EGI-InSPIRE

EGI PLATFORMS ROADMAP

EU MILESTONE: MS518

Document identifier:	EGI-InSPIRE-MS518-final.docx
Date:	**13/06/2014**
Activity:	**SA2**
Lead Partner:	**EGI.eu**
Document Status:	**FINAL**
Dissemination Level:	**PUBLIC**
Document Link:	https://documents.egi.eu/document/2232

Abstract

This document introduces the platform-driven service delivery model to the EGI community. It defines the term platform and how IT platforms fit into the current and emerging EGI ecosystem.

After providing an overview of the EGI platform architecture, the first part of the document describes the different platforms in more detail. The second part of the document provides a roadmap on the adoption of the platform-based architecture and the service delivery model used by the European Grid Infrastructure.

I. COPYRIGHT NOTICE

II. DELIVERY SLIP

	Name	Partner/Activity	Date
From	Michel Drescher	EGI.eu	01-06-2014
Reviewed by	**AMB& PMB**	EGI.eu	06-06-2014
Approved by	**AMB & PMB**		13-06-2014

III. DOCUMENT LOG

Issue	Date	Comment	Author/Partner
1	18 Feb 2014	Initial version.	Michel Drescher, EGI.eu
2	02 Apr 2014	Intermediate version	Michel Drescher, EGI.eu
3	30 Apr 2014	Intermediate version	Michel Drescher, EGI.eu
4	02 Jun 2014	Intermediate version	Michel Drescher, EGI.eu
5	05 Jun 3014	Final draft	Michel Drescher, EGI.eu

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/. Additional definitions of terms may be found in the ITIL 2011 Glossary [R 12] and the EGI Technology Glossary [R 13].

The following table provides a set of terms that are used in this document.

Term	Description
EGI Platform model	The EGI Platform model refers to business models that may emerge by utilising any of the IT platforms that are described in the EGI Platform architecture.
EGI Platform architecture	Describes how the individual platforms (see below) are embedded in the EGI ecosystem, and how they are technically integrated with the current EGI production infrastructure.
EGI Platform	In EGI, a Platform is an IT service comprising of technical and human services of value, enabling its users to integrate some or all of its components as they see fit into their own infrastructure for consummation, or as underpinning supply for higher-level services offered to other communities.
EGI Core Infrastructure Platform	The EGI Core Infrastructure Platform consists of services that are necessary to establish and operate a federated, distributed computing and data infrastructure (DCDI). This platform does *not* include services that provide access to computing or data resources. In EGI, any other platform that is deployed and operated by EGI *must* integrate with the EGI Core Infrastructure Platform.
EGI Cloud Infrastructure Platform	The EGI Cloud Infrastructure Platform provides a federation of IaaS Clouds deployed in the EGI ecosystem. Integrated with the EGI Core Infrastructure Platform, it provides consistent and standardised access to virtualised compute, storage and networking resources.
EGI Collaboration Platform	The EGI Collaboration Platform provides IT Infrastructure and Services that facilitate collaboration between Research Communities without being a core infrastructure service for Research Communities.
Community Platform	Community Platforms (there may be more than one) consist of services that are specific to or preferred by the respective community. Community Platforms may be deployed on top of the EGI Core Infrastructure Platform, or on top of the EGI Cloud Infrastructure Platform.

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 document provides an update of the second EGI Platform Roadmap [R 4], by updating the roadmap to document changes lead by the use cases collected by prospective user communities, which mainly lead to the introduction of the Cloud Infrastructure Platform and to the evolution of the Core Infrastructure Platform.

The EGI Platform Architecture has matured into its current form as described in the first part of this document. It comprises the *EGI Core Infrastructure Platform*, which is generic and scoped around the mechanisms and tools needed for federating any type of Platforms, the new *EGI Cloud Infrastructure Platform*, which complements it and focuses on Cloud Computing as a means to provide federated consistent access to e-Infrastructures across Europe, and *Community Platforms*. The services provided within the Unified Middleware Distribution (UMD) are examples of Community Platforms that are integrated with the EGI Core Infrastructure Platform and deployed in the EGI production infrastructure, but owned and maintained by the respective communities themselves.

Components of the cloud infrastructure platform are being implemented thanks to the support of task SA2.6 and to several mini-projects in activity SA4.The Cloud Infrastructure Platform has been evolving thanks to the drive of many user community proofs of concept.

What was considered a concept before is now swiftly becoming an agreed architecture of the EGI production infrastructure, both partly driven *and* validated by the EGI Cloud Infrastructure Platform.

With existing and new services finding their place in this architecture the composition of each EGI Platform allows for formalisation and persistent documentation of the IT landscape in EGI. The first steps have been made in this document, by providing the general architecture and integration between components across platforms. The next steps will have to be further decomposing the architecture into finer details.

Taking the platform architecture's abstraction level concept, EGI will utilise the Technology Collaboration Board and the newly formed UMD Release Team to drive the deployment of ICT service that implement the Platform roadmap documented in this deliverable. The EGI Platforms Roadmap will *also* focus on further development of the IT services based on the technical components described in this document.

With the increasing maturity of the technical platforms, the stakeholders in the EGI community become better defined, and the roles they may play in different deployment and integration scenarios become clearer than before. This is a classic example of an evolution of feedback based circular platforms, where the evolution of a community is influenced by the evolution of the used computing platform and vice versa.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

1   INTRODUCTION

The EGI Strategy [R 1] outlines how EGI is gradually yet continuously aligning itself with the EC's vision of H2020. Consequently the EGI Strategy is built around implementing three pillars of services to the European Research Area (ERA); (1) Provide a pan-European e-Infrastructure, (2) Provide Community and Coordination services, and (3) Foster the development and support for Virtual Research Environments in Europe. Within this strategy, the technical infrastructure forms the first pillar of EGI's support for the researchers and research communities that make up the ERA. Part of this overarching goal is to define the scope for the EGI platforms and mature their respective technical architecture into a model that allows access to and integration with these platforms as a marketable asset to research communities within the EGI ecosystem.

The EGI Platform Roadmap document, in its updated editions (i.e. this most recent version, and [R 3] and [R 4]), describes the technical architecture of the EGI production infrastructure; still on an overview level it provides information about the key building blocks of the architecture, and how the various platforms EGI is offering are composed.

The support of the long tail of science as well as of large Research Infrastructures, is driving the evolution of the EGI Platforms, and the collection of multiple use cases[1] from research communities has driven in the past project year the extension of the platform roadmap to include a Cloud Infrastructure Platform, whose production phase will be launched in May 2014. This third edition of the Platform Roadmap reflects the changes introduced in the architecture to meet the requirements of new user communities.

This document is complemented by a number of other materials so that all these documents together provide the reader with a comprehensive description of the technical aspects of delivering the EGI strategy as follows. The EGI Strategy document [R 1] describes EGI's future direction, and the activity areas that will help delivering this strategy, in a very high-level overview. This document provides the most overview, and the least detail, across the entirety of EGI. The EGI Technical Roadmap [R 2] (and the soon to be published newer edition) provide the most detail on planned technical improvements supporting the *entire* EGI strategy – that explicitly includes technical activities helping delivering both community support services, and support for Virtual Research Environment development.

This document is organised as follows.

Section 2 provides and overview of the stakeholders in the EGI Platform Architecture.

Section 3 describes the EGI Platform Architecture itself, its building blocks and all necessary definitions of stakeholders and actors.

Section 4 describes activities that are more venturing, or crosscutting than those described for specific platforms, leading the evolution of the platforms. Proofs of Concept for new potential platforms are described in this section, as well as activities that affect many platforms at the same time, for example Proofs of Concept around advancement in the Authentication and Authorisation Infrastructure.

---

[1] https://wiki.egi.eu/wiki/Fedcloud-tf:Users#Current_FedCloud_Users_and_Communities

2 STAKEHOLDERS AND ACTORS

Transitioning from a vertical service delivery model to a horizontal platform deployment model requires re-examining and identifying stakeholders and actors that together deliver end-to-end services for the research communities. This ensures that also requirements of those are addressed and satisfied in the newly proposed EGI ecosystem.

While it is possible to hide the distinction of roles and stakeholders in a singular vertical service delivery model, service platforms enable a much greater independence of actors operating at the various platforms built on top of each other. By carefully defining and scoping the different roles and identifying relevant stakeholders in a horizontally organised open platform-oriented EGI ecosystem, identification of business opportunities and orchestration of the various activities become much clearer. Synergies can be thus much more clearly leveraged.

This section will define the stakeholders and actors in the EGI ecosystem. Following sections will apply these to the specific platforms discussed later-on. Table 1 provides an overview of the terms and verbs that are explained as follows.

Stakeholder
Resource Infrastructure Provider (RP)
European Grid Infrastructure (EGI)
Technology Provider (TP)
Platform Integrator (PI)
Research Community (RC)

Actor Name	Actor Verb
Platform Owner	Own
Platform Contributor	Contribute
Platform Packager	Package
Platform Deployer	Deploy
Platform Operator	Operate
Platform Users	Use

Table 1: Stakeholders and Actors in the EGI Platform ecosystem

2.1 Stakeholders

2.1.1 Resource Infrastructure Provider

Infrastructure providers are those partners in EGI who provision the tangible resources (compute, storage and network) in the EGI Infrastructure Platform providing transient access to these resources for a certain amount of time to a known set of users.

Resource infrastructure providers carry the risks and responsibilities of ownership of those resources, but at the same time have the control on whom they allow access to these resources. Resource infrastructure providers are interested in a sustainable customer base that does not threaten their business models should one or more customers terminate the business relationship. Therefore infrastructure providers require a platform that exposes their resources securely, yet allows for flexibility and uniformity irrespective of how customers are actually making use of the leased resources – tying one's business models into the customer's business models potentially threatens sustainability of an infrastructure provider as a whole.

2.1.2 European Grid Infrastructure (EGI)

EGI is a federation of Resource Infrastructure Providers (mostly NGIs and EIROs), captures EGI Federation members' vested interest in pan-European services. In the context of this document, this comprises central technical services that members of the EGI federation outsource to one of their members, coordinated through EGI.eu.

2.1.3 Technology Provider

Technology Providers develop software according to available requirements and needs. Architecture and design of the software may be community specific, or of general purpose that may serve any consumer. Depending on their involvement in the EGI ecosystem, business and service models, or the software's main aims, the actual interest in the EGI ecosystem may vary across Technology Providers. For example the Apache Foundation may have very little interest in the EGI ecosystem as such, yet it must be considered as a Technology Provider in the Platform Integrator's choice of software suppliers. On the other hand a Technology Provider may have strong interests in providing software tailored to the needs of a supported community. The EGI Application Database provides many examples of applications written and maintained by Technology Providers with dedicated, specific community scope. Current Technology Providers for the EGI community provide software that is deployed in the current EGI production infrastructure.

Typically, Technology providers deliver software as source code, binaries compiled for a specific execution platform, or both (just like the software registered in EGI AppDB today). Delivery of software ranges from online code repositories that are either self-managed or externally managed to shipped media. Support services for the provided software varies greatly, depending on the Technology Provider's business model.

Engagement with Technology Providers may happen at all platforms present in the EGI ecosystem, from the EGI Infrastructure Platform to the various EGI Community Platforms built on top of it (see below).

2.1.4 Platform Integrator

Platform Integrators architect and design a platform according to identified requirements delivered by their customers. During that process Platform Integrators match the requirements against available software and select the most suitable software components according to additional criteria (such as ease of customization & configuration). An important aspect of this platform design process is the actual selection of a suitable lower-level platform to integrate with. Depending on customer requirements, available software, engineering skill sets and licensing models (next to many other potential selection criteria) a Platform Integrator may choose one, or many lower-level platforms for integration.

In an ideal world, a Platform Integrator may choose from available software that behaves perfectly well according to documented interfaces and deployment guidelines. In reality, however, this is often not the case, and software "glue" (e.g. adapters for certain incompatible functionality) is required to be able to integrate two components. That glue software is hidden, and not included in the official external public platform interface and documentation. The extent of the required integration effort has a strong influence on the selected components, ranging from near to zero integration effort of perfectly interoperable components to significant integration effort for components that are used beyond their original intent.

2.1.5 Research Community

In the EGI ecosystem, Research Communities are groups of collaborating researchers that sustain a distinct (perhaps dedicated) management function that coordinates activities within the Research Community, and maintains relationships with other, external stakeholders within or without the EGI ecosystem.

Research Communities pursue strategic goals for the benefit of the collaborating scientists and research projects the Research Community participates in. As such, Research Communities are interested in platforms that deliver exactly the functionality they need, and responds efficiently and timely to evolving needs.

2.2 Actors

2.2.1 Platform Owner

A Platform Owner typically drives and controls the overall shape of a platform. The term "owning" does not always imply legal or financial ownership; typically for EGI owned platforms the responsibility and platform ownership is delegated from EGI's stakeholders to EGI.eu.

2.2.2 Platform Contributor

The Platform Contributor describes an actor capturing the role of Technology Providers that develop and maintain software that is included in any platform in the EGI ecosystem. The term *contributor* does not imply an active role in the development and maintenance of any given platform. In fact, it is often the case that contributing is a *passive* role in this ecosystem, particularly for platforms that depend on Open Source software and projects. For example, contributors are often Platform Integrators who use a bespoke Open Source solution (e.g. Open Stack) and contribute patches and extensions to the Open Stack repository for all other contributors and users to share.

2.2.3 Platform Packager

Platform Packagers turn the documented architecture and design of a given platform into artefacts that can be deployed on a target platform. Depending on the scope and definition of the platform those artefacts may be binary code packages such as RPM archives, or a larger structure and set of packages that together deliver a service as part of the platform.

With that, Platform Packagers take care of the technical platform lifecycle. This begins with assembling and publishing the initial release of the platform as a whole, including the technical documentation. The packager then monitors the development activities within the individual lifecycles of the included components. If required the packager plans and initiates updates to the platform components, thus creating a lifecycle of their own for the deployable platform components.

Platform Packagers often re-use components in order to simplify the process of aggregating low-level functionality into a higher-level service. Re-using software component also reduces the number of dependencies and effort necessary to monitor and track the development of the selected components. On the other hand, Platform Packagers must keep an eye on the quality of the selected components in terms of software defects (software problems and vulnerabilities), since each re-use of a component raises the impact of any of those software defects in the deployed platform.

2.2.4 Platform Deployer

Often overlooked as a distinguishable role, the Platform Deployer takes care of rolling out the components of a chosen platform at a specific time and place and configuration. This is documented in detailed roll-out plans that align with generally planned maintenance cycles of the production infrastructure (or may warrant a specific maintenance cycle if required).

In due time the Platform Deployer then implements the planned roll-outs (or updates the deployment plans). Each rollout of a platform component is documented in the "roll-out history" of that component to reflect the most current configuration state of that element for post-rollout consultation and troubleshooting. This is often referred to as a "configuration item" whose current state is maintained in a configuration management database (CMDB)..

The role of a Platform Deployer is often assumed by Platform Operators since the topic of their duties is identical. However, viable scenarios separate those roles where the Infrastructure Provider conducts the platform deployment (as a service) while the consuming Research Community is assuming the role of the Platform Operator.

2.2.5 Platform Operator

Platform Operators – as the name suggests – operate a deployed and initially configured platform on top of its selected infrastructure. This day-to-day activity includes monitoring the platform infrastructure, administering changes if operational metrics are outside of acceptable upper or lower bounds, and reporting for pro-active platform provisioning.

Platform Operators are therefore interested in platforms that are easy and efficient to manage. Their requirements on a platform focus on scalability, reliability, accuracy and efficiency of the platform management infrastructure.

2.2.6 Platform User

Generally, all those individuals that access a given platform, or any type of software, are summarised as "end users". For the purpose of (at least) this document, these users are described as Platform Users. Platform Users (primarily from various research communities) use the chosen Community Platform; they consume its services and underlying resources, without maintaining the business relationships that make this possible.

In an "end-to-end" description of service delivery (see Figure 4), Platform Users would be located at one end of the service value chain, while the Resource Infrastructure Providers are located at the other end of that chain. In a platform oriented service delivery model this notion does not change for the Platform Users, they still consume the services that were deployed for their direct use.

This description does not imply a passive or receiving-only role. Instead, Platform Users are the main suppliers of functional requirements that reflect the needs of the respective community. By proxy, the Research Community ensures that the Community Platform, either by requesting a change to existing platform components or by having them replaced by a better alternative, meets these requirements. These requirements are the main drive the virtuous cycle of continuous service delivery to the Platform Users.

3 EGI PLATFORM ARCHITECTURE

The EGI production infrastructure is a mix of a number of platforms deployed across its entirety. All services deployed in EGI are part of a particular platform, as illustrated in Figure 1.
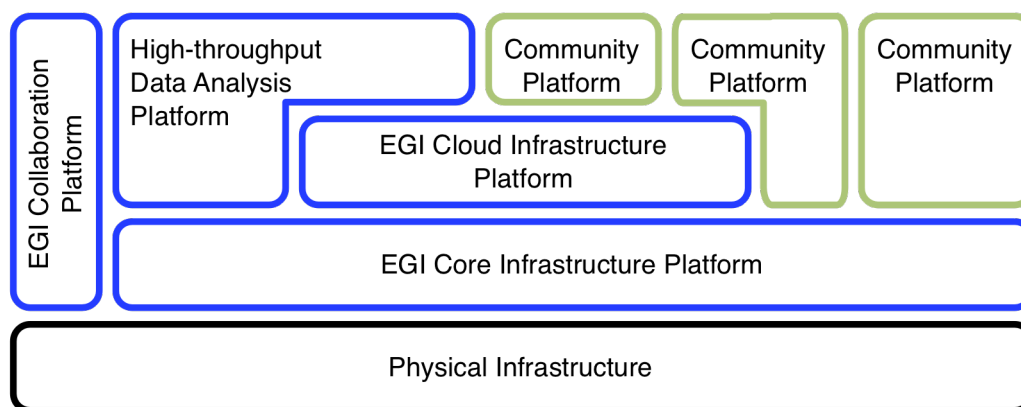

Figure 1: The EGI Platform Architecture

Each platform in this architecture follows a clear scope and definition as provided later in this document. The colour coding in Figure 1 follows a particular schema as follows:

- **Black** denotes direct legal and financial ownership of the physical infrastructure as exerted by the Resource Infrastructure Provider within the EGI production infrastructure (typically NGIs and EIROs).
- **Blue** platforms are owned by EGI, i.e. EGI is the Platform Owner, and offers these for integration and use by its research communities.
- **Green** platforms are *not owned* by EGI; the mapping of stakeholders to actors varies across these.

Note that the architecture diagram is *not to scale* with respect to deployment and adoption. That said, some platforms might be larger or smaller than depicted, relative to others; at the same time there may be many more platforms deployed on top of the EGI-owned platforms.

This is the fundamental benefit of the platform architecture: It allows to re-use proven processes and interfaces at scale.

Following, very brief descriptions of each EGI-owned platforms are given. Together, these span EGI's technical service portfolio ready for its research communities to consume.

3.1 EGI Core Infrastructure Platform

The **EGI Core Infrastructure Platform**'s mission is to "*enable operational federation of any platform deployed in EGI*". It includes only services that are necessary to establish and operate a federated distributed computing infrastructure. This implies that, unlike before, there are no core resource access services included in the EGI Core Infrastructure Platform. The underlying strategic decision was introduced in the EGI Strategic Plan [R 1] and the accompanying Technical Roadmap [R 2]. From there on this concept was further evolved and discussed with major stakeholders, e.g. at the EGI Technical Forum [R 6] and the Evolving EGI workshop in January 2013 [R 7]. This division

allows the EGI Core Infrastructure Platform to be offered as a service to other e-Infrastructures and Research Infrastructures in Europe and around the world, without imposing a certain resource access and usage paradigm to the partnering e-Infrastructures or Research Infrastructure (see also the online version of the EGI solutions portfolio[2]).
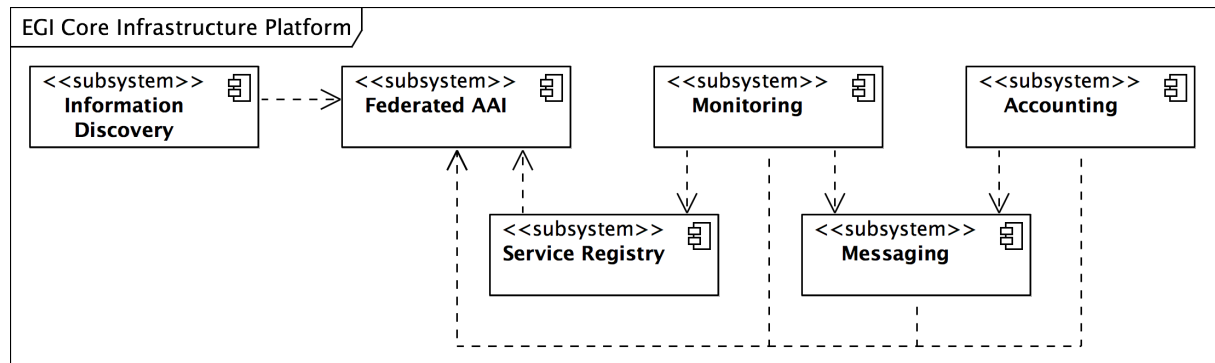

Figure 2: Components of the EGI Core Infrastructure Platform

The EGI Core Infrastructure Platform (CIP) is deployed directly on top of the physical hardware owned by the Resource Infrastructure Provider in the management infrastructure part of the provider's physical infrastructure. The EGI Core Infrastructure Platform is owned, deployed, packaged and operated by the Resource Infrastructure Providers federated into EGI. These activities are coordinated through management and collaboration boards and groups that are part of the EGI IT Service Management domain.

The EGI CIP comprises of the following seven components:
- Messaging
- Federated Authentication and Authorisation Infrastructure (AAI)
- (Service Availability) Monitoring
- Accounting
- Central Services Catalogue
- Information Discovery Service
- Metrics visualisation

---

[2] http://www.egi.eu/solutions/fed-ops/

Figure 2 illustrates the principal composition of the EGI Core Infrastructure Platform. The composition of the diagram indicates components that form the integration surface of the EGI CIP, and which are designed to be operational services (i.e. Metrics visualisation and Central Service Catalogue). At the time of writing not all five available integration components (Information Discovery, Monitoring, Federated AAI, Messaging, Accounting) are compulsory for Community Platforms to integrate with. Currently, Community Platforms are expected to integrate with the Federated AAI, Monitoring and Accounting components, while the Central Service Catalogue will have to accommodate for the registration of the Community Services. From a technical perspective, the Messaging component is used only as a sub-component of Accounting and Monitoring, but it has in and by itself potential as a first class component in the EGI Core Infrastructure Platform, providing value to *any* Community Platform deployed on top of EGI Platforms. A specific service model would need to be developed, but is expected to align with provisioning specific messaging topics and queues according to agreed service levels (e.g. throughput, retention, reliability, points of presence).
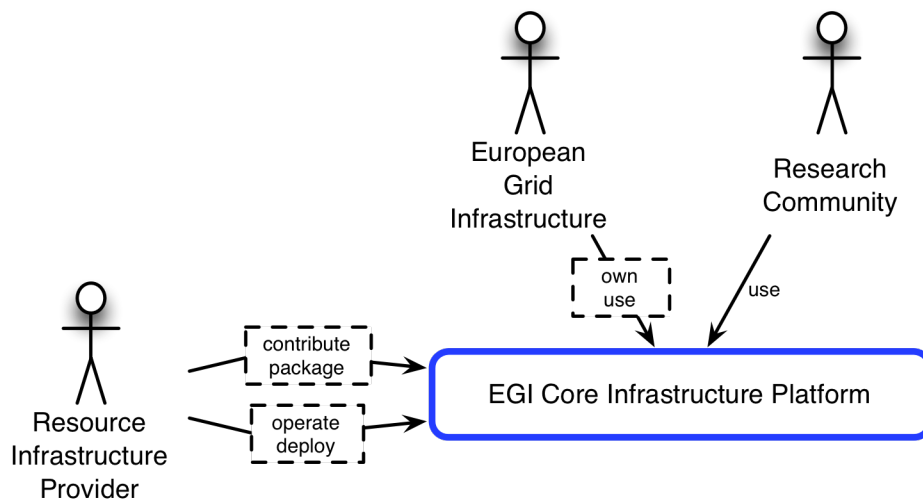

Figure 3: Actors for the EGI Core Infrastructure Platform

Figure 3 illustrates the relationships for the EGI Core Infrastructure Platform (CIP). While EGI, represented through EGI.eu owns and evolves the EGI CIP, the Resource Infrastructure Providers are deploying and operating the CIP. As it is the foundation for the entire EGI production infrastructure, the CIP is the most widespread and most used platform in EGI.

Research Communities might directly integrate with the CIP in platform models that deploy directly on the physical infrastructure (which, for the purpose of this document, includes POTS operating systems like some of the many different Linux distributions).

Most dominantly, Resource Infrastructure Providers are also Technology Providers for the CIP, funded partially through the EGI-InSPIRE project (JRA1 & JRA2 work packages), and other public funding streams.

3.2 EGI Cloud Infrastructure Platform

To provide generic, consistent and flexible access to EGI resources, EGI initiated a strategic activity to establish a federation of locally deployed IaaS Clouds. The **EGI Cloud Infrastructure Platform**'s mission is to provide "*consistent access to federated IaaS Cloud resources*". It directly supports EGI's strategic alignment with the European Commission's Horizon 2020 strategy [R 8]. While EGI will continue to support and maintain its existing relationships with research communities, it will provide the Cloud Infrastructure Platform in support of new research communities who wish to deploy their own Virtual Research Environment on the resources that they are able to access. This enables less structured research collaborations and individual researchers stemming from the so-called "long tail of science" to easily access and configure EGI's resources.

At the heart of the EGI Cloud Infrastructure Platform (CLIP) are locally deployed Cloud Management stacks. In compliance with the Cloud computing model, the EGI CLIP does not mandate deploying any particular or specific Cloud Management stack; it is the responsibility of the Resource Providers to research, identify and deploy the solution that fits best their individual needs for as long as the offered services implement the required interfaces and domain languages.

Consequently, the EGI Cloud Infrastructure Platform is built around the concept of an *abstract* Cloud Management stack subsystem that is integrated with components of the EGI Core Infrastructure Platform (see Figure 4).
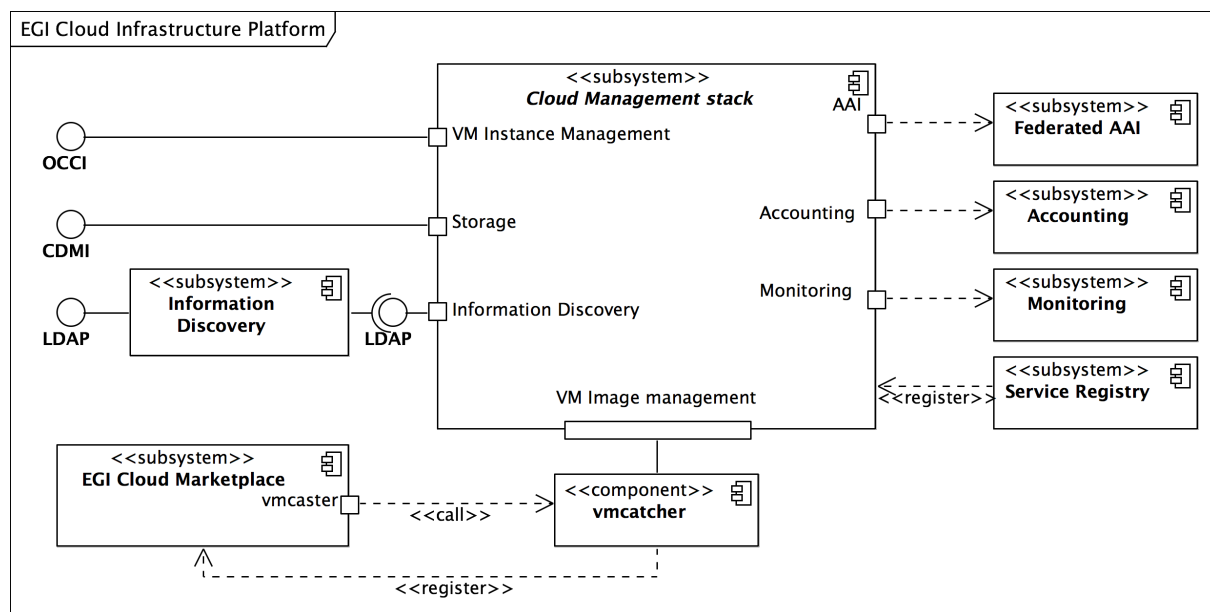

Figure 4: Architecture of the EGI Cloud Infrastructure Platform

As already indicated in section 3 the EGI CLIP is owned by the EGI Resource Infrastructure Providers through EGI. However, taking the nature of Cloud Computing into account, allows EGI to define the CLIP as a relatively thin layer of federation and interoperability around local deployments and integrations of Cloud Management stacks.

This architecture is modelled using an *abstract* Cloud Management stack subsystem (see Figure 4) defining interaction ports with a number of services from the EGI Core Infrastructure Platform, and the EGI Collaboration Platform. At the same time, it defines the required external interfaces and corresponding interaction ports. All these ports will have to be realised by local Cloud Management stack deployments.

The main interaction points of the EGI CLIP thus are:

- Integrate with the EGI Core AAI
- Integrate with the EGI Core Accounting system
- Integrate with the EGI Core Monitoring system
- Provide a standardised Cloud Computing interface (OCCI)
- Provide a standardised Cloud Storage interface (CDMI)
- Provide a standardised interface to an Information Service

Additionally, by means of using the Appliance Repository and the VM Marketplace from the EGI Collaboration Platform (COP) the EGI CLIP is providing VM image sharing and re-use across EGI Research Communities.

At the time of writing the EGI CLIP is maintaining its own, separate Information Discovery system. Even though it is using the GLUE2 schema, some extensions are not yet compatible with the canonical GLUE 2 specification. Hence Cloud Resource providers maintain local LDAP endpoints (usually deployed as a resource BDII) aggregated into a Cloud Platform Information Discovery service, which in turn allows access to the data using LDAP v3.

Figure 5 provides an overview of the current realisations of the abstract Cloud Management stack subsystem in the EGI CLIP. It illustrates that each existing realisation inherits the obligation to implement the interaction points from the generalised parent Cloud Management stack. At the same time, the EGI Federated Clouds Task (funded through the EGI-InSPIRE project) gives Resource Providers a platform to share their implementation solutions for a commonly deployed specific Cloud Management stack (e.g. OpenNebula and OpenStack).
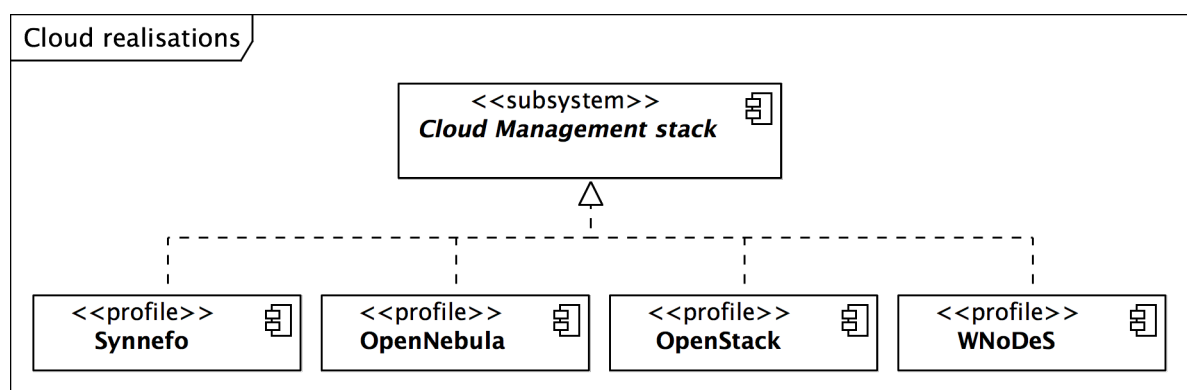

Figure 5: Current realisations of the abstract Cloud Management stack component

Through this collaboration, Resource Providers gradually develop and mature deployment and configuration profiles around common Cloud Management stacks. Consequently, the stakeholder relationships are different for the EGI Cloud Infrastructure Platform (see Figure 6).

As one of the key platforms, the CLIP is owned by EGI, while it is deployed and operated by the member Resource Infrastructure Providers. However, the used software and services are almost

entirely independent Open Source projects. Independent Technology Providers contribute the bulk of the software, without being responsible for the actual contents of the platform (which is shared between EGI and the RPs). However, the used Open Source software needs to be adapted and integrated, so that it operates well with the other elements of the EGI production infrastructure. These contributions to the platform are provided by the RPs themselves, often in a shared and collaborative fashion grouped around the Cloud Management Frameworks that are deployed within EGI.
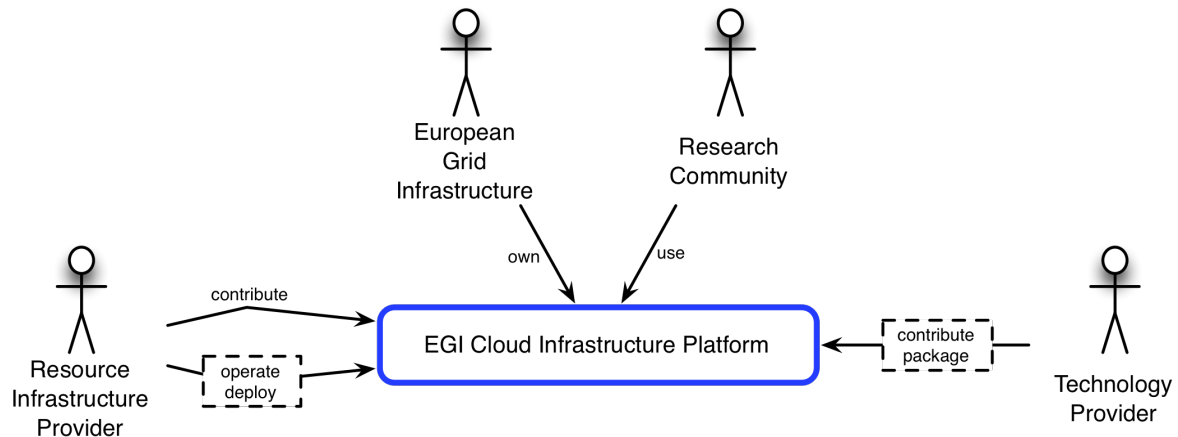

Figure 6: Stakeholder collaborations around the EGI Cloud Infrastructure Platform

Research Communities may directly use the CLIP to deploy their own software and services in the EGI production infrastructure (as part of their community platforms), though it is not expected that *all* EGI Research Community will do so.

3.3 EGI Collaboration Platform

Complementing the two previously outlined platforms, the **EGI Collaboration Platform** "*offers services enabling and facilitating collaboration across EGI's research* communities". It facilitates synergies between research communities by encapsulating services that are common across multiple communities and are not critical to the operation of the EGI production infrastructure, therefore are outside of the Core and Cloud platforms. The EGI Collaboration Platform complements the other platforms and contributes to their efficient use. As a feature of the Platform architecture, EGI's platforms also make use of services stemming from the Collaboration Platform.

The EGI Collaboration Platform (CoP) is a collection of components that facilitate collaboration in the EGI community. Broadly, these services can be further structured into two categories of services:

- Social collaboration, and
- Technical collaboration.

While technical collaboration tools are designed and offered for integration with technical components of any number of Community Platforms deployed in EGI's production infrastructure, the social collaboration components are typically tools and services that require direct user interaction (see Figure 7).
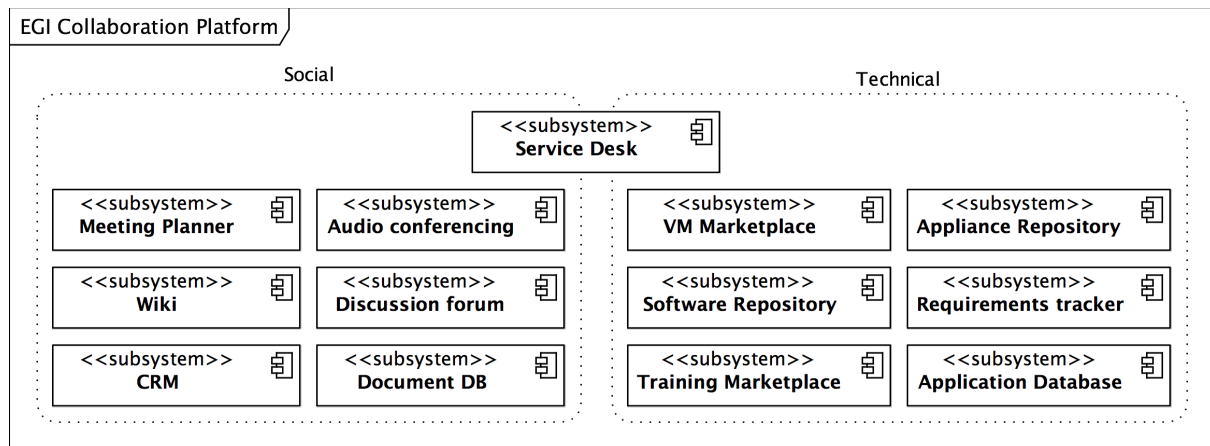
Figure 7: EGI Collaboration services capture social and technical collaboration needs

Similar to the Core Infrastructure Platform, the EGI Collaboration Infrastructure originated from the same original software portfolio developed in the EDG and EGEE series of projects. Hence the relationships to the contained services are identical, though with exceptions in the details (c.f. the audio conferencing service, which is sourced in as an entirely external service).
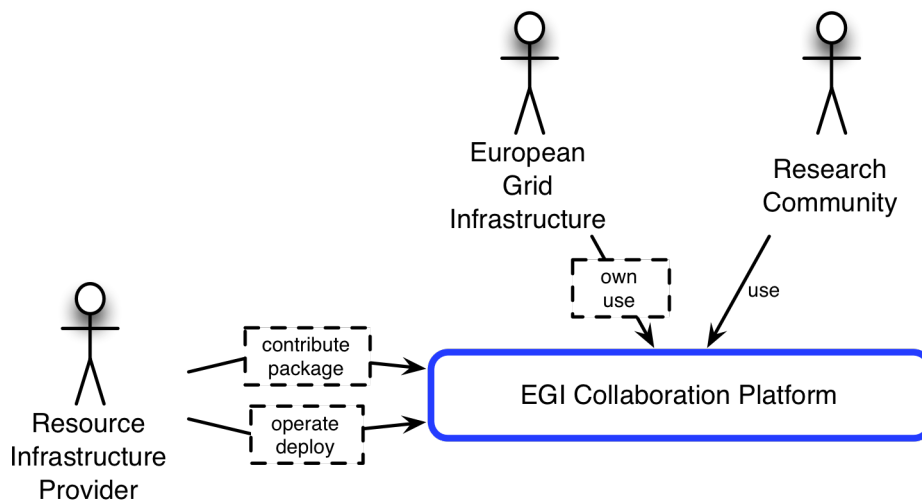

Figure 8: Relationships for the EGI Collaboration Platform

Through this legacy, some of the services grouped in the Collaboration platform are home-grown solutions that did not exist at the time these solution were conceived, or did not fit the requirements at that point in time. Many are based on Open Source solutions (e.g. DocumentDB, Discussion Forum, CRM, Wiki, Requirements Tracker, etc) that were integrated into the Collaboration Platform.

3.4 EGI High-Throughput Data Analysis Platform

The High-Throughput Data Analysis Platform is the successor of the series of EMI platforms deployed via the UMD in the EGI production infrastructure. Re-using a reduced set of components

from the latest EMI-3 platform release, the HT Data Analysis Platform is essentially a Community Platform sustained by its user community stakeholders under permanent stewardship of EGI.
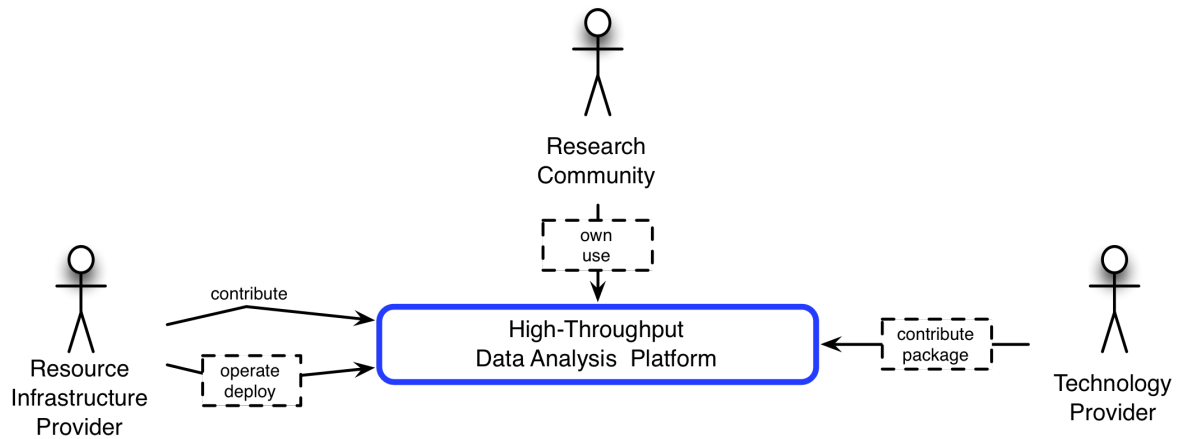

Figure 9: Relationships for the High-Throughput Data Analysis Platform

4 ROADMAP

Naturally, with the evolution of EGI, research infrastructures in Europe, and technology itself, the EGI Platforms will be subject to changes themselves to align with new requirements, new technologies, and other changes at large. Clearly, not all platforms are at the same level of maturity hence some will show more activity in terms of extending its portfolio than others.

The following sections indicate activities and strategic directions of evolutions for each platform as an executive overview of the entire activities with further reading indicated where applicable.

4.1 Core Infrastructure Platform

As is expected the Core Infrastructure Platform is mature and stable. New inclusions or evolution will be carefully managed in a Technology Insertion process, as any change in the Core Infrastructure Platform will have great impart on the entire production infrastructure.

4.1.1 Permanent Identifiers

The development of an open data infrastructure for ERA will drive the evolution of the Core Infrastructure Platform, which will need to include services for resolving identifiers to permanent digital objects. The extension towards this direction will be lead by the requirements of user communities about data discovery and dissemination, and will be considered in 2015.

4.1.2 Authentication and Authorisation Infrastructure (AAI)

The AAI is one of the very few essential components whose absence is fatal for any meaningful production infrastructure.

For many years, the EGI production infrastructure relied on the mature and stable service provided by the International Grid Trust Federation (IGTF) and its regional peers such as the EUGridPMA for the currently X.509 v3 certificate-based Authentication infrastructure.

For as long as EGI and its predecessors served research communities with very similar requirements, the existing AA infrastructure was enough. With the decision to diversify and support many more research communities, new requirements for a more flexible AAI need to be supported and implemented.

EGI's AAI will be subject to careful examination to extend the currently supported authentication and authorisation mechanisms with additional means for identifying and managing user information where and when required.

Indicative timeline:

2014:   Conduct a Proof of Concept to examine a number of alternatives as candidate technologies for inclusion

2015 – 2016: Collaborate with GEANT and Terena to integrate existing solutions and implementing services addressing remaining gaps, and support early adopter infrastructures and research communities

2017:   Begin to roll out extensions to EGI's AAI into production.

4.2 Collaboration Platform

Few activities around the Collaboration Platform are targeting mainly the Help Desk services (GGUS) and the Application Database as described below.

4.2.1 GGUS

The GGUS system itself is stable; most of the work performed is maintenance and careful service evolution hence strategic longer-term activities are not foreseen at the time of writing. However, some of the medium term activities are noteworthy as they may inform and influence activities in other platforms.

Shibboleth authentication: GGUS will support Shibboleth as a federated authentication infrastructure, allowing maintainers and supporters for xGUS instances to re-use their institutional credentials.

Expected delivery: 1H 2014

Migrating main GGUS to xGUS: The xGUS framework has proven sufficiently scalable to more the main GGUS instance to use the xGUS framework. With this shift, the entire GGUS service makes a significant step towards being considered a cloud service; perhaps coined as "Help desk as a service".

Expected delivery: 1H 2014

Cross e-Infrastructure Integration: the integration of GGUS with other e-infrastructure helpdesks will be considered if required by users.

RI-EGI Integration: GGUS with its central interface to a distributed set of regional and local helpdesks is modular and can be extended by integrating with Research Infrastructure-specific helpdesks systems. This was already accomplished in PY4 for the MAPPER user community as an outcome of the EGI-MAPPER collaboration.

4.2.2 Application Database (AppDB)

The Application Database is currently being heavily extended towards becoming an integrated key service of the EGI Cloud Infrastructure Platform. This is part of a longer-term strategy to evolve the Application Database into a marketplace for federated cloud services that is expected to prevail in the academic Cloud e-Infrastructure provided by EGI in the future. Continued development of the tool is ensured during PY5 thanks to the project extension.

Indicative timeline:

Q2 2014:   Full availability of the Cloud Marketplace for researchers and VOs to register public virtual appliances. AppDB takes on the role of either creating and publishing image lists on a user's or VO's behalf, or republish these into the EGI federated Cloud infrastructure.

Q2 2014:   Full support of private image lists (i.e. registered in AppDB but not available for others to re-use or download)

Q4 2014:   Tighter integration with the Resource Providers in terms of VO enabling at the provider side, automating image list subscriptions and feedback to users.

4.3 Cloud Infrastructure Platform

The Cloud infrastructure is currently the platform with the most development and evolution of its services.

The most noteworthy milestone for the Cloud Infrastructure Platform is its production launch in May 2014 in the context of the EGI Community Forum in Helsinki, Finland. With this, EGI will offer federated compute and storage services in the Cloud in production quality as a first class resource in its solutions portfolio. Mirroring the fast-paced nature of the Cloud sector, a number of activities and Proofs of Concepts are taking place to assess which future extensions to the EGI Federated Clouds solution portfolio need to be tackled as follows.

Digital Preservation as a Service: EGI is currently involved in a number of EC funded projects such as DCH-RP, Civic Epistemologies, SCIDIP-ES, EISCAT-3D and other projects that are facing preservation issues as part of the project, or the respective sustainability strategy. With support and collaboration of other European projects such as DCH-RP, APARSEN and Open-AIRE, EGI is conducting a Proof of Concept experiment to provide a preservation platform as a service to the digital cultural heritage community, underpinned by the EGI Cloud infrastructure Platform. Demonstrating feasibility of such a platform, the outcome of this PoC together with results from other projects will feed into the assessment of a production quality preservation as a service product offered to the European Research ERA as a service in the future.

High-throughput data analysis in the Cloud: Together with the DIRAC INTERWARE community EGI is investigating how the pilot job deployment pattern can be adapted to utilise Cloud infrastructures. By implementing the OCCI interface natively into the VMDIRAC extension, any user community integrating the DIRAC framework into their workflows will be able to transparently utilise the EGI Federated Cloud. This collaboration has led to EGI and the DIRAC community discussing the integration roadmap for DIRAC into the High-Throughput Data Analysis Platform provided by EGI to its users.

Brokering, auto-scaling and dynamic VM placement: The EGI Federated Cloud collaboration is currently examining a number of alternative solutions that are able to support a variation of these capabilities. The expected outcome is not necessarily intended to be included in EGI's Cloud Infrastructure Platform. Rather, they will be offered to users as solutions and services known to "work with EGI" with out further support from EGI. Currently there are three solutions under examination: Slipstream, VMDIRAC and COMPss.

MapReduce platforms as a service: Through collaborating with the Peachnote community EGI is investigating how a number of alternative Map&Reduce platforms such as Hadoop and HBase can be provided to user communities in the future.

Continuous Integration / DevOps as a Service: Together with the WS-PGRADE project, EGI is examining how a federated Cloud infrastructure may be used to support Software Engineering at a European scale. As a spin-off, a DesktopGrid infrastructures may extend into the Cloud using the WS-PGRADE/gUSE framework.

Science Gateways as a Service: In collaboration with the CHAIN-REDS project, the e-Science Gateway Framework, developed at INFN-Catania and a first class product in CHAIN-REDS, might lead to EGI's Federated Cloud portfolio include an easy to use gateway service for its users that do not want to use the IaaS interfaces directly. Already, a number of applications for such a service are devised and planned for going into production in 2014 via the CHAIN-REDS project.

RNA sequencing: In collaboration with the University of Wuerzburg in Germany, EGI is conducting a Proof of Concept on providing RNA sequencing services in the Cloud.

In summary, there are many more activities and Proofs of Concepts taking place in context of the Cloud Infrastructure Platform[3]. A projection or reliable timeline for these activities are mostly impossible, as they are often conducted on a voluntary or best effort basis, or external project timelines and roadmaps change without further notice to EGI. Even though the Cloud Infrastructure Platform is not in production yet, it already has sparked a wide variety of collaborations and Proofs of Concepts in this area that may lead to potentially rapid expansion of the Federated Cloud solution portfolio in the future.

4.4 High-Throughput Data Analysis Platform

The High-Throughput Data Analysis Platform is in itself mature, and the successor of the EMI portfolio of services in a slightly changed setup and configuration. Regular maintenance is conducted as part of ordinary operational maintenance activities conveying updates via the UMD provisioning activities in EGI.

Currently, the stakeholders of the High-Throughput Data Analysis Platform are discussing the inclusion of DIRAC (including its extension VMDIRAC) in the platform as a workload distribution system next to EMI WMS (Workload Management System). With its extension VMDIRAC, DIRAC is a promising candidate to seamlessly extend high-throughput data analysis into the Cloud.

The main evolution for the High Throughput Data Analysis Platform will be new tools, or the consolidation and wider adoption of existing tools, for the workload management, that is the capability for users to have a single entry point.

The most promising among the tools is the DIRAC Workload Management System. This pilot jobs-based WMS provides all the necessary components to ensure efficient execution of user jobs using heterogeneous computing resources. The job management makes a special emphasis on operations in a distributed computing environments, and provide fault tolerant mechanisms to compensate instability of some of the distribute resources.

The DIRAC system is a pilot-jobs-based system, where user payload is submitted through a robot-certificate, reducing in this way the complexity of handling grid certificates to new users, and is executed by pilot jobs submitted to the resource centres that are more reliable and where the probability to have a successful execution is higher. The DIRAC framework is suitable for large user communities with complex internal policies of computing resource usage for massive data processing, and small user groups requiring a simple interface to heterogeneous resources.

On the storage management side, DIRAC Data and Storage Management Systems, DMS & SMS, provide tools for seamless access to various types of data storage, data cataoging and for massive data

---

[3] https://wiki.egi.eu/wiki/Fedcloud-tf:Users#Current_FedCloud_Users_and_Communities

replication. DIRAC system can be used directly by grid users for job submission, or connected to their VRE to improve the job submission efficiency. Some big user communities already use DIRAC in production, and in the second half of PY4 several pilots have been deployed in collaboration with the DIRAC developers and other VOs, with good results and feedback from the users.

5 REFERENCES

R 1	Secure Federated Data Analysis Capability for the European Research Area, EGI Council Strategy, September 2013 (http://www.egi.eu/news-and-media/publications/EGI-GrandVision-V6.pdf)
R 2	D2.31 – EGI Technical Roadmap, https://documents.egi.eu/document/1094
R 3	MS510 – EGI Platforms Roadmap, 1st edition, https://documents.egi.eu/document/970
R 4	MS514 – EGI Platforms roadmap, 2nd edition, https://documents.egi.eu/document/1624
R 5	D2.33 – EGI Technical Roadmap, to be published
R 6	EGI Technical Forum 2012, Prague, Czech Republic, 17-21 Sep 2013, http://indico.egi.eu/indico/conferenceDisplay.py?confId=1019
R 7	Evolving EGI Workshop, Amsterdam, The Netherlands, 29-30 Jan 2013, http://indico.egi.eu/indico/conferenceDisplay.py?confId=1252
R 8	EC Horizon 2020 strategy, http://ec.europa.eu/research/horizon2020
R 9	Post-EMI/IGE support for Technology Providers, https://documents.egi.eu/document/1499
R 10	TCB Terms of Reference, https://documents.egi.eu/document/109
R 11	URT Terms of References, https://documents.egi.eu/document/1618
R 12	ITIL 2011 Glossary, http://www.itil-officialsite.com/InternationalActivities/ITILGlossaries_2.aspx
R 13	EGI Technology Glossary, https://wiki.egi.eu/wiki/Technology_Glossary_Contents