

**EGI-Engage**

Infrastructure tests and best usage practices
for life science service providers

D6.10

|  |  |
| --- | --- |
| **Date** | 31 May 2016 |
| **Activity** | SA2 |
| **Lead Partner** | CSC |
| **Document Status** | DRAFT |
| **Document Link** | https://documents.egi.eu/document/2802 |

Abstract

ELIXIR Competence Centre (CC) aims to bring the EGI resources, especially the EGI Federated Could, better available to the ELIXIR user community. This document sums up the experiences of those ELICIR CC members who are providing resources for EGI Federated Cloud and/or have utilized EGI resources for providing life science services.

**COPYRIGHT NOTICE**



This work by Parties of the EGI-Engage Consortium is licensed under a Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/). The EGI-Engage project is co-funded by the European Union Horizon 2020 programme under grant number 654142.

**DELIVERY SLIP**

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***Name*** | ***Partner/Activity*** | ***Date*** |
| **From:** |  |  |  |
| **Moderated by:** |  |  |  |
| **Reviewed by** |  |  |  |
| **Approved by:** |  |  |  |

**DOCUMENT LOG**

|  |  |  |  |
| --- | --- | --- | --- |
| ***Issue*** | ***Date*** | ***Comment*** | ***Author/Partner*** |
| **v.1** | 29/Apr/2016 | Document skeleton | G. Sipos / EGI.eu |
| **v.2** | 2/May/2016 | ToC | K. Mattila / CSC |
| **v.3** | 3/May/2016 | Corrected file | K. Mattila / CSC |
| **v.4** | 4/May/2016 | Restructuring after ELIXIR CC teleconference | G. Sipos / EGI.eu |
| **v.5** | 4/May/2016 | Initial text into Introduction, ECP description | G. Sipos / EGI.eu |
| **v.6** | 13/May/2016 | Some draft chapters added | K. Mattila / CSC |

**TERMINOLOGY**

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

**Contents**

1 Introduction 5

2 The ELIXIR Compute Platform; Role of service providers 6

3 Integration status and plans 8

3.1 CSC 8

3.2 EMBL-EBI 9

3.3 CESNET 9

3.4 CNRS 9

3.5 GRNET 10

3.6 SURFsara 10

3.7 JetStream 10

3.8 Other providers? 10

4 Integration guidelines and tips for service providers 11

4.1 Generic concepts and installation guidelines EGI 11

4.2 Installation guideline for OpenStack providers EGI 12

4.3 OpenStack-specific experiences, recommendations, tips EBI 12

4.4 Installation guideline for OpenNebula providers EGI 12

4.5 OpenNebula-specific experiences, recommendations, tips CESNET 12

4.6 Installation guideline for Synnefo providers EGI 13

4.7 Synnefo-specific experiences, recommendations, tips GRNET 13

4.8 JetStream integration experiences Indiana University 13

5 Report on AAI integration 14

5.1 Integration of ELIXIR AAI with EGI AAI proxy Christos 14

5.2 Integration of GOCDB with the EGI AAI proxy David 14

5.3 Integration of AppDB with the EGI AAI proxy Marios 14

Appendix I. Appendix example 15

**Executive summary**

ELIXIR[[1]](#footnote-1) is a pan-European research infrastructure in agreement between 17 European governments to build a sustainable European infrastructure for biological information, supporting life science research and its translation to medicine, agriculture, bioindustries and society.

EGI[[2]](#footnote-2) is a pan-European e-infrastructure that delivers integrated computing services to European researchers, driving innovation and enabling new solutions to answer the big questions of tomorrow.

The ELIXIR Competence Centre (CC) of the EGI-Engage project evaluates, adopts and promotes technologies and resources from EGI to the wider ELIXIR research community. This report is the second outcome of this effort. The document (1) describes the concept of the ELIXIR Compute Platform and responsibilities of service providers participating in it, (2) provides a status update about the completed, ongoing and planned integration of resources in this platform, and (3) presents integration guidelines for life science providers who are wishing to participate in this infrastructure.

The ELIXIR Compute Platform is a reference technical architecture to support a vast range of data analysis activities. EGI is currently contributing to the platform development with several services and technologies from EGI – all relating to the management and access of a cloud federation.

Within this federation <status of providers>…

The protocol to join the federation is …

# Introduction

ELIXIR[[3]](#footnote-3) is a pan-European research infrastructure in agreement between 17 European governments to build a sustainable European infrastructure for biological information, supporting life science research and its translation to medicine, agriculture, bioindustries and society.

EGI[[4]](#footnote-4) is a pan-European e-infrastructure that delivers integrated computing services to European researchers, driving innovation and enabling new solutions to answer the big questions of tomorrow.

Life science is a fast moving field. For the EGI services to become relevant and help keep European Life Sciences competitive globally, it is important to develop mechanisms that allow the research infrastructure to flexibly meet new challenges and respond to new scientific and technical developments.

The ELIXIR Competence Centre (CC) of the EGI-Engage project evaluates, adopts and promotes technologies and resources from EGI to the wider ELIXIR research community. This is achieved with an iterative approach:

1. Bringing together designated life science experts from ELIXIR and technical experts from EGI within the CC.
2. Identify life science use cases that could benefit from EGI services and could make big impact on ELIXIR and EGI communities. Analyse the e-infrastructure requirements of the use cases.
3. Implement the use cases as demonstrators based on EGI e-infrastructure services. Collaborate during implementation with relevant EGI and ELIXIR partners, such as the EUDAT[[5]](#footnote-5) to create a generic infrastructure, the ‘ELIXIR Compute Platform’ that can underpin demonstrators and production applications from/for the ELIXIR community.
4. Demonstrate and evaluate the implementations. Disseminate the experiences gained with the use cases towards ELIXIR, EGI and other relevant communities. Decide about the long-term adoption of EGI services within ELIXIR based on the pilot experiences.

This document is a deliverable produced by stage 3 of this process. It captures the goals, current status and plans for the ELIXIR Compute Platform, and provides guidelines for interested service providers to join the platform with cloud services.

The document was written by life science and e-infrastructure experts from ELIXIR and EGI who are brought together within the CC.

# The ELIXIR Compute Platform; Role of service providers

During 2015 the ELIXIR community – in collaboration with various e-infrastructures and other service providers – initiated the development of the reference architecture for ELIXIR, called the ‘ELIXIR Compute Platform’ (ECP). The prime role of the ECP is to support the use cases of the ELIXIR-EXCELERATE H2020 project, however, the platform is expected to serve other ELIXIR-related use cases from ELIXIR and other biomedical sciences Research Infrastructures. The demonstrator use cases of the CC (documented in M6.3) will also use this platform.

The need for an ELIXIR reference technical architecture was first discussed during a BioMedBridges e-Infrastructure workshop in May 2014, where reference was made to the MONARC report[[6]](#footnote-6) that formed the basis of the Tiered model that was initially adopted by WLCG community to serve the needs of High Energy Physics. Following on from work by the ELIXIR Authentication and Authorization Infrastructure (AAI), Storage and Cloud Task Forces to define a set of Technical Use Cases, a workshop was held in Amsterdam (12-13th March 2015) to discuss with representatives of ELIXIR nodes, European e-Infrastructures and other service providers, how the ELIXIR‑EXCELERATE Scientific Use Cases could be mapped onto the Technical Use Cases and thereby define the ELIXIR Compute Platform. Through a series of presentations and breakouts the technical aspects of the Scientific Use Cases were identified and mapped to a number of Technical Use Cases. As a result of these discussions, a number of recommendations have been made for technical solutions that together will provide an ELIXIR Compute Platform. The platform can not only support the ELIXIR-EXCELERATE Scientific Use Cases, but a vast range of other data analysis activities that will be found within the ELIXIR research community. Such as:

* Hosting portals that enable users to select and launch virtual machines onto an available cloud resource (e.g. for training activities).
* Hosting web tools that deploy a network of virtual machine images onto distributed cloud resources operated for ELIXIR users for large scientific analysis.
* Provisioning ‘Desktop as a Service’ where researchers are able to obtain a desktop image (e.g. BioLinux) in a cloud that they can use for their data analysis activities that is always on for their use.

The role of ELIXIR and the ELIXIR-EXCELERATE proposal is not to undertake middleware development. Instead the focus is on leveraging the investment that has already been made in services that can be integrated for our needs and steer future development priorities. Essentially, our role is to define a minimal ‘neck’ of an hourglass that ELIXIR Researchers and Application Developers can build upon and that ELIXIR Nodes and other infrastructure service providers can deploy and support. The ECP is envisaged to consist of the following service groups:

* Basic Identity Environment: authentication and authorization related infrastructure (“AAI”) to provide user identity and access management services[[7]](#footnote-7) for ‘ELIXIR infrastructure services’ (all other services). The basic ELIXIR AAI environment is available since the end of 2015 and further developments and refinements are coming during 2016.
* Core Enabling Infrastructure Services: provide capabilities to store and effectively transfer data (storage management and file transfer services). ELIXIR and EUDAT are working together in the EUDAT2020 project to identify, test and deploy services for this area.
* Basic Infrastructure Services: Cloud IaaS, Cloud Storage or HTC/HPC Cluster resource may be operated from within the ELIXIR community. ELIXIR is working with EGI in the context of the CC to implement this service area using technologies and know-how from the EGI Federated Cloud solution[[8]](#footnote-8). Priority focus is on cloud provisioning, and this is exactly in the main scope of this document: Providing guidelines for cloud resource providers about how to federate their services into the ELIXIR Compute Platform. Section xxx provides the related guidelines for service providers who are wishing to participate in the provisioning of the basic infrastructure services.
* Integrating Infrastructure Services: providing a federating structure that ensures a consistency of operation and behaviour across all resources and services of the ECP. ELIXIR and EGI are working together to implement this service area using technologies and know-how from the EGI Federated Operations solution[[9]](#footnote-9). The use of the GOCDB service registry together with the ELIXIR basic identity environment is the high priority integration activity in the CC. This would enable simple and reliable discovery of the integrated infrastructure services discoverable by life science users and applications. Section xxx of this report summarises the outcome of this work.
* Higher-Level Services: solutions that expand the platform to better serve specific use cases or use case categories. Competition among similar solutions is expected in this area. ELIXIR is working with EGI to bring in solutions into this area. High priority activity in the CC is integrating and using the EGI Virtual Machine/Virtual Appliances Marketplace[[10]](#footnote-10) in the ELIXIR Compute Platform. Section xxx of this report summarises the outcome of this work.

# Integration status and plans

## CSC

CSC provides an IaaS cloud service called cPouta for its customers (https://research.csc.fi/cpouta). This OpenStack based cloud service, funded by the Finnish Ministry of Culture and Education, allows users to launch their own virtual machines to a server environment running in the CSC computing center in Kajaani.

Currently the service contains nearly 3500 computing cores, used by over 200 computing projects. The majority of users are academic researchers working in Finland and requested access is granted by the CSC resource allocation board. There are also some commercial users, who buy the cPouta capacity directly from CSC.

As a member of EGI, CSC is of course interested in being part of the EGI Federated Cloud. However, as cPouta IaaS cloud is a production service that primarily supports Finnish researchers, joining the EGI Federated Cloud should not affect CSC's service level. In particular, EGI requirements of usage policies or maintenance work must have restricting impact on CSC's services.

Note: CSC already provides cloud services to ELIXIR pilots, and to our knowledge they are actively in use.

The current key issues that need to be solved before CSC can formally join the EGI federated cloud are:

1. **Nova / Keystone integrations and compatibility with CSC's set-up**

It seems that these have improved lately, and now use more standard ways of integrating to OpenStack. The EGI's OpenStack integration will need to be tested by CSC to ensure compatibility with CSC's configuration. This also includes testing OpenStack release dependencies. This is to ensure that they work as expected, and that they don't cause unexpected issues with our other authentication and authorisation services. This also includes

testing OpenStack release dependencies.

1. **Architectural design**Previously, the Keystone LCMAPS integration broke CSC's design on stateless API nodes. This is (apparently) fixed, but CSC needs to ensure that there are no similar issues which break CSC's production architecture.
2. **Maintenance burden of supporting services**

The EGI Federated Cloud has supporting services like accounting, monitoring and service discovery. The integration and maintenance burden of these must be evaluated.

1. **User/group mapping issues**

The cPouta cloud uses project based access control where all resources are shared within a project. It appears that for the EGI Federated Cloud, users can only access their own resources. This and other possible differences in policies need to be resolved. The granularity of VO to OpenStack project mapping needs to be resolved.

1. **Quotas, accounting, billing**

The quotas, accounting and billing need to be resolved. How will this be integrated with CSC's billing system? (While the CSC system is primarily funded by the Finnish Ministry, outside access will be billed.)

1. **Timetable**

The production-level integration is not trivial, and this needs to be scheduled into CSC's development work. EGI's expectations need to be clarified. (Again, note that CSC already provides cloud services to an ELIXIR pilot.)

Operating an OpenStack cloud is resource intense. Adding extra integrations to a production OpenStack cloud increase the complexity of maintaining the cloud. To ensure the long term viability of this platform, CSC would like to be sure that it doesn't add undue burden to the maintenance and update process. Even once the above points have been considered, and potentially solved, more could arise.

## EMBL-EBI

What’s your status and plans for federating cloud and storage resources into the ELIXIR Compute Platform? What are the open questions (if any)?

## CESNET

What’s your status and plans for federating cloud and storage resources into the ELIXIR Compute Platform? What are the open questions (if any)?

## CNRS

What’s your status and plans for federating cloud and storage resources into the ELIXIR Compute Platform? What are the open questions (if any)?

## GRNET

GRNET operates Infrastructure as a Service ~okeanos via large datacenters (84 racks, 1200+ servers, 10000 Virtual Machines active, 5 Petabytes of storage). GRNET is also developing Synnefo, the cloud software for ~okeanos. The service was initially conceived and designed with the Greek Research and Academic community in mind, which comprises the natural user base of GRNET. Soon it became evident though that there is a wider potential and that this effort can be exploited in a broader environment. Towards this, ~okeanos has joined the EGI Federated Cloud activities with the aim to enhance its interoperability features and enable the offering of computing resources to the high-productivity federated infrastructure offered by EGI. ~Okeanos is fully integrated with EGI Federated Cloud and is offering the following services:

OCCI via the SNF-OCCI an implementation of the OCCI specification on top of synnefo’s API kamaki

CDMI via the SNF-CDMI an implementation of the CDMI specification on top of synnefo’s API kamaki

Accounting via the SNF-SSM implementation of the SSM accounting mechanism.

VM Image management via SNF-VMCATCHER.

GRNET supports the images required by ELIXIR CC such as Chipster. The plans for the next year is to update snf-occi to support OCCI 1.2 specification and to extend it capabilities.

## SURFsara

What’s your status and plans for federating cloud and storage resources into the ELIXIR Compute Platform? What are the open questions (if any)?

## JetStream

What’s your status and plans for federating cloud and storage resources into the ELIXIR Compute Platform? What are the open questions (if any)?

## Other providers?

Aaaaaaa

# Integration guidelines and tips for service providers

The ELIXIR Compute Platform uses the integration approach and integration technologies from the EGI Federated Cloud to establish the ‘basic infrastructure services’ layer (See section 2 for details). This section provides service deployment and configuration guidelines for cloud providers who are wishing to participate in the ELIXIR Compute Platform ‘basic infrastructure services’ layer. The section begins with a description of the EGI Federated Cloud concept, then provides integration information and tips for OpenStack, OpenNebula and Synnefo cloud service providers[[11]](#footnote-11).

## Generic concepts and installation guidelines EGI

The EGI Federated Cloud integrates public and community clouds into a scalable computing platform for data and/or compute driven applications and services. Its architecture is based on the extension of the Cloud Management Frameworks deployed at the resource centres to provide a set of agreed uniform interfaces to the user communities and to the EGI Core infrastructure federator services. The federation of IaaS Resource Providers in EGI is built upon the extensive autonomy of resource centres in terms of ownership of the exposed resources. EGI does not mandate deploying any particular or specific Cloud Management Framework, providers should deploy the solution that fits best their individual needs whilst ensuring that the offered services implement the required interfaces.

EGI provides the services and technologies to extend the Cloud Management Frameworks (currently integration of OpenNebula, OpenStack and Synnefo are supported) to create federation of clouds. The cloud resource provider installation manual[[12]](#footnote-12) provides all the steps to deploy and configure the software components to support the federation on the supported Cloud Management Frameworks. Whenever possible these software components are designed and developed to not interfere the usual deployment of the cloud services but to use the already existing public interfaces and simply act as a client for those. The following services help to achieve the federation:

* **Federated AAI**, using X.509 proxy certificates and VOMS extensions with information on the VO of users. Integration with the new EGI AAI is currently under development.
* **Accounting**, usage information is collected via a secure messaging infrastructure in a centralised repository and displayed in a web portal where both individual users and communities can monitor their own resource/service usage across the whole federation.
* **Service Registry**, where providers register the different services offered to the federation.
* **Information Discovery**, so users and tools can retrieve a real-time view of the actual capabilities of the infrastructure.
* **VM Image catalogue and replication**. EGI AppDB provides a catalogue of Virtual Machine Images that encapsulate software appliances relevant for a given community. These images are automatically replicated to the local catalogues of the CMFs supporting the community.
* **Availability Monitoring**, to collect availability and reliability statistics about the providers that can be used to monitor SLAs and OLAs agreed with user communities and resource providers.
* **Standard Interfaces for IaaS**. OCCI and CDMI provide an interoperable interface across the different CMF, so users and applications can interact with the services offered with a single API.

Cloud providers joining the Federated Cloud follow EGI procedure to register and certify a Resource Centre (RC)[[13]](#footnote-13), which makes the EGI infrastructure aware of the new resources you offer, and takes care of validating and testing the behaviour of the services. In the context of the registration, the Resource Centre will become part of a Resource Infrastructure such as a National Grid Initiative (NGI), an EIRO, or a multi-country Resource Infrastructure.

## Installation guideline for OpenStack providers EGI

Aaaa

## OpenStack-specific experiences, recommendations, tips EBI

The deployment of EGI Federated Cloud integrator components has started at EMBL-EBI in late 2015. The goal was to federate the EMBL-EBI Openstack Kilo site into the EGI Federated Cloud, using the installation documentations[[14]](#footnote-14) that were available from EGI at that time. The first feedback from this experience was reported to EGI in the second half of February 2016. The feedback pointed out weaknesses in the installation manual and in the quality of some of the software components. The relevant EGI teams have implemented improvements both in the software and in the installation manuals. Based on the new versions EMBL-EBI successfully federated its OpenStack cloud site during April-May 2016. The site is based on OpenStack Kilo at the time of writing.

other info about the experiences of EBI?

## Installation guideline for OpenNebula providers EGI

Aaaa

## OpenNebula-specific experiences, recommendations, tips CESNET

Aaaaaaa

## Installation guideline for Synnefo providers EGI

Aaaa

## Synnefo-specific experiences, recommendations, tips GRNET

Aaaaaaa

## JetStream integration experiences Indiana University

Aaaaaaa

# Report on AAI integration

This subsection provides details about the integration activities that were performed by EGI and the ELIXIR AAI Task Force to integrate the EGI GOCDB service registry, and the EGI Virtual Machine/Virtual Appliance marketplace (AppDB) with the ELIXIR Authentication-Authorisation Infrastructure (AAI). The integration was requested to enable members of the ELIXIR Community to interact with GOCDB and AppDB using their ELIXIR user identities. In GOCDB these ELIXIR members can register and manage the registration of basic infrastructure resources, while in AppDB they can register new Virtual Machine Images, new Virtual Appliances for sharing these with the broader community. The AAI integration was implemented through the recently released EGI AAI proxy service[[15]](#footnote-15).

The below diagram shows the high level architecture of the integrated AAI components.

….

….

Figure X: Integration of AAI components between ELIXIR and EGI

## Integration of ELIXIR AAI with EGI AAI proxy Christos

Aaaaaaa

## Integration of GOCDB with the EGI AAI proxy David

Aaaaaaa

## Integration of AppDB with the EGI AAI proxy Marios

Aaaaaaa

1. Appendix example

1. <http://www.elixir-europe.org/> [↑](#footnote-ref-1)
2. <http://www.egi.eu/> [↑](#footnote-ref-2)
3. <http://www.elixir-europe.org/> [↑](#footnote-ref-3)
4. <http://www.egi.eu/> [↑](#footnote-ref-4)
5. <http://www.eudat.eu/> [↑](#footnote-ref-5)
6. It is worth noting that the MONARC report is now considered outdated as following the initial experience of running the WLCG and the advances in the capability of the international networks, an evolved technical architecture is now being established that is less hierarchical in its data flows. [↑](#footnote-ref-6)
7. ELIXIR AAI – Requirements and Design: <https://docs.google.com/document/d/1CMY1np3GyvPD8LcKvXljXcRO04V2zu3n_Jcg19jgNOw/edit> [↑](#footnote-ref-7)
8. <https://www.egi.eu/solutions/fed-cloud/index.html> [↑](#footnote-ref-8)
9. <https://www.egi.eu/solutions/fed-ops/index.html> [↑](#footnote-ref-9)
10. <https://appdb.egi.eu/browse/cloud> [↑](#footnote-ref-10)
11. OpenStack, OpenNebula and Synnefo are those cloud management frameworks for which integration components already exists. Integrator components can be developed for additional cloud system if needed. [↑](#footnote-ref-11)
12. https://wiki.egi.eu/wiki/MAN10 [↑](#footnote-ref-12)
13. https://wiki.egi.eu/wiki/PROC09\_Resource\_Centre\_Registration\_and\_Certification [↑](#footnote-ref-13)
14. <https://wiki.egi.eu/wiki/Federated_Cloud_resource_providers_support#Join_as_a_Resource_Provider> [↑](#footnote-ref-14)
15. The EGI AAI Proxy service was developed by the JRA1.1 task of EGI-Engage: [https://wiki.egi.eu/wiki/EGI-Engage:WP3#TASK\_JRA1.1\_Authentication\_and\_Authorisation\_Infrastructure](https://wiki.egi.eu/wiki/EGI-Engage%3AWP3#TASK_JRA1.1_Authentication_and_Authorisation_Infrastructure) [↑](#footnote-ref-15)