



EGI-InSPIRE

EGI TECHNICAL ROADMAP

EU DELIVERABLE: D2.31

Document identifier:	D2.31-TechnicalRoadmap-V7.docx
Date:	08/05/2012
Activity:	NA2
Lead Partner:	EGI.eu
Document Status:	FINAL
Dissemination Level:	PUBLIC
Document Link:	https://documents.egi.eu/document/1094

Abstract

The EGI Technical Roadmap describes the currently planned activities within the next two years of the EGI-InSPIRE project. The roadmap aligns activities within EGI to its main strategic priorities of:

- Community & Coordination
- Operational Infrastructure
- Virtual Research Environments

Each section provides an overview of the key activities and references supporting activity specific material where available. The document will be updated in April 2013.



I. 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.

II. DELIVERY SLIP

	Name	Partner/Activity	Date
From	Steven Newhouse	EGI.eu/NA2	19/04/2012
Reviewed by	AMB & PMB		08/05/2012
Approved by	AMB & PMB		08/05/2012

III. DOCUMENT LOG

Issue	Date	Comment	Author/Partner
1	19/04/2012	First draft	Steven Newhouse
2	01/05/2012	Second draft	Steven Newhouse
3	02/05/2012	Final version	Steven Newhouse

IV. 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.

V. 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>

VI. TERMINOLOGY

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



VII. 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.



VIII. EXECUTIVE SUMMARY

This first EGI Technical Roadmap draws together the individual project activities for PY3 and PY4 (where available) and describes the interactions with other EC funded projects. These technical plans have been structured around EGI's strategic priorities of Community & Coordination, Operational Infrastructure and Virtual Research Environments. This enables supporting activities around marketing and communications, events, strategy and policy and NGI outreach to be oriented around the key technical activities emerging from developments around the:

- **Core Infrastructure Platform:** The management and uniform delivery of services hosted in different administrative domains to different European research communities.
- **Cloud Infrastructure Platform:** The prototyping integration of interfaces to virtualised resources using the Core Infrastructure Platform to provide European research communities with a federated cloud resource built upon resources in the public sector.
- **Collaborative Infrastructure Platform:** A set of services, some of which have been developed over the last decade and some of which will be delivered through new partnerships, that may be used internally by the Infrastructure Platforms but are also valuable for research communities to use within their Virtual Research Environments.
- **Virtual Research Environments:** The set of distributed software services, applications and tools that research communities use to undertake their daily science. Most of the software used within Virtual Research Environments is developed and managed outside of EGI-InSPIRE through collaborations managed through EGI or directly by the individual research communities.

EGI-InSPIRE will continue to support the operational infrastructure and the development of the services needed to deliver it while integrating these services so that they can be adopted by other research communities to monitor their distributed service infrastructure. Alongside the evolution of the operational infrastructure, the current functional services will continue to be integrated from the EMI and IGE projects into UMD and verified in staged rollout before being deployed into production. The work coming from StratusLab and other projects is being used to prototype a Federated Cloud that provides access to a resource centre's virtualised resources that can be offered alongside EGI's existing services

While much of the development work that relates to Virtual Research Environments takes place outside the EGI-InSPIRE project, there are services and tools within EGI-InSPIRE that are being developed and used within one research community, that can be re-used as the basis of building Virtual Research Environment for other communities.



TABLE OF CONTENTS

1	INTRODUCTION	7
2	COMMUNITY & COORDINATION	8
2.1	Introduction	8
2.2	Events and Workshop	8
2.3	Marketing and Communications	9
2.4	Strategy and Policy	10
2.5	NGI International Liaisons	11
2.6	Business Development Strategy	11
2.6.1	Community & Coordination	12
2.6.2	Operational Infrastructure	12
2.6.3	Virtual Research Environments	13
2.7	Summary	14
3	OPERATIONAL INFRASTRUCTURE	15
3.1	Introduction	15
3.2	Core Infrastructure Platform	15
3.2.1	Operational Security	16
3.2.2	Operations Integration	17
3.2.3	Operational Tools	17
3.2.4	Accounting	18
3.2.5	Availability and core services	18
3.2.6	Network Support	19
3.3	Cloud Infrastructure Platform	20
3.3.1	Management	20
3.3.2	Authentication and Authorisation Infrastructure	21
3.3.3	Accounting	21
3.3.4	Information Discovery	21
3.3.5	Monitoring	21
3.4	Collaborative Infrastructure Platform	22
3.4.1	Helpdesk	22
3.4.2	Software Repository	22
3.4.3	Software Verification	23
3.4.4	Staged Rollout	23
3.4.5	VM Image Marketplace	23
3.4.6	Applications Database	24
3.4.7	Training Marketplace	24
3.4.8	Client Relationship Management (CRM) System	25
3.4.9	Federated Authentication and Authorisation Infrastructure	25
3.4.10	Data Movement as a Service	26
3.5	Summary	26
4	VIRTUAL RESEARCH ENVIRONMENTS	28
4.1	Introduction	28
4.2	Generic Technologies and Services	29
4.2.1	Job Management Frameworks	29
4.2.2	Dashboards	29



4.2.3	Science gateways.....	29
4.2.4	Workflows.....	31
4.3	Community Platforms	31
4.3.1	UMD	31
4.3.2	High Energy Physics (HEP)	33
4.3.3	Life Sciences (LS).....	35
4.3.4	Astronomy & Astrophysics (A&A)	35
4.3.5	MPI	36
4.4	Summary.....	36
5	CONCLUSIONS	37



1 INTRODUCTION

The focus of the EGI-InSPIRE project is on delivering a pan-European production infrastructure to its supported research communities and to develop that infrastructure in response to their needs, and the needs of research communities new to EGI. Many of the software components that the infrastructure depends on are developed outside of EGI-InSPIRE but within the EGI ecosystem by other projects and organisations. This Technical Roadmap aligns the contributions from the whole EGI ecosystem into the three strategic areas defined in the EGI Strategy. Namely:

- Community and Coordination
- Operational Infrastructure
- Virtual Research Environments

As a result, this roadmap (which will be updated yearly) identifies the activity taking place within EGI-InSPIRE and in related projects that it is dependent on. It effectively provides an update to the DCI Collaborative Roadmap¹ that was developed in September 2010 identifying how the outputs of the DCI projects have contributed to EGI, and how projects funded outside of the immediate ‘DCI projects’ are helping to develop the policies, technical infrastructure and community around EGI.

This roadmap does not contain detailed technical plans. The individual activities within EGI-InSPIRE and the projects with which EGI-InSPIRE collaborates with produce their own detailed technical roadmaps at different intervals. The latest versions of these roadmaps are referenced and have been used to develop this high-level view of community wide activities.

¹ <https://documents.egi.eu/document/172>



2 COMMUNITY & COORDINATION

2.1 Introduction

The EGI community and the ecosystem that it supports have been developing over the last decade. However, it is only in recent years as it has matured that issues relating to structure and sustainability have really been considered. The establishment of EGI.eu and the EGI Council provided a coordinating focus and governance body for the structuring the delivery of the production infrastructure. The governance of the individual product teams that delivered the middleware components and the Virtual Research Communities (VRCs) that use EGI has been revealed in many cases to be less mature without a project structure to bring them together. It is not seen as EGI.eu's role to govern these external activities, but to work with those structures where they exist as part of an open ecosystem, and to support the establishment of these structures where they do not exist.

Further information on the activities in this section is described in the 'Annual Report on the External Relations Activity' in EGI-InSPIRE.

2.2 Events and Workshop

EGI Technical and Community Forums: The bi-annual Forum model will continue as a key tool in establishing the EGI ecosystem within the EGI community. The Forums are face-to-face meetings taking place every 6 months alternate between a technical and community focus, where the wider community of people involved in using, managing and contributing to the infrastructure get together to present and share experiences that are mutually relevant. These Technical and Community Forums are one week-long events, built on core EGI sessions, coupled with partner events and relevant co-located meetings. Over time the Forums will evolve towards an umbrella model with many inter-related and loosely coupled co-located events. These will include VRC-related events, other projects' events and, potentially, commercial events

- PQ10: Technical Forum, Prague, 17-21st September 2012
- PQ12: Community Forum, Manchester, 5-9th April 2013
- PQ14: Technical Forum – to be decided September 2012
- PQ16: Community Forum – to be decided April 2013

EGI workshops: EGI workshops provide an open and inclusive approach to defining specific community needs and processes, build consensus on technology roadmaps of interest to the community, gather new or clarifying existing requirements, collecting and sharing experiences and best practices for existing technologies. These EGI.eu facilitated workshops will:

- Be planned to take place during the EGI Forums, co-located with NGI events, with community specific events or at EGI.eu Amsterdam.
- Planned to be primarily face-to-face existing online video conferencing technologies such as WebEX or EVO will be explored as an alternative or supplement.
- Use EGI's network of contacts and outreach tools to bring inter-dependent individuals together.

An outline programme of events ranging from the domain specific to community wide cross-cutting issues is foreseen, including:

- PQ10: EGI Champions Briefing: Discussion of the proposed EGI Champions network and how it can be established:
- PQ10: Digital Cultural Heritage: Services and processes coming from EGI that can be used to support the DCH community.



- PQ10 & PQ14: Workflows Workshop: Integration of complex scientific workflow systems with EGI's services.
- PQ10, PQ12, PQ14 & PQ16: Federated Cloud Workshop: Workshop series to expand and report on the adoption of virtualised resources within EGI.
- PQ12: EGI Champions Workshop: Follow on from earlier briefing and virtual team activity to describe how the EGI Champions network will operate.
- PQ12: RI Workshop: Workshop of Research Infrastructure interested in adopting EGI's operational tools to support their own distributed production infrastructure.
- PQ12: Science Gateways Workshop: Bringing together web portals for research communities currently accessing EGI's resources and the technology providers active in this field.
- PQ14: Data Services Workshop: Workshop held in collaboration with EU-DAT to define how EU-DAT services could benefit from EGI's operational infrastructure.

2.3 Communications and Marketing

The Communications and Marketing Plan² is to outline strategies and means for successfully communicating EGI's outcomes across Europe and internationally by:

- Outline the stakeholders and participants in EGI communications
- Define the target audiences for the project and the project's results
- Summarise the message and means for reaching out to EGI's target audiences
- Describe techniques for monitoring and evaluating the progress in reaching out to these audiences

As an internationally funded project, it is important that any communications and marketing undertaken by EGI-InSPIRE reaches as wide a range of audiences as possible and also that messages are targeted effectively at these audiences. A range of means for communications will be used to reach out to the targeted audiences including:

Plans during PQ9-12 include:

- **Events** - EGI will focus on "going to the user", researching and selecting events focusing on individual scientific disciplines and promoting the benefits of grid to that community through targeted case studies including working with the VRCs to select events and generate the materials.
- **Website** - The engagement with the website through blog comments, social media sharing and time on the site will be tracked through the web analytics. Feedback will be sought from attendees at events and through surveys, so that adjustments can be made where needed. Further images and case studies will be added to the website during the course of the year, and Virtual Teams will be set up around new content areas, such as EGI's federated clouds initiative.
- **Materials and publications** – The EGI-InSPIRE presentation and paper will be updated and the EGI.eu communications team will also continue to produce a quarterly newsletter distributed by email, online and on collaborating websites such as scienceomega.com. Articles will be targeted at the 8000 subscribers to *iSGTW*, and to policy focused publications such as *Pan European Networks* and *Public Service Review*.
- **Media and public relations** – Continue to develop material for those new to EGI and the general public called "Stories from the grid" and will include posters, brochures and videos with links to more in-depth information depending on individual interest/technical knowledge.

² <https://documents.egi.eu/document/1070>



These will be marketed to the general public through associated articles and press releases. Media partnerships are in place with Tabor Communications, iSGTW and Hostingtecnews.com, the EC's media partners.

- **Social media** – EGI will build on its presence on FaceBook and Twitter, using them to launch further competitions and build community engagement. The EGI blog will be promoted and content across these channels will be segmented by audience interest: Policy, User Community, Technology & Operations and General. The new website also makes it easier for people to share EGI content via social media channel(s) of their choice.

These activities will enable engagement with EGI's primary audiences:

- New user communities (social sciences, environmental sciences, humanities etc.)
- Existing user communities (life sciences, physics, earth sciences etc.)
- Journalists
- General public
- National Grid Infrastructures (NGIs) and European International Research Organisations (EIROs)
- Resource providers
- Collaborating projects
- Decision makers
- Governmental representatives

and EGI's secondary audiences:

- Secondary schools, educational institutions
- Local communities in the partner countries

2.4 Strategy and Policy

The Strategy and Policy Team at EGI.eu coordinates and facilitates the development of internal and external policy activity within EGI. It also supports the definition and execution of the EGI strategy.

Part of this activity is the routine support given by the team to:

- General MoU activities to help establish an EGI ecosystem (e.g. establish new ones, track current ones, reporting, etc.)
- PQ12 & PQ16: Produce annual reports on MoU achievements.
- Contributing to the EGI Community and Technical Forums through sessions relating to EGI's governance, strategy, policy, sustainability and business models
- Policies and Procedures
 - PQ10: Review and subsequent revision of the Policy Development Process (PDP)
 - PQ12: Creation of the EGI Policy/Procedure manual
 - Security Policy Group:
 - PQ10: Top-Level Security Policy
 - PQ10: Accounting Policy
 - PQ11: SPG Glossary integration into EGI Glossary
 - PQ12: Data Protection Policy
 - PQ12: Revise VO Operation Policy and VO Registration Security Policy
- Community Communications
 - Articles published through external channels (e.g. iSGTW, e-IRG)
 - Articles in the EGI Newsletter (quarterly)
 - EGI Blog Posts (quarterly)

- Social Media (e.g. FB, Twitter)
- PQ10: Updated report/table on EGI's contribution to EU2020
- Community Resources
 - PQ9: Finalise the EGI Glossary and regularly review and update
 - PQ12: Annual review and update of the EGI Glossary
 - PQ16: Annual review and update of the EGI Glossary

Other activities being undertaken by the team are:

- EGI Compendium 2012: The data collection process for the EGI Compendium 2012 is already underway.
 - PQ9: Analysis of the collected data.
 - PQ10: Produce compendium report.
 - PQ10: Explore the development of the web tool to enable stakeholders to update the information directly.
- EGI Compendium 2013:
 - PQ11: Data collection starts
 - PQ12: Analysis
 - PQ13: Produce compendium report.
- EGI Compendium 2014:
 - PQ15: Data collection starts
 - PQ16: Analysis
 - PQ16: Produce compendium report.
- European Research Infrastructure Consortium (ERIC) activity
 - PQ9: ERIC Workshop
 - PQ10: Report on EGI transition to ERIC in D2.11 (PM29)
- PQ9: Define the EGI.eu Service Portfolio
- PQ10: Development of Balanced Scorecard Framework for EGI and EGI.eu
- PQ10: Development of the EGI.eu Business Model

2.5 NGI International Liaisons

Established during PY2 the NGI International Liaisons provides a structure for the development and coordination of non-operational activities. This role, and the resources that it can marshal within the NGI to support technical engagement with new research communities in conjunction with other NGIs, is becoming a vital asset in developing EGI's activities in areas involving policy, marketing, communication, events and technical outreach.

- PQ9+: Continue to develop the NIL model and establish short-lived virtual teams as required.
- PQ10: NIL meeting at the EGI Technical Forum
- PQ10: Development of plans to extend the national outreach of NILs through a 'Champions' network

2.6 Business Development Strategy

The development of the EGI Strategy highlighted the need to more clearly articulate the 'products' and 'services' that EGI offers to different consumers and to identify who those consumers are. Alongside the activities taking place within EGI that offer direct value, there are other activities that



have less-direct value, but are never the less valuable in establishing and growing a broad and open community.

2.6.1 Community & Coordination

The activities taking place to coordinate and develop the EGI community are valuable but it is challenging to define who the specific beneficiaries of the services are. Much of the work in this area is to bring together individual national activities taking place across Europe into a coherent European activity to support EU2020 objectives of an integrated European Research Area with open trans-national services. Therefore a mixed mode of support is foreseen:

- **Governance:** By adopting a federal model much of the governance is devolved to a national level. Bringing these national representatives together to provide European governance within their own community will continue to be supported through EGI.eu.
- **Events:** While individual events are expected to operate on a non-profit basis, experience over the last decade has shown that a persistent event secretariat that can provide professional coordination, management, marketing and communication around an event significantly increases the quality and therefore the effectiveness of the event. As many of these events are explicitly to structure the pan-European development of the infrastructure or its engagement across research communities, the continued support of the EC will be sought.
- **Marketing & Communications:** As has been noted, the main success of the large EGI events can be attributed to a professional marketing and communications team at the heart of the event secretariat. In addition, such a team provides a resource to the whole EGI community and its ecosystem to provide professional services. While direct funding from an EC project to provide such European level activity is seen as the main source of funding, the provision of professional services to organisations within the ecosystem (paid for by either private or public funds) and providing a service to its direct stakeholders within EGI are business options that need to be developed.
- **Policy:** The provision of accurate, understandable policy material for internal and external decision makers is essential and needs to be supported by those who derive the most value from the information. Internally, the EGI community should support the generation of material that is of benefit to its internal processes and stakeholders. Externally, collecting material that supports external policy making activities or accelerates the integration of national infrastructures into the ERA (e.g. the EGI Compendium) should be supported by the EC.
- **NILs:** Establishing a grassroots network throughout all NGIs, which is integrated across EGI provides, an excellent framework within which to connect geographically to researchers across Europe and across different disciplines. The network itself represents a significant investment in human capital that needs to be developed, but a framework from within which human capital of the research community could be developed. The growth of such a human network would underpin the digitalisation of the European Research Area and its integration at a European level. The establishment of such a network would need to be supported heavily by the EC.

2.6.2 Operational Infrastructure

The Operational Infrastructure provides coordination services (either to internal or external consumers) that are used to ensure the provisioning and delivery of high-quality services to researchers across different disciplines. These services are grouped into platforms to provide distinct products to different consumers.



- Core Infrastructure Platform: These are the services used by NGIs and their resource centres to integrate their local functional services targeted at consuming research communities into national infrastructures, and then to federate their national infrastructures hosting these local functional services into a uniform integrated European infrastructure.
- Cloud Infrastructure Platform: The adoption of a cloud model by EGI would enable research communities, or platform operators working on their behalf, to access virtualised resources at individual resource centres (where authorised) directly for the purposes of provisioning services for individual research communities. This would allow platform operators to provide customised solutions for the research community they were supporting independently of any generic service the resource centre might choose to provide.
- Collaborative Infrastructure Platform: These services help distributed communities (beyond the resource centres and the production infrastructure) to work more effectively together. While the services can benefit a similar type of consumer as the Core Infrastructure Provider, they can also enable collaboration and interaction between research communities and research communities and resource providers.

The use of these Infrastructure Platforms enables the resource providers to provide services to their consuming research communities. To promote adoption by these research communities the long-lived operation of these services needs to be assured by its providers. Therefore the cost of routinely operating these services needs to be assured through the EGI community with routine national operation being assured by national funds and routine European coordination to deliver the operational infrastructure to be assured through EGI.eu through the participant' contribution. The development and introduction of new innovative capabilities and operational techniques into the infrastructure through short-term projects needs to be supported by external funding sources such as the EC.

2.6.3 Virtual Research Environments

Each research community has different needs around the physical infrastructure they require for their data analysis and the software environments needed to undertake this work. Supporting different virtual research environments (VREs) for different research communities within the production infrastructure is essential in order to increase its attractiveness and uptake. Although there may be many different VREs, they may have many common services coming from independent technology providers that are assembled by platform integrators for operation by platform operators on behalf of particular research communities. Some of these services will be provided centrally specifically for the research community, some services may come from the Collaborative Infrastructure Platform available to all, and other services that are expected as part of the VRE will need to be deployed at each participating resource centre as part of that community's Community Platform.

One of the purposes of the Collaborative Infrastructure Platform is to provide services that help organisations undertaking a platform integration role to discover, assemble and deploy the VREs into production. Once in operation, there are services in the Collaborative Infrastructure Platform that can assist the platform operators in delivering the VRE to the consuming research community. There are two mechanisms by which these community platforms can be brought into production use:

- Local Deployment: The Community Platform is managed by the local resource centre based on a software release made by a single platform integrator. This is the current mode of operation used in EGI where most centres deploy a single Community Platform.
- Remote Instantiation: The Community Platform is encapsulated in a virtual machine that can be created and instantiated by the platform operator (on behalf of the research community) on the virtualised resources provided by the resource centre.



While there is a responsibility for EGI to ensure that an open ecosystem is established that allows platform integrators, platform operators and research communities to have equal access to these services provided by EGI, the quantity of resources that are allocated to a research community is governed outside of EGI, as is the funding to support a particular research community and the platform operators & integrators, and the technology providers that they may work with.

2.7 Summary

Community and coordination provides an essential framework for the development of a range of non-operational activities that are necessary to provide a balanced production infrastructure that engages openly with the ecosystem that needs to develop around it. Strengthening the openness and transparency of EGI's activities and its governance is essential in building the necessary confidence as a strategic partner with research infrastructures and research communities. Defining and articulating the value that EGI offers and engaging with the consumers of these services will be a major focus over the next two years.



3 OPERATIONAL INFRASTRUCTURE

3.1 Introduction

The Operation Infrastructure and related services can be factorised into three different platforms which have different focuses, target consumers and levels of maturity. These are:

- Core Infrastructure Platform: For resource centres wanting to ensure the services they operate for communities as part of distributed collaborations operate consistently across different resource centres. These services could be deployed or used by research infrastructures to manage the community services being run by their distributed data centres. This platform represents the current operational infrastructure.
- Cloud Infrastructure Platform: Exposing the academic institution private clouds that are beginning to be deployed by many resource centres for consistent use within research collaborations (the same way as private batch clusters are exposed through grid middlewares) is becoming of interest to many research communities. An integration within the current operational infrastructure is currently underway. This collaboration between resource providers, technology providers and research communities will build understanding and help establish best practice which can inform a wider-scale deployment into production. This platform could provide a new capability within EGI that could allow research communities to directly deploy and operate their own services more flexibly than is currently possible. Results so far have been promising but no estimate can yet be made as to when such a capability might enter production.
- Collaborative Infrastructure Platform: These services support collaboration within the other Infrastructure Platforms but are also available for research communities, platform operators and platform integrators and others to use. Some of the services are mature and others are still developing their functionality. Offering these as an integrated platform is a new aspect to this work.

The main work items for these platforms in PY3 and PY4 that relate to the operational infrastructure and its tools are described below. These have been sourced from:

- The workplan approved by the Operations Management Board during the meeting held in Amsterdam in January 2012³ and are described in more detail on the wiki⁴.
- The activity milestones defined by the partners of the Federated Cloud Task Force⁵.
- The operational tool development roadmap described in the MS708⁶.

The operational services herein mentioned and the related ITIL processes are defined in the EGI Operations Architecture: Grid Service Management Best Practices⁷.

3.2 Core Infrastructure Platform

The Core Infrastructure Platform with some of the services in the Collaborative Infrastructure platform represents the current stable operational infrastructure. A non-technical focus in PY3 will be to

³ <https://www.egi.eu/indico/conferenceDisplay.py?confId=618>

⁴ https://wiki.egi.eu/wiki/EGI-inSPIRE_SA1#2012

⁵ <https://wiki.egi.eu/wiki/Fedcloud-tf:FederatedCloudsTaskForce>

⁶ <https://documents.egi.eu/document/962>

⁷ <https://documents.egi.eu/document/763>

identify research infrastructures with a need to monitor their own communities distributed services in their own distributed resource centres and to explore how they can use the services within the Core Infrastructure Platform to operate their own operational infrastructures. Formal uptake of these services by individual research communities will also be explored.

A consistent theme of the Core Infrastructure services and tools for the near future will be consolidation and standards orientation. Many services (e.g. Operations Integration, Accounting) are planning to integrate with the EGI Messaging Broker Network based on Apache MQ, which in turn implements JMS 1.1, AMQP and STOMP⁸. STOMP emerges as the preferred messaging protocol, serving test-oriented rendering of domain-specific payload modelling languages. These, in turn, consolidate around OGF UR and StAR⁹ for accounting and OGF GLUE 2.0¹⁰ for describing EGI's distributed computing infrastructure. In the networking area, activity will be devoted to verify the conformance of the software components to IPv6.

3.2.1 Operational Security

- PQ9-PQ10
 - Service Security Challenge (SSC) 5:
 - Extension of access Monitor Module
 - One NGI run of SSC5
 - Integration of more job-submission methods
 - Improvement of reporting module
 - Security Dashboard
 - Optimization of alerts
 - New version of Pakiti rolled to production
 - Proposal for site-wide security monitoring
 - Software Vulnerability Group
 - EMI Vulnerability Assessment of VOMS Core completed
 - Update of the EMI Vulnerability Assessment plan including status report.
 - Security Threat Risk Assessment (as described in D4.4)
 - CSIRT
 - F2F meeting
 - Improvement in restricted access ticketing system and of the internal issue handling procedure
- PQ11-PQ12
 - SSC6
 - Running SC6
 - Evaluation of SSC6
 - Update of site certification procedure

⁸ JMS 2.0 – <http://jcp.org/en/jsr/detail?id=343>

AMQP – <http://www.amqp.org>

STOMP – <http://stomp.github.com/>

⁹ OGF UR – <http://www.ogf.org/documents/GFD.98.pdf>

OGF StAR – <https://forge.gridforum.org/sf/go/artf6532?nav=1>

¹⁰ OGF GLUE 2.0 – <http://ogf.org/documents/GFD.147.pdf>

- CSIRT
 - Define the Operational procedure for compromised certificates
 - Nagios: CRL checking on services that have gridftp (CEs/SEs) and checking for known vulnerable file permissions via gridftp
 - Security training at the EGI Technical Forum
- Software Vulnerability Group
 - F2F meeting
 - Revise and improve issue handling procedure
 - EMI Vulnerability Assessment of WMS completion
 - EMI Vulnerability of CREAM begins

3.2.2 Operations Integration

- PQ9-PQ10
 - Integration of desktop grid software into EGI operational tools
 - Definition of tests for integration of GLOBUS and UNICORE
 - Finalization of integration of UNICORE tests into SAM
- PQ11-PQ12
 - Completed integration of MAPPER software platform

Additional integration activities will be undertaken according to the demand of EGI's research communities.

3.2.3 Operational Tools

- PQ9-PQ10
 - EGI.eu Central Operations Tools
 - SAM monitoring
 - High Availability configuration implemented for all central instances
 - Automatic notification when a central instance fails (ALARM tickets)
 - Message Broker Network
 - Implementation of a test instance
 - Improvements in scalability
 - Improvements in availability
 - Staged upgrades
 - Operations Portal
 - Decommissioning of old instance
 - Deployment of VO operations dashboard
 - Major upgrade of the regional package
 - Refactoring of the Operations Dashboard
 - First implementation of an Availability /Reliability module
 - GOCDB
 - Virtual sites handling
 - Redevelopment of XML output module
 - GLUE2 compatibility
 - Site Availability/Reliability automated monitoring and follow-up

- Update to the metrics portal from PY2 experience
- PQ11-PQ12
 - Operations Portal: Provide a mobile version
 - Operations Portal: Improved synchronization between various operations tools: SAM, message broker network and the Operations Portal
 - Message Broker Network: Migrate to a secure implementation with authenticated-only connections that will require clients to be: identified and registered, providing each client with credentials and enforcing authenticated access in the client. Authorization rules will be evaluated and enforced¹¹.
 - GOCDB: Re-assessment of requirements for GOCDB regionalization and of its integration with other regionalized tools.

3.2.4 Accounting

- PQ9-PQ10
 - Repository: New Secure StoMP Messenger (SSM) -based infrastructure in production
 - Portal: Implementation of NGI usage reports
 - Portal: Database migration (after other accounting providers are migrated to SSM)
 - Repository: Improved handing of cloud records
 - Repository: Prototype storage records implementation
 - Portal: Contributed CPUs view
 - Portal: Parallel Job accounting
- PQ11-PQ12
 - Portal: Receiving message-based summaries
 - (EMI) EMI APEL client using SSM released
 - Repository: Regional distribution available

3.2.5 Availability and core services

- PQ9-PQ10
 - Assessment of OPS VOMS server high-availability needs
 - Operational Level Agreements
 - Second release of Resource infrastructure Provider OLA
 - Extension of Resource infrastructure Provider OLA to include more NGI grid core services
 - First draft of EGI.eu OLA
 - Availability/Reliability reporting for NGI core services:
 - Discussion of specifications with operations portal team
 - Migration of grid middleware probes from SAM packages to the EMI packages
 - MyEGI: Maintenance and stability improvements of the central monitoring service
 - MyEGI: Improving the existing documentation (adding MyEGI user guide)
 - POEM: Improvements to the first version released in PQ8

¹¹ https://wiki.egi.eu/wiki/Plan_2012_SA1.4#Security_implementation_in_messaging_system.

- Messaging:
 - Check which instances are still using unauthenticated client connections
 - Setup credentials / request clients to migrate to authenticated connections
 - Request to EGI OMB to approve authenticated only connections to PROD message broker network
- PQ11-PQ12
 - First release of EGI.eu OLA
 - Decommissioning of legacy ROC groups in the DTEAM VO
 - Messaging: Enforce authentication to all clients if OMB approves it
 - Messaging: Enforce authorization rules if OMB approves it

3.2.6 Network Support

- PQ9-PQ10
 - Network support:
 - IPv6 testbed in place for site services
 - HINTS troubleshooting tool for ia64
 - TPM: Review and revision of effort, duties, role of TPM
 - Improvement of the procedure and tools for a graceful migration from testing to production, to facilitate the running of new sites into production
 - Testing of Nagios probes to assess under-performing sites
- PQ11-PQ12
 - Solution of the UNKNOWN status problem
 - Change to underperforming sites followup-procedure
 - Networking Support
 - IPv6 testbed support for operational tools
 - Update on perfSONAR MDM developments at the Technical Forum
 - Report on IPv6 testing activities
 - EGI.eu OLA enforcement
 - ROD training and education

3.2.7 Technical Outreach

EGI has a monitoring system that enables a research or other community to deploy generic and service-specific test probes and execute these on services deployed in resource centres for their community. The monitor can collect statistics on service availability, reliability and correctness and it can raise alarms within the EGI Helpdesk in case of a service found erroneous by a probe.

The EGI monitoring system has been traditionally used by resource centres to monitor grid middleware services. A new version of this monitor is expected to be ready before in PQ9 that can target both community-specific services and the currently supported middleware services of EGI. Whereas middleware services need to be tested across a large number of sites in a uniform way, community-specific services need custom probes that are typically used by a small number of sites and users. Once the new monitor is available with documentation from JRA1 the UCST at EGI.eu will use it to support user communities for community-specific services and use cases. This support will integrate effort from NA2 (TNA2.5), SA1 and JRA1, and the following steps are envisaged:

- PQ9: Move the current probes (service availability probes) used by the TNA2.5 services (AppDB, Training Marketplace, CRM) into the user community area of the monitoring service.
- PQ10: Update/replace if necessary the service availability probes of the AppDB, Training Marketplace and CRM services to service-specific probes.
- PQ10: Develop tutorials for NGI user support teams and collaborating projects about the community service monitoring capabilities of the EGI monitoring system. Promote and/or deliver these tutorials to NGI user support teams and collaborating projects.
- PQ11: Support projects that provide community specific software services in EGI to develop and deploy service specific probes for their services using the EGI monitoring system (for example SHIWA workflow repository, SCI-BUS Generic Gateway Technology).
- PQ12: Setup and provide a repository of reusable and customisable community-specific service probes.

3.3 Cloud Infrastructure Platform

The Cloud Infrastructure Platform reuses services from the Core Infrastructure Platform to federate the virtualised resources starting to emerge in individual resource centres into a federated cloud that is available to support European research communities. The definition of this platform is in an exploratory phase that is establishing the capability that can be achieved within EGI-InSPIRE and related European and national activities.

From the beginning, the Cloud Infrastructure Platform was designed to rely on Standards promotion (e.g. SIENA) and development (e.g. OGF, SNIA, OASIS, DMTF). Integrating with the EGI Core Infrastructure (section 3.2) this platform will reuse existing services and standards wherever possible, for example UR and StAR for accounting integration, Cloud management access will be provided through OCCI and CDMI¹² for virtualised compute and storage management; OVF¹³ emerges as a commonly accepted container for VM images and appliances. Access interfaces to stored data consolidates around HTTP based protocols (plain HTTP, WebDAV) mapped into POSIX drivers available for Cloud storage clients. Access control to Cloud Management services will be based on X.509 certificates, but is foreseen to integrate with SAML/XACML based AAI infrastructures.

3.3.1 Management

Management solutions enable resource providers to expose virtualised compute, storage or network resources to user communities without the need to dedicate any physical resources to one particular community or customer. While virtualised compute resources are the most mature class of cloud resources, storage and even more so network resources are lagging behind in maturity and available implementations. The overall goal for the medium term future is enabling Resource Providers to select one solution out of potentially many that fits well with their specific local requirements. To reach that goal, the following intermediate milestones are defined:

- PQ10: Adoption of distributions with native support for OCCI1.1 instead of the current proxy model to access virtualised compute resources within EGI.
- PQ10: Confirm the adoption of CDMI as Cloud storage management interface within EGI.
- PQ12: Assess OVF as VM image and appliance container.
- PQ12: Assess virtualised network resources.

¹² OGF OCCI – <http://ogf.org/documents/GFD.183.pdf> et al.

SNIA CDMI – <http://cdmi.sniacloud.com/>

¹³ DMTF OVF – <http://www.dmtf.org/standards/ovf>

- PQ12: Assess integration issues of commercial Cloud management solutions alongside EGI's resources.

3.3.2 Authentication and Authorisation Infrastructure

A separation is made between the authentication and authorisation needed to access the Cloud Infrastructure Platform (the focus of this section) and the functional services that are run through that platform which are the responsibility of the platform integrator and platform operator to meet the needs of the consuming research community (see section 3.4.9). It is expected that X.509 certificates will remain the primary mechanism for accessing the Core Infrastructure Platform.

- PQ10: Integrate VOMS based authorisation using X.509 based authentication with the cloud management solutions
- PQ12: Provide XACML support through (if needed) a simplified SAML/XACML profile.
- PQ12: Explore use of ARGUS and lightweight ID federations

3.3.3 Accounting

It is expected that the new virtualised resources will be integrated with the current EGI accounting system.

- PQ10: Define the accounting data that needs to be inserted into the EGI Accounting system and prototype accounting clients to extract this data from the adopted cloud management distributions.
- PQ10: Refine the early version of the Cloud Accounting Record (CAR), which is a subset of the existing OGF Usage Record format, and adopt it within the accounting clients.
- PQ14: Provide a CAR experience document to OGF.
- PQ12: Use EMI's Storage Accounting (StAR) submission to OGF as a basis for incorporating storage into the CAR and update the prototype accounting clients.
- PQ12: Define the requirements for network accounting
- PQ14: Provide an implementation plan for network accounting within EGI.
- PQ14: Based on the experience obtained to date of the integration of virtualised resources into EGI record the lessons learned and define requirements for further work.

3.3.4 Information Discovery

The Information Discovery capability will reuse EGI's existing information services to provide information about the virtualised resources and the capabilities offered by a Resource Provider.

- PQ10: Use the GLUE2 model to define the information model in EGI that will be used to describe virtualised resources.
- PQ12: Review the suitability of the LDAP based information service model currently used in EGI within the Federated Cloud testbed.

3.3.5 Monitoring

Monitoring of a site resource is essential in ensuring the availability and reliability of a site to deliver the expected service interface.

- PQ10: Derive availability and reliability definitions for Cloud management solutions
- PQ12: Expand monitoring probes to validate OCCI, CDMI, OVF and other interfaces as required
- PQ14: Integrate probes with the production infrastructure

3.4 Collaborative Infrastructure Platform

The Collaborative Infrastructure Platform is being established as a set of services that can be used by resource providers and optionally by other partners in the EGI ecosystem (such as platform operators, platform integrators, technology provider, research communities and others) to collaborate effectively with each other. An ongoing activity is to develop the business models around these services and around the platform as a whole.

Reflecting the early stages of the Collaborative Infrastructure Platform, the provided services operate in relative isolation. Repository-based services converge towards package or appliance based repositories, providing different access interfaces. Classic package based repositories provide RPM/YUM, or DEB/APT access for package management related activities, while the VM Image Marketplace provides HTTP/WebDAV access to download binary artefacts (VM images, etc.). The VM Image Marketplace provides an RDF/SPARQL query interface, which might get adopted across higher-level repositories such as the Applications Database and Training Marketplace.

3.4.1 Helpdesk

The EGI helpdesk is a mature service that provides an issue tracking system that links resource providers, research communities, technology providers, platform integrators, platform operators together in order to provide a single integrated entry point for all adopting users.

- PQ9-PQ10
 - First prototype of the report generator
 - Discussion on specific workflows to support operations security
 - Implementation of a high-availability configuration supporting auto-switching
 - Report generator in production
 - Widget for integration into VO portals
 - Running of new xGUS instances according to demand
 - Specific work flows for CSIRT/Security (under discussion)
 - Authentication through Shibboleth (deadline to be determined)
- PQ11-PQ12
 - Completed high-availability configuration for the DBMS and for GGUS
 - Completed by an intrusion Prevention System configuration for GGUS
 - Improvements of service monitoring
 - Planning the DBMS migration to a virtual environment
 - First production version of report generator

3.4.2 Software Repository

The EGI Software Repository provides the infrastructure for provisioning software (i.e. Quality Control Verification and Staged Rollout), and the final repository from where the federated EGI Resource Centres are installing product updates into the production infrastructure.

- PQ9
 - Provide repository infrastructure for SL5, SL6 and Debian 6
 - Provide stable repository configurations for provisioning repositories beyond production (“Untested”, “Testing”, “Release Candidate”)
 - Repository download statistics for UMD based on Metapackage downloads.
- PQ10
 - Repository download statistics for other hosted software (e.g. SAM, CA certificates)



- PQ11
 - Expand provisioning infrastructure for additional Technology Providers
- PQ13
 - Simplify the infrastructure and processes in order for it to become a sustainable community service

3.4.3 Software Verification

Software Verification activity developed over PY1 & PY2 is scoped around acceptance testing of software that is delivered to EGI against the defined Quality Criteria. The Quality Criteria Verification team will monitor the releases from the EGI Technology Providers and source the updated products into the Software Provisioning infrastructure using a small Web Application.

- PQ9, PQ11, PQ13 & PQ15: Updated Quality Criteria.
- PQ9: Refine verification process for SL5, SL6 and Debian 6 platforms
- PQ10: Update verification planning and allocation to allow for monthly UMD releases

3.4.4 Staged Rollout

Staged Rollout has been developed over the last two years in response to the use of external technology providers and platform integrators to prepare the software deployed at local resource centres. Verified releases are run in production at early adopting resource centres as a final acceptance test before being released into production.

- PQ9-PQ10
 - Preparation of Early Adopting sites for EMI 2.0 products for SL5, SL6 and Debian
 - Improvement of workflow in case problems are found after UMD release
 - Re-assessment and improvement of overall software provisioning workflow
 - Web application to supply Staged Rollout reports
 - Mechanism to protect Availability / Reliability statistics in case of Early Adopter failures
- PQ11-PQ12
 - OSG-WLCG-EGI cooperation on interoperability testing
 - Implications of the Federated Cloud Task Force on software provisioning, in particular on Staged Rollout

3.4.5 VM Image Marketplace

The VM Image Marketplace (based on the output of the StratusLab project) is a location where individuals and projects participating in the EGI ecosystem can register and advertise the existence of VM images available for other communities to integrate into their own Virtual Research Environment.

- PQ9
 - Register VM images by uploading a signed XML metadata document (using any IGTF approved certificate) describing the Virtual Machine image while maintaining public read access to the Marketplace at all times
- PQ10
 - Support basic image endorsement by multiple endorsers at the same time
- PQ11
 - Explore partial and multiple signing of image metadata to support more complex image endorsement scenarios

- PQ12
 - Identify endorsement scenarios (including semantics and formal metadata token) for security, privacy, etc. related endorsement needs
- PQ13
 - Explore support for OVF
 - Explore merging with Applications Database

3.4.6 Applications Database

The EGI Applications Database (AppDB) stores tailor-made computing applications for scientists and grid application developer tools for software developers. It embraces all scientific fields, from resources that simulate exotic excitation modes in physics, to applications for complex protein sequences analysis. Storing pre-made applications and reusable tools means that scientists and grid application developers can achieve their goals with EGI in a shorter time.

- PQ9: Provide a list of workflow frameworks in the AppDB that can be integrated into the EGI website (see the science gateway section¹⁴ as an example)
- PQ9: Provide a list of desktop grid frameworks in the AppDB that can be integrated into the EGI website (see the science gateway section¹⁵ as an example)
- PQ9: Use the recently developed features, such as notification and “broken link detector” to inform users about erroneous entries to improve the quality of the AppDB content.
- PQ9: Review the current AppDB features to facilitate “EGI software interest groups” around the EGI-related software entries. These interest groups would comprise developers, supporters and users software and be used to facilitate communication and information exchange between the members.
- PQ9: Start sending regular quarterly updates about the AppDB system (“AppDB newsletters”) to registered users.
- PQ9: Deliver a webinar presentation about AppDB to the EGI community, particularly for NILs, NGI, VRC and VO user support teams.
- PQ10: Extend AppDB with a write API to enable its integration with third party science gateways and repositories.
- PQ11: Create a new or extend the current AppDB Gadget with write-enable capabilities.
- PQ13: Investigate the possible integration of the AppDB into the EGI Collaboration Platform as (or part of) a VM Image Marketplace and the possibility of extending AppDB with code repository function for hosting source-codes of registered applications.

3.4.7 Training Marketplace

The Training Marketplace is a tool to view and advertise training events, resources and requirements. It is aimed at supporting cooperation between trainers and users in different localities and projects by connecting the groups through the activities that are established within the NGIs and scientific clusters. A stable infrastructure and services has reduced NGI based training activities in recent years, however this situation is expected to change as the need for training services and resources is expected to increase because of the transition of the infrastructure to the platform based model and because of

¹⁴ <http://go.egi.eu/sciencegateways>

¹⁵ <http://go.egi.eu/sciencegateways>

the introduction of new services for infrastructure operators, developers and research communities. The Training Marketplace team under the coordination of EGI.eu UCST will prepare for this new situation by extending the current system with:

- PQ9: Adding support for tags that would enable the creation of training topics from the registered resources.
- PQ9: Improving some of the existing services in order to provide more coherent and intuitive interfaces across the various sections of the marketplace.
- PQ9: Collaborating with PRACE to create a roadmap for the creation and delivery of integrated training services for European structured scientific communities.
- Increasing the use of the marketplace by:
 - PQ9: Delivering a webinar presentation about the service to the EGI community, particularly for NILs, NGI, VRC and VO user support teams.)
 - PQ9-PQ11: Collaborate with members of the EGI Federated Cloud Task Force, and users of the EGI Federated Cloud testbed on the development and registration of new training materials.
 - PQ12: Investigate possible uptake of the Training Marketplace within commercial user communities through the UK NGI.
- PQ10: Developing new gadgets that enable the customisation of the training marketplace to national websites.
- PQ12: Providing a read-write API for the marketplace to enable the federation of the marketplace with third party registries (such as Indico installations) and enable the development of custom web gadgets.

3.4.8 Client Relationship Management (CRM) System

The EGI CRM supports EGI's outreach to new communities by providing a database with a web interface where potential user communities, personal leads to those communities and key needs of those communities identified through the leads can be recorded.

- PQ9: Survey to get feedback from NIL and other early users
- PQ10-12: Monitor update and usage and respond to recorded feedback.

3.4.9 Federated Authentication and Authorisation Infrastructure

Currently, most EGI services use X.509 certificates for user authentication, authorisation and single sign-on purposes. Short term proxy certificates generated from long term certificates are used as "tokens" by applications, jobs and files to request access to EGI services. Through requirements collected from several communities it is evident that many research communities that are potential adopters of EGI services look at certificate based access as a barrier of uptake of EGI. Technical solutions that exploit Federated Identity models exist (e.g. Terena certificate server, some science gateways) but are not widespread. Federated Identity is the most attractive solution to scientific communities as it provides institution based identity management model that can be used across multiple service providers within a distributed virtual research environment:

- PQ9: Completion of the Federated Identity Virtual Team¹⁶ established to understand the state of federated identity providers within the EGI community. The report is expected to clarify

¹⁶ https://wiki.egi.eu/wiki/VT_Federated_Identity_Providers_Assessment

next steps towards a wider and more harmonised adoption of federated identity management solutions within the NGIs and their clients. An important first step will be to and share those science gateway enabling technologies that provide support for the mapping of federated user identities to grid certificates accepted by EGI sites (typically to robot or other special certificates).

- PQ9+: Next in a series of “Federated Identity systems for scientific collaborations” workshops¹⁷ that brings together representatives of various scientific groups and technology providers with the goal to explore the requirements for federated identity management across the different disciplines, compare the functionality, operational constraints and state of deployment of current technologies, and formulate a roadmap for how such a service could be established in the future.¹⁸
- PQ10: Assess the activities that EGI would need to undertake to support a Federated AAI model as a service to its consuming communities.

3.4.10 Data Movement as a Service

Currently, data movement services, that allow requested data transfers to be prioritised and managed, are deployed at key resource centres within the production infrastructure. Developing and packaging software for general deployment may become a cost that is unsustainable for the community. An alternative model that is now technically feasible is to provision such a capability centrally as a service.

- PQ10: Establish a trial European entry point for Globus Online¹⁹ through collaboration with the IGE project.
- PQ12: Evaluate uptake of this trial service to decide if this should be offered as an EGI Collaborative Infrastructure Service

3.5 Summary

During PY3 the first EGI security risk assessment will be completed and the tools supporting operations security will be enhanced (security dashboard, RTIR, GGUS and Pakiti). As gLite 3.2 reaches its end of life at the end of April 2012, a major infrastructure upgrade will be undertaken and various obsolete services will be decommissioned. GLUE 2.0 information will be validated and a transition plan will be defined.

The EGI Operations Level Agreement (OLA) framework including three documents: the Resource Centre OLA, the Resource infrastructure Provider OLA and the EGI.eu Central Services OLA will be completed and enforced. The reporting of service level targets for NGI services and EGI.eu will require many extensions in the current operational tools, whose implementation is part of the JRA1 plan.

NGI Core Services will be monitored and consolidated through the deployment of high-availability and/or load balanced configurations. The deployment of GLOBUS and UNICORE will be consolidated in a few NGIs. The accounting infrastructure will be migrated to SSM and IPv6 readiness tests will be conducted in collaboration with JRA1 and WLCG.

¹⁷ <https://indico.cern.ch/conferenceDisplay.py?confId=129364>,
<https://indico.cern.ch/conferenceDisplay.py?confId=157486>,
<http://indico.cern.ch/conferenceDisplay.py?confId=177418>

¹⁸ <https://cdsweb.cern.ch/record/1442597>

¹⁹ <https://www.globusonline.org/>



Many services and tools that are part of the Operational Infrastructure are beginning to consolidate towards a succinct set of publically available set of standards based interfaces. A complete overview of Standards in use in the Operational Infrastructure (and the Virtual Research Environments) is maintained as a publically available live reference at <https://wiki.egi.eu/wiki/Standards> while the overall EGI interoperability context has been defined in the last iteration of the Standards Roadmap deliverable²⁰.

²⁰ <https://documents.egi.eu/document/721>



4 VIRTUAL RESEARCH ENVIRONMENTS

4.1 Introduction

Virtual research environments (VREs) represent the distributed services used by a research community to undertake their data analysis and computational modelling within EGI. Some of these services may be run centrally by the research community or their partners, centrally by the production infrastructure or their partners (as part of the Collaborative Infrastructure Platform), or deployed on individual resource centres (as a specific Community Platform).

Integrating and operating a VRE for a particular community is a skilled task. It can be undertaken by the infrastructure on behalf of the whole infrastructure, but experience has shown that the need to balance conflicting priorities coming from different research communities may mean delivering a solution that fully satisfies none of them. Alternatively, expecting each research communities to have these specialised skills ‘in-house’ is unlikely, but these could be developed over time. A practical approach is for experts within the DCI community is to offer their platform integration and platform operation skills as a service to research communities.

A strong and vibrant set of generic and community specific technology providers, built on the open source software model, is central to platform integrators assembling VREs to meet the needs of individual research communities. The majority of the technology components used the Community Platform that EGI assembles for use in the production infrastructure, the Unified Middleware Distribution (UMD), is currently drawn from the EMI and IGE projects. The platform integration activities provided by these projects in addition to the community specific technology that they provide means the EGI can adopt a lightweight integration and verification activity on top of their work (see the Collaborative Infrastructure Platform services such as Software Repository, Software Verification and Staged Rollout provided by SA2²¹). With the end of the EMI and IGE projects at the end of PY3, EGI may have to find additional effort to integrate the components that are relevant to its use of this is not sustained within the respective EMI and IGE projects.

Through SA3, EGI-InSPIRE supports generic and community specific technologies and services that contribute to the VREs used by different research communities²². The detailed technical plans for these components in PY3 are available²³. In addition to the technology coming from within EGI-InSPIRE there are other Community Platforms that could be used as the basis of a community’s VRE. If EGI’s Core Infrastructure Platform is deployed, research communities could use these resources to instantiate the VENUS-C platform (or any other virtual machine appliance) as part of their VRE or use a subset of the EMI or IGE technologies customised for their own particular needs.

²¹ <https://documents.egi.eu/document/1015>

²² <https://documents.egi.eu/document/742>

²³ <https://documents.egi.eu/document/746>



4.2 Generic Technologies and Services

4.2.1 Job Management Frameworks

4.2.1.1 GANGA

Ganga is an end-user tool for creating and managing computational tasks. Its mature Python codebase provides for a stable yet extensible framework, which is consistently used by over 400 unique end-users each month. It remains a popular environment for running grid analysis jobs within the LHCb and ATLAS experiments, and is known to be used by a further 10 user communities across a range of scientific disciplines.

Developments for PQ9-PQ12:

- Facilitate the adoption of Ganga as an analysis platform through its adoption into the mainstream Linux community through EPEL.
- Review and minimise the external software dependencies.

After PQ12 when this software is no longer part of the EGI-InSPIRE project, the software will rely on effort from its multiple user communities to provide development and support effort. This is promoted by its modular architecture which reduces the cost of maintaining the core framework while allows development to be shared by different developers at different institutes through the open community.

4.2.2 Dashboards

The Experiment Dashboard system provides monitoring of the WLCG infrastructure from the perspective of the LHC experiments and covers the full range of their computing activities, such as data transfer, job processing, and site commissioning. In contrast to many other monitoring systems, the Experiment Dashboard is not coupled to any particular middleware, workload management, or data management systems. It is shared by several LHC virtual organizations and works transparently across various middleware platforms.

PQ9-12:

- Refinement of the Global WLCG transfer monitor in response to user community feedback.
- A new version of the task monitoring application based on the hBrowser framework.
- All Dashboard data repositories will be migrated to new hardware and ORACLE version 11g to improve performance.
- The functionality of the Site Status Board application will be extended in order to provide better automation of the operations of the distributed infrastructure and improve sustainability.
- Maintenance and support of the Dashboard cluster, services, and software.

After PQ12 when this software is no longer part of the EGI-InSPIRE project it will be supported with a core development and support team at CERN, and collaborating teams from the LHC experiments. There are also contributions from the institutes participating in the LHC project from Russia, Taiwan, India, and the UK. The reengineering supported by EGI-InSPIRE has left a modular framework that is well suited to this collaboration model.

4.2.3 Science gateways

An EGI science gateway is a 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 the European Grid Infrastructure. These gateways can support a variety of capabilities including workflows, virtualisation software and hardware, visualization as well as resource



discovery, job execution services, access to data collections, applications, and tools for data analysis. A science gateway enables community members to define and perform custom research scenarios.

The EGI.eu User and Community Support Team, through its partners in the NGIs and various projects, will facilitate the development and integration of community specific gateways, especially web based science gateways through a number of mechanisms:

- **Setup a sub-community within EGI:** Kept together by an email list²⁴ it can be used to connect and coordinate the effort of portal developers, operators, and technology providers in a particular domain.
- **Setup specific sections on the EGI website:** These can be used to promote EGI science gateways (for end users) and gateway enabling technologies (for developers)²⁵. The site lists gateways and technologies from the EGI Applications Database through tagging and web gadgets. The mechanism is flexible, scalable and open, providing a ‘marketplace’ gateway providers and potential users. The same or similar lists can be embedded into any user support website through the gadgets.
- **Extended the EGI Applications Database application with ‘Try it’ links:** These can be used to connect ready-to-use web based gateways to AppDB entries which simplifies provisioning and access for new users.
- **Links to Technology:** Match make available gateway technologies and user support (from NGIs, communities, projects e.g. SCI-BUS) with the needs of VRCs and projects (e.g. DRIHM and ESFRI projects).
- **International Collaboration:** Initiate collaboration with other international initiatives, especially XSEDE on joining efforts in exchanging and promoting science gateway technologies, in further developing these technologies.

As next steps the team will:

- PQ9: Promote the new science gateway sections of the website to those who could benefit from it:
 - Providers and operators of gateways
 - Potential user communities, who needs simple interfaces to interact with EGI services
- PQ9-PQ10: Prepare support material, a gateway primer, to gateway developers. The primer would bring together recommendations and best practices about the tools, methods, and policies that one needs to use to create custom gateways. In order to pull effort from the NGIs and projects, the primer will be developed by a new Virtual Team.
- PQ9-PQ14: Assess the feasibility of extending the science gateway concept to mobile platforms by providing mobile applications to the EGI user support services: AppDB, Training Marketplace, CRM, Requirements Tracker.
- In collaboration with the SCI-BUS project²⁶:
 - PQ10 & PQ14 Organise a joint event on science gateways and enabling technologies at the EGI Technical Forums.
 - PQ9-10: Assess the feasibility of using the portlet repository that SCI-BUS uses to exchange reusable portlets within the EGI community.

²⁴ portal-community@mailman.egi.eu - <https://www.egi.eu/sso/groupView/portal-community>

²⁵ <http://go.egi.eu/sciencegateways>

²⁶ EGI-InSPIRE – SCI-BUS Memorandum of Understanding: <https://documents.egi.eu/document/1009>



SOMA2 is a versatile modelling environment for computational drug discovery and molecular modelling supported through EGI-InSPIRE SA3.

Plans for PQ9-12 include:

- Provide a pilot service for international user communities
- Advertise the service to different user communities.
- Expand the selection of scientific applications from different virtual organizations.
- Continue assessing feature requests from the user community, bug fixes, and other enhancements.

While CSC currently plans to continue to develop and maintain SOMA2 for national use, the open modular nature of the source code and the engagement of new research communities through the EGI-InSPIRE based pilot service, will help grow an international developer community that can contribute to its long-term sustainability.

4.2.4 Workflows

Workflow systems enable scientific communities to compose and execute a series of computational or data manipulation steps, or a workflow, in a scientific application on resources of the European Grid Infrastructure. Workflows facilitate the reuse of software components, of complete applications and increase the repeatability of scientific calculations. Several NGIs, projects, communities within EGI develop workflow systems and workflows. The EGI.eu User and Community Support Team with its partners within NGIs, VRCs and projects aims to facilitate the exchange of workflow systems, workflows, development best practices, and further developments by:

- PQ10: Setup a dedicated section on the EGI website about EGI workflow systems, similarly how the science gateways section is implemented (using tags and web gadgets). Besides workflow systems the page will also list workflows that are available for these systems in the EGI AppDB and in external repositories (e.g. MyExperiment, SHIWA repository).
- As a continuous activity broker and matchmake available workflow systems and support from NGI and community-specific teams and projects (e.g. SHIWA) and between VRCs, projects that need such services (e.g. DRIHM and ESFRI projects).
- PQ10 & PQ14: Organise community events relating to workflow systems in 2012 and 2013 at the Technical Forum.

4.3 Community Platforms

4.3.1 UMD

Within a year both of the projects that support EGI's main Technology Providers will come to an end. Currently, the individual sustainability strategies of these projects are not clear. It appears that the software currently provided through EMI may be continued by some of the current product teams but only at a level sufficient to provide minimal maintenance or to support developments for community that they have specific funding for. Some of the NGIs within EGI have recognised the importance of some of the software components and may be willing to contribute to their sustainability. For software delivered by IGE, the most likely candidate for future software maintenance is currently the Globus Alliance, but it is not clear if support will be provided for components that were developed specifically by IGE. At the moment there is no reason to assume fundamental changes current approach to how software delivery and provisioning is executed.

Detailed planning for the upcoming UMD releases is based on the currently available plans from the contributing Technology Providers:

- PQ9: Release UMD-1 updates including backwards-compatible components that are included in EMI-2.
 - Contents of EMI-1 plus the following new components: Hydra, WNoDeS, EMI Registry (EMIR), Common Authentication Library (CANL), Pseudonymity, and EDGI bridges for all EMI CEs (CREAM, ARC CE, UNICORE/X).²⁷
 - Hydra and WMS are candidates for later EMI-2 updates; consequently they may appear in UMD at a later stage.
 - Supported OS platforms: SL5, SL6 and Debian6 (UI clients)
 - Unified LRMS support for all EMI CEs including LFS, GridEngine, PBS & PBS-Pro.
 - GLUE2 will be mostly supported across all EMI products with the following initial exceptions: EMI (common) data library, WMS, gLite JDL, gLite clients, ARC Compute clients
 - Complete support for EGI's monitoring infrastructure through Nagios plugins
 - Include IGE-2.1 which extends IGE 2.0 with:
 - GridSAFE, an accounting framework supporting UR 1.0,
 - OGSA-DAI REST for distributed data access and integration,
 - An LCAS/LCMAPS update for Globus, introducing SAML2/XACML2 support (for integration with ARGUS and/or GUMS)
 - A GridWay update supporting CREAM and any BES compliant Job Execution service
 - Enhanced security supporting RFC6277 and RFC5280 (IETF Proxy Certificate profile)
- PQ10: Release UMD-1 update including backwards compatible or new components:²⁸ CEMon (v1.14.0), CREAM (v2.0.0), and DGAS sensors (v4.1.0) which are expected to be provided in later EMI-2 updates, among other product updates.
- PQ10-PQ11: Follow monthly EMI-2 update schedules. As the contents of each update are known only a couple of weeks ahead UMD update releases will be made as required by the user and operations communities.
- PQ10-PQ12: Include IGE-3 in UMD-1 with:
 - GridSafe providing full integration with EGI Accounting infrastructure
 - GRAM5 providing core support for BES
 - LCAS/LCMAPS supporting low-level extraction of accounting records
 - GridWay supporting DRAMAA v2
- PQ12: Include EMI-3 in UMD. No complete details of EMI-3 are currently (April 2012) available.

²⁷ The currently planned EMI-2 contents is a *moving target*, i.e. components may be added or removed at any time: http://go.egi.eu/EMI-2_components

²⁸ The initial release of EMI-2 is a subset of all components of a general EMI-2, and is tracked at <https://savannah.cern.ch/task/?28140>.

4.3.2 High Energy Physics (HEP)

The strategy adopted by HEP since many years is for collaborative development by a number of partners. This has been proven to provide a satisfactory level of isolation from changes in funding and / or resourcing from individual partners in each collaboration: the effort from each typically varies slowly over time due to evolving priorities of the partner involved.

4.3.2.1 HammerCloud

HammerCloud is a Grid site testing service developed around Ganga. HammerCloud uses frequent short jobs to validate a site's availability and functionality, and also delivers on-demand stress tests to aid in site commissioning or general benchmarking. Ganga was developed with the ATLAS LHC experiment and that collaboration continues to use the service actively. The CMS and LHCb LHC experiments also have HammerCloud plug-ins developed.

During PQ9-12:

- Migration of the ATLAS instance to the new version, merging development with the CMS and LHCb instance.
- Development of features for new database backends based on NoSQL engines (such as CouchDB) for the use case of high demand storage (mainly metrics and job results).
- Extension of the auto-exclusion feature that has been tested in production with ATLAS and proved effective, to the CMS and LHCb experiments.
- Decommission of Job Robot for CMS, switching to HammerCloud as the main Grid testing service for CMS along with ATLAS.
- Optimization and generalization of submission mechanisms to reduce load.
- Deeper integration with LHCb DIRAC and further deployment for LHCb Grid testing.
- Integration of LHCb testing with the LHCbDIRAC resource status system.
- Working on a standard benchmark (or benchmarking suite) to evaluate the storage elements (e.g. standard set of ROOT test jobs and datasets).
- Investigation of error and performance correlations across VOs (at multi-VO sites).
- Development of a multi-VO testing interface for users to easily test all experiment workflows.
- HC installation simplification and packaging for other-VO usage.

After PQ12 when this software is partially supported by the EGI-InSPIRE project, the software will continue to rely on effort from the remaining partners to provide development and support effort.

4.3.2.2 CRAB and CRAB Analysis Server

The CMS Remote Analysis Builder (CRAB) was the first analysis tool in CMS to aid users in configuring CMS applications for distributed use, by discovering the location of remote datasets and submitting jobs to the Grid infrastructure. CRAB has progressed from a limited initial prototype nearly 5 years ago to a fully validated system that was critical to the production of published physics results.

During PQ9-PQ12 the development plan includes::

- To rewrite all the monitoring implementation, building centralized service, based on CouchDB, able to collect monitoring documents from the distributed agents.
- To implement advanced functionality like user data publication, support for the user output merging.
- To evolve the current client allowing its usage as a python library. The aim is to enable the possibility to be used by external bot/services like HammerCloud, which will be the first case we will support.
- To rewrite the RESTful based APIs using a better modular approach, which means a set of common libraries among various Web Services of the CMS Workload and Data management. The new APIs are going to be rewritten mostly to address the security aspects.



- To evaluate and eventually integrate and use frontier on worker node to cache UserFileCache
- To perform a long scale test, including the whole stack which compose the CRAB3 system, namely: Client, REST Interface, UserFileCache, WorkQueue and distributed Agents.

After PQ12 when this software is partially supported by the EGI-InSPIRE project, the software will continue to rely on effort from the remaining partners to provide development and support effort.

4.3.2.3 ATLAS Distributed Data Management

The ATLAS Distributed Data Management (DDM) is the project built on top of the WLCG middleware and is responsible for the replication, access, and bookkeeping of the multi-Petabyte ATLAS data across the grid, while enforcing the policies defined in the ATLAS Computing Model. The current ATLAS DDM software is now in a mature state and the present work is focused on maintenance and support operations.

During PQ9-PQ12:

- Ongoing maintenance and support operations.
- Architecture discussions to solve current shortcomings and scalability issues.

After PQ12 when this software is partially supported by the EGI-InSPIRE project, the software will continue to rely on effort from the remaining partners to provide development and support effort.

4.3.2.4 Persistency Framework

Persistency and detector conditions cover two essential areas for HEP data processing and analysis, namely the handling of the event data and detector conditions data of the LHC experiments.. In three of the LHC experiments (ATLAS, CMS and LHCb), some or all of these types of data are stored and accessed inside data processing jobs using one or more of the three software packages developed by the Persistency Framework project: CORAL, COOL and POOL.

PQ9-12:

- Ongoing user support, service operation and software maintenance
- Capability of the CORAL software to restore the database connection and session after a network glitch.
- R&D to prototype an improved monitoring infrastructure for CORAL
- Development of CoralServer and CoralServerProxy components in the context of their usage for the ATLAS HLT system.

After PQ12 when this software is no longer part of the EGI-InSPIRE project, the different software packages will follow different support models:

POOL: A sustainable support model has been agreed by phasing out its use. LHCb has already stopped using POOL, after replacing it by a new software layer, inspired from POOL but maintained internally, which is also able to read existing data stored in POOL format; as a consequence, LHCb no longer needs support from POOL through EGI-InSPIRE and its software has no pending dependencies. ATLAS will continue to use POOL for as long as the 2012 production version of the ATLAS software is actively used, but it will no longer use it or need POOL support through EGI-InSPIRE for the 2013 production version, where a custom software package derived from POOL will be built and maintained by ATLAS as part of their internal software.

CORAL & COOL: A sustainable future support model is still being discussed with all its users (ATLAS and LHCb for both, CMS only for CORAL) and other relevant stakeholders. An important step towards sustainability was achieved in 2011 by transferring the full responsibility for CORAL and COOL release build and validation, from the EGI-InSPIRE funded IT team to the team in the CERN Physics department that takes care for software infrastructure issues common to all LHC experiments.



4.3.2.5 DIRAC

The DIRAC system provides a complete solution for using the distributed computing resources of the LHCb experiment. DIRAC has a complete framework for data production and analysis; including workload management, data management, monitoring, and accounting.

PQ9-12:

- Consolidation of the service for data consistency checks between file catalogues and SEs, both in the development and operational aspects.
- Support and possible improvement of the system for the accounting of storage resources usage developed and put in production during the past year.
- Adapt when necessary the DIRAC DMS to change in the data management middleware (FTS, SRM, LFC etc.).
- Continue the development of the popularity service to provide metrics to assess the data-sets popularity and provide a ranking of the most popular data-sets (i.e. data most frequently accessed by users).
- Provide general support for LHCb computing operations on the grid, both for production and private users' activity.

In advance of PQ12 when this software is no longer part of the EGI-InSPIRE project, the LHCb DIRAC project has improved its sustainability model for developing software, certifying and making releases. By making the code more modular and providing a more systematic process for testing from developers, the time required for certification and for producing new releases has reduced.

4.3.3 Life Sciences (LS)

To coordinate their efforts and sustain their activity, members from the Life Science community self-organized into the project independent "Life Sciences Grid Community" (LSGC).

- PQ9 Deploy a redundant VOMS server.
- PQ10: Deploy a Hydra keystore server.
- PQ9-12: Development of a HUC users database and management tools to assist VO administrators that will interface to VOMS, the EGI Applications Database and the LSGC Dashboard.

After PQ12 when this activity is no longer part of the EGI-InSPIRE project, the LSGC will rely on the support of the contributing NGIs that has been developed over the three years of EGI-InSPIRE.

4.3.4 Astronomy & Astrophysics (A&A)

The coordination of activities within the A&A community and their use of DCIs for both small-scale and large-scale projects, in particular the ESFRI ones (i.e. SKA and CTA) remains the primary concern.

- PQ9-12: The implementation of a VisIVO web portal for gLite
- PQ9-12: The production of a MPI-compatible version of VisIVO for gLite
- PQ9-12: The production of a CUDA-compatible version of VisIVO for gLite.

After PQ12 when this activity is no longer part of the EGI-InSPIRE project coordination will take place where necessary with the individual activities rather than any community coordinating function.



4.3.5 MPI

The Message Passing Interface (MPI) standard provides a widely adopted model for building scientific applications that can be easily ported to many different computing platforms: Recurring activities by the MPI team during PQ9-12 include:

- Improved end-user documentation, addressing MPI application development and job submission in ARC, gLite, and UNICORE,
- Quality controlled MPI site deployment documentation,
- Outreach and dissemination at major EGI events and workshops,
- User community, NGI and site engagement, gathering direct input,

Specific work items during PQ9-12 include:

- End-user support for both MVAPICH (an MPI implementation with advanced Infinband networking) and generic parallel job support.
- Investigating additions to the Information System to support GPGPU clusters.
- Upgrade the predominant job scheduler “MAUI” due to a node allocation bug once /problem has been discovered in the standard distribution.

After PQ12 when this activity is no longer part of the EGI-InSPIRE project support will be absorbed into general SA1 activities.

4.4 Summary

Virtual Research Environments are a recognition that for research communities to engage effectively in the use of distributed computing infrastructures for their analysis activities they need to be able to customise these environments for their own needs. With the current deployment model where EGI physically deploys a single Community Platform (UMD) for all its supported research communities, these communities have had to develop their own job management frameworks, dashboards, science gateways and workflows to build the VRE that is tuned to the needs. Some communities have established their own community activities (sometimes specific services and/or the monitoring and operation of their resources) to ensure an effective environment for their researchers. This has been established within EGI-InSPIRE for high energy physics, life sciences, astronomy & astrophysics and a generic technology platform (MPI) that is needed across many research communities. The inclusion of desktop grid resources that provides a generic technology platform for some workloads is already being integrated into the production infrastructure.



5 CONCLUSIONS

This roadmap brings together the technical plans that EGI-InSPIRE has established for PY3 and PY4 (where available), the activities taking place around these technical plans for engagement with new user communities, communication and marketing activities, strategy and policy development and the building of links within the NGIs and in other collaborating projects. Comprehensive detail is provided in supporting technical documentation and these plans remain under regular review.

These technical plans have been structured around:

- **Core Infrastructure Platform:** The management and uniform delivery of services hosted in different administrative domains to different European research communities.
- **Cloud Infrastructure Platform:** The prototyping integration of interfaces to virtualised resources using the Core Infrastructure Platform to provide European research communities with a federated cloud resource built upon resources in the public sector.
- **Collaborative Infrastructure Platform:** A set of services, some of which have been developed over the last decade and some of which will be delivered through new partnerships, that may be used internally by the Infrastructure Platforms but are also valuable for research communities to use within their Virtual Research Environments.
- **Virtual Research Environments:** The set of distributed software services, applications and tools that research communities use to undertake their daily science. Most of the software used within Virtual Research Environments is developed and managed outside of EGI-InSPIRE through collaborations managed through EGI or directly by the individual research communities.

Drawn together, these summaries of the individual technical plans provides an overview of the evolution of EGI's activities during PY3 and provide opportunities for improved interaction between activities taking place within the project and with other activities in the EGI ecosystem.