrastructure to support distributed data analysis and connecting researchers with reliable and innovative ICT services they need for their work in a digital online future. The objectives of this paper are to describe EGI's methodology for defining its strategic contribution in bringing the digital ERA online, identifying the overall benefits of this work within the Europe 2020 strategy for social-economic impact and analysing value networks to define the necessary roles and functions needed to transform EGI into an open ICT ecosystem [1] for long-term sustainability.

3. Methodology

As described, EGI is a rich ecosystem of research communities, resource and technology providers, coordinating bodies and funding agencies each offering varying sets of services for the benefit of the whole. But, in order to better capture the value exchange between the participating actors and to better define the best way forward in bringing the digital ERA online while aligning with EU-wide priorities, a number of strategic analysis tools have been used over the last year, which helped shape the future direction of EGI defined in the EGI2020 Strategy [2].

First, the EGI ecosystem was defined in terms of roles, responsibilities and value exchange to clarify the relevant interactions and understand its possible evolution. Such an analysis required several iterations with the key stakeholders and led to the identification of two new roles, platform operator and platform integrator, that potentially could be performed by independent organisations within their own business space. In terms of value exchange, it should be noted that EGI is not a linear value-chain, but it is more structured into a multi-stakeholder value creation network (Figure 1). This value was therefore analysed using the Value Network Analysis [3] and validated through a survey questionnaire [4]. The latest description of the ecosystem was presented in the EGI 2020 strategy.

Following the EGI ecosystem and value network analysis, a SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis was performed for each of the roles

within the ecosystem and for EGI as an integrated service provider. The matching of strengths to opportunities and the evaluation of threats and weaknesses have been the base to further develop business models for EGI and its European coordination body, while describing the business space and business modelling framework for resource and technology providers [5].

In parallel to these activities, a EU2020 actions analysis was conducted with focus on the two flagships initiatives: Digital Agenda for Europe (DAE) and Innovation Union (IU). The actions relevant for EGI have been identified to analyse and understand its potential contribution to the EU-wide priorities.
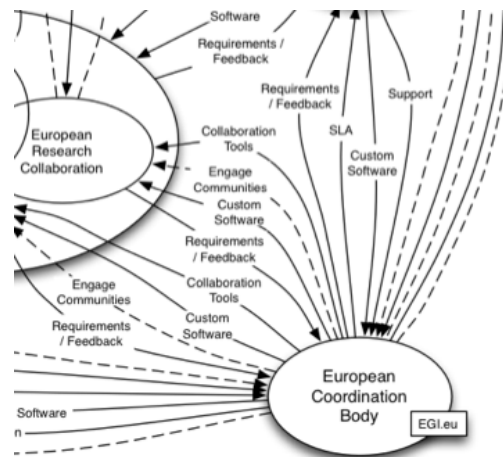

Figure 1 - EGI Value Network Analysis - Snapshot

4. Business Case Description

EGI provides uniform access to large scale computing, storage and data resources across Europe through a federation of national resource providers that allows scientists from all fields of research to make the most out of the latest computing technologies for the benefit of their research activities.

4.1 Services

Through EGI, research collaborations in Europe can manage their scientific data across different independent resource centres and perform high throughput computing across organisational boundaries, thus being able to scale out the capacity of the single data centre to meet their needs. Technical services are offered through open APIs and a number of scientific gateways are flourishing (i.e. community-specific set of tools, applications, and data collections that are integrated together via a web portal or a desktop application, providing access to resources and services from EGI).

EGI also offers community-oriented tools such as a virtual organisation membership service to create named list of users organised in groups and sub-groups to which access permissions to resources are assigned following a role-based model. Through this mechanism, resources from distributed resource centres are allocated to a named collection of users and the manager of the collaboration can add/remove users in an autonomous way. Other services available to users are an application marketplace to enable the individual researchers to share their problem solving tools with other and stimulate reuse, and a training marketplace to share/discover learning material for the usage of the infrastructure.

4.2 Applications

EGI supports over 21,000 researchers working on a wide range of scientific disciplines. The discovery of the Higgs boson at the Large Hadron Collider particle accelerator in CERN is the most high profile example of the benefits of grid computing to science. Without distributed computing facilities across Europe as part of EGI and in collaboration with Open Science Grid in the US [6], it would have been impossible to share, analyse and interpret the experimental data. In the run up to the 4 July 2012 announcement, the scientists of the ATLAS collaboration processed data from one quadrillion collision events (that is one million billion). Only a 24/7, reliable grid infrastructure is able to cope with this challenge.

The Life Sciences are the second most active research community in EGI, following High Energy Physics. EGI resources are being used every day to identify respiratory diseases still unknown to science [7], help doctors to monitor the onset of Alzheimer's Disease or design better and smarter antibiotics to fight multi-resistant fungal infections with minimal side effects [8]. Grid computing is especially helpful when researchers need to run multiple small calculations in parallel. Structural biology benefits immensely from this model, because it allows researchers to investigate the properties of millions of molecules in a reasonable timeframe. This is very instrumental in the search of new drugs against infectious diseases (e.g. malaria), or new anaesthetics for use in hospitals [9].

Grid computing also has an impact in the Earth Sciences. Seismologists have used EGI resources to calculate peak ground acceleration and other parameters useful to help local authorities to highlight the areas more likely to be damaged by earthquakes [10]. Palaeontologists can also gain new insights from fossils: a grid-enabled simulation has helped a team to conclude that some dinosaurs were not as fast as we once thought [11]. Full descriptions of these use cases and others are available online [12].

4.3 Users and Target Groups

The rich EGI ecosystem was analysed and broken down into roles and responsibilities in order to improve clarity on value exchange between different actors and set out the context for opening it up to improve efficiency and flexibility (Figure 2).
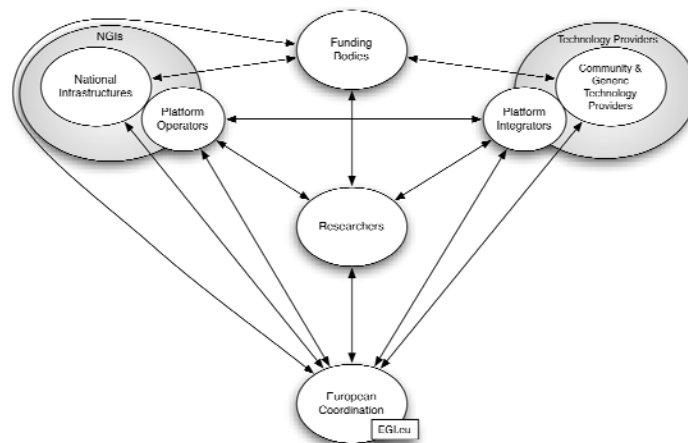

Figure 2 - The EGI Ecosystem

At the core of the ecosystem are the 'researchers', i.e., consumers of e-Science services that are supported by e-Infrastructures to perform their digital research; they are interested in services that can rapidly adapt and integrate with their workflows to conduct their

research, achieve faster results, publish first and gain the recognition of their peers. They can be organised in research collaborations either on a local national, European level.

EGI's main stakeholders are the 38 National Grid Infrastructures. The governance model slightly varies within each country, but each follow a general set criteria. It starts by each country identifying a 'National Coordinating Body' as a recognised organisation given the exclusive responsibility for strategic decisions and overall management of a National Grid Infrastructure. The national infrastructures then comprise one or many 'Resource Infrastructure Providers', which manage and deliver the national operational infrastructure and 'Resource Centres' (i.e. the actual resource providers), which provide a minimum set of local or remote services compliant to well-defined IT capabilities necessary to make resources accessible to the researchers. Both are established within the national infrastructure and follow a defined Operational Level Agreement, which defines the minimum set of operational services and the respective quality parameters that an EGI resource provider or centre is required to provide. The role of 'Platform Operators' for particular research communities ensures that any research community specific service provided by the resource centres are operating effectively. In essence, this can be a position held within a resource centre or outsourced.

Regarding technology, EGI does not develop the software deployed in the grid infrastructure, all upgrades and new programmes are produced elsewhere, by independent technology providers (TPs). 'Technology Providers' deliver and support open-source or commercial software components that are put together by 'Platform Integrators' to meet the needs of particular research communities.

As EGI is a publically funded infrastructure for scientific research, all costs are covered through contributions by different 'Funding Bodies' such as the European Commission and national research councils or other organisations, which define policies and funding schemes to support the digital research.

The multi-stakeholder nature of an initiative such as EGI and the provision of services in the form of a partnership require strong leadership to build consensus and effective coordination activity. For this reason, a 'European Coordination Body' was envisioned and created as a legal entity under the Dutch law with the name of EGI.eu. This organisation established in February 2010 provides coordination across different national infrastructures and community building by driving forward the strategic direction of service delivery, attracting new research communities, exploring collaborations with both the public and commercial sectors and promoting EGI's role within Europe (e.g. DAE, ERA).

The EGI.eu business model [5] describes the provision of a number of global services for community coordination and infrastructure operation that are supported by the annual participation fees paid by the participants, National Grid Initiatives (NGIs) or European Intergovernmental Research Organisations (EIROs). Through participation in or fostering of EC-funded projects, EGI.eu will also lead focused innovation activities for the community. The establishment of EGI.eu is a key milestone in the overall EGI evolution and provides a similar governance role as a European Research Infrastructure Consortium (ERIC) legal entity [13]. The design and planning of EGI.eu started before the availability of this legal framework, therefore the ERIC option was not an available option, nevertheless, this will be reassessed in the near future [2].

4.4 Opportunities

According to the delivery model strategy defined in ITIL [14], EGI follows a partnership or multi-source model where a number of service providers have made an agreement to work together to provide an integrated set of services. Therefore, EGI is seen from a researcher's

perspective as a single point of access to a powerful ICT infrastructure to support digital research and collaboration.

The various roles and functions previously described contributing to the EGI ecosystem have their own autonomy and independence, each of which is able to clearly define their own individual business model.

Overall, the EGI vision is to support the digital ERA 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. Consequently, the EGI mission is 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.

It is important to note that EGI is a publically funded e-Infrastructure intended for scientific research with varying restrictions on commercial activities. However, there are several opportunities through Public-Private Partnerships, especially in the area of cloud computing that are being explored (e.g. pre-competitive research, joint technical development). As 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, supporting uniform standards-based transnational cloud access is a new capability that will not only attract new research communities, but business opportunity for commercial organisations. In addition, several companies have been created based on the innovation resulting from the infrastructure such as DCore providing 'Big Data' virtualisation and analytics IT services; SixSq offering agile cloud automation solutions and consultancy; Constellation Technologies selling big data bioinformatic products and services into the life science sector; and ScienceSoft assisting scientific communities in finding the software they need.

5. Future Developments

Having defined the vision and mission for EGI and its role within EU2020, strategic directions have been developed over the last year and focus on three main areas [2]. The first area is community building and coordination, which focuses on opening up the ecosystem to promote better 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.

The second area is the operational infrastructure. While continuing to provide the services for the currently supported research communities, the operational infrastructure needs to be evolved to allow other European scale research communities to monitor, manage and account their own services operating at their distributed facilities. This is a capability that EGI has developed for its needs over the last decade that is becoming essential for other research infrastructures with distributed services. In addition, the operational infrastructure needs to support the uniform transnational federation through standards of virtualised resources (such as institutional private clouds) from the public sector and public clouds from the commercial sector. This new capability will increase the flexibility of the infrastructure to respond to the needs of individual research communities and allow researchers to personalise the virtual research environment they use for their own research activities on top of a strong reliable infrastructure platform. This increased flexibility and customisation will ultimately result in the broadening of EGI's supported research base and contribute to its long-term sustainability.

The third area is Virtual Research Environments (VREs). Research communities need VREs that interface the users to the e-Infrastructure to reduce the technical barriers to accessing EGIs resources. This software should be personalised to the users' needs and

should be composed of open extensible software solutions that can be reused across communities.

The EGI 2020 Strategy [2] defines in more detail the initiatives that could take place within the EGI community over the next few years with coordinated investment in innovation from national and European funding bodies. Adaptation of this plan will allow the rapid evolution of EGI's activities to be a key enabling foundation of the digital ERA 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. Due to the size, scope and complexity of the ecosystem, implementation needs to be calculated and methodical taking place over the course of the next couple years. Each step has been outlined through a Technical Roadmap [15]. Ultimately, these activities will significantly reduce the technical barriers to accessing EGI's resources faced by research communities new to the infrastructure, and allow for the rapid customisation by the communities themselves for their essential applications that they want to run at scale across Europe.

In the area of scientific impact, EGI is also working towards improving the traceability of scientific publications that benefit from EGI. Tracking the scientific outputs based on EGI has been always difficult because of the geographically dispersed communities that use the infrastructure and the lack of well-defined processes and tools. A dedicated team is working to mitigate this issue by setting up tools and processes that enable the accurate tracking of scientific publications [16]. In this work, available solutions such as CiteULike [17] and Mendeley [18] are being considered as well as the open access policy for EC-funded research and the related OpenAIRE tool [19].

6. Benefits

EGI provides significant added value to individual researchers and research groups (national, European and international) who need to process, share and analyse large data sets generated from their research through its production infrastructure and services contributed by the participating organisations.

Overall, EGI:

- Ensures that researchers have access to uniform and reliable computing resources, enabling faster scientific results and avenues of multi-disciplinary research otherwise not possible;
- Allows researchers to focus on their research rather than managing their e-Infrastructures needs;
- Guarantees an integrated, reliable and uniform service provided across organisational and national boundaries thanks to the monitoring and operational services;
- Provides effective use of resources and ensures the most effective return on Europe's e-Infrastructure investments;
- Enables knowledge transfer and sharing of solutions through community events, an application database and a training marketplace;
- Promotes open science through the availability, accessibility and reuse of scientific data and results, the promotion or integration of web-based tools that facilitate scientific collaboration and by ensuring public access to research.

In support of the Digital Agenda for Europe, EGI will:

- 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 for e-Infrastructures through the promotion and adoption of open standards [20];
- Reduce inefficient research spending and stimulate innovation across Europe by optimising the investments from different sources into a common federated infrastructure and community;
- 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 that will allow 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.

7. Conclusions

To ensure a sustainable infrastructure to support open science for decades to come, EGI has laid out the framework for establishing itself as an open ICT ecosystem. This will ultimately be achieved through a combination of policies, strategies, processes, information, technologies, applications and stakeholders that together make up a technology environment for the main beneficiaries. By allowing researchers to customise the services within their virtual research environments and to provide them with the collaboration tools and platforms needed to build these virtual research environments will be critical for bringing the digital ERA online, and through it, will contribute to the Europe 2020 strategic agenda and deliver a substantial socio-economic impact.

There are a number of researchers, collaborations and infrastructures currently available throughout Europe, therefore, for the EU to become a smart, sustainable and inclusive economy within the next decade it cannot afford to work in silos. Over the past decade extensive knowledge and expertise, as well as technology, applications and tools and resources have been developed, which should be reused and leveraged where possible. The ultimate value of EGI is the collaborative spirit where researchers can come together and conduct science in ways otherwise not possible and continue to bring the social-economic impact on society at large.

Overall, this paper presented the methodology in defining the roles and functions of the ecosystem, established the rationale for service redesign to increase user flexibility, and the added value of EGI for the European research communities. The set of strategic activities performed over the last year define the long-term vision for EGI to complete the transition from its origins as a monolithic project-based infrastructure to a modular and open network of organisations and resources with defined roles and responsibilities to deliver value with personalised business models.

The final goal for the community is that by the end of the EGI-InSPIRE support project (April 2014), the implementation of the initial phases of the EGI2020 strategy [2] will have established an open ecosystem where individual actors supported through public and/or private funding will have been able to define their own added value and business models to support their activities. A technical roadmap for its delivery has also been defined [15]. This foundation will enable EGI to continue to play a key role in sustainably bringing digital the European Research Area online.

Acknowledgments

References

[1] Roadmap for Open ICT Ecosystems, Open ePolicy Group, Harvard Law School, http://cyber.law.harvard.edu/epolicy/roadmap.pdf (last accessed 23 May 2012)

[2] EGI 2020 Strategy, https://documents.egi.eu/document/1098 (last accessed 23 May 2012)

[3] V. Allee, O. Schwabe - Value Network Analysis and the true nature of collaboration – Digital book: http://www.valuenetworksandcollaboration.com/ (last accessed 23 May 2012)

[4] Value Creation in the EGI Ecosystem, https://documents.egi.eu/document/987 (last accessed 23 May 2012)

[5] Evolving the EGI Business Model, https://documents.egi.eu/document/1040 (last accessed 23 May 2012)

[6] https://www.opensciencegrid.org

[7] Hunting for viruses: finding a needle in a haystack, http://go.egi.eu/virus

[8] Designing better antibiotics, http://go.egi.eu/antib

[9] Stories from the grid: the cone snail and the search for powerful new anaesthetics, http://go.egi.eu/conco

[10] Modelling earthquakes in Thessaloniki, http://go.egi.eu/thess

[11] How fast could a T-rex run?, http://go.egi.eu/t-rex

[12] EGI Case Studies, http://www.egi.eu/case-studies/ (last accessed 23 May 2012)

[13] Community legal framework for a European research infrastructure consortium (ERIC), Council Regulation (EC) No 723/2009 of 25 June 2009 http://ec.europa.eu/research/infrastructures/pdf/council_regulation_eric.pdf

[14] The Information Technology Infrastructure Library (ITIL), http://www.itil-officialsite.com

[15] EGI Technical Roadmap, https://documents.egi.eu/document/1094 (last accessed 25 May 2012)

[16] EGI Virtual Team Scientific Publications Repository, http://go.egi.eu/vt-scipubrep

[17] Citeulike, http://www.citeulike.org/

[18] Mendeley, http://www.mendeley.com/

[19] OpenAIRE Project: Open Access Infrastructure for Research in Europe, http://www.openaire.eu/

[20] EGI Standards Roadmap, https://documents.egi.eu/document/721 (last accessed 23 May 2012)