

**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 |
|  |  |  |  |

**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 8

3.3 CESNET 8

3.4 CNRS 8

3.5 GRNET 8

3.6 SURFsara 8

3.7 JetStream 8

3.8 Other providers? 8

4 Integration guidelines for service providers 9

4.1 Generic concepts and generic guidelines EGI 9

4.2 OpenStack-specific guidelines (with experiences, recommendations, tips?) EBI 9

4.3 OpenNebula-specific guidelines (with experiences, recommendations, tips?) CESNET 9

4.4 Synnefo-specific guidelines (with experiences, recommendations, tips?) GRNET 9

4.5 JetStream-specific guidelines (with experiences, recommendations, tips?) Indiana University 9

5 Some services utilizing or planning to utilize EGI resources 10

5.1 4.1 using EGI Federated Cloud to provide access to Chipster platform 10

5.1.1 4.1.1. Chipster platform 10

5.1.2 4.1.2 Fitting Chipster to EGI Federated Cloud 10

5.1.3 4.1.3 Pros and Cons 11

5.2 4.2 BioShaDock 11

Appendix I. Appendix example 12

**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 scope of this document: Providing guideliens for cloud resource providers about how to federate their services into the ELIXIR Compute Platform.
* 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). These services – particularly the GOCDB service registry – make the federated cloud resources discoverable and usable by life science users.
* 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, primarily in connection to the Federated Cloud, into this area (for example the Virtual Machine/Appliances Marketplace of the EGI Applications Database).

# Integration status and plans

## CSC

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

## 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

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

## 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 for service providers

## Generic concepts and generic 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[[10]](#footnote-10) 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)[[11]](#footnote-11), 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.

## OpenStack-specific guidelines (with experiences, recommendations, tips?) EBI

Aaaaaaa

## OpenNebula-specific guidelines (with experiences, recommendations, tips?) CESNET

Aaaaaaa

## Synnefo-specific guidelines (with experiences, recommendations, tips?) GRNET

Aaaaaaa

## JetStream-specific guidelines (with experiences, recommendations, tips?) Indiana University

Aaaaaaa

# Some services utilizing or planning to utilize EGI resources

## 4.1 using EGI Federated Cloud to provide access to Chipster platform

### 4.1.1. Chipster platform

Chipster is an easy to use data analysis platform for bioinformatics. It provides an uniform graphical interface for over 350 commonly used bioinformatics tools including several R/Bioconductor based tools and standalone programs (e.g. BWA, TopHat).

Chipster is based on a client-server system where the user runs locally a Chipster-client that submits analysis tasks to a Chipster server. Even though Chipster is an open source tool, there is no public Chipster server that would be open for everybody. Due to that, a researcher needs to have an access to some of the existing Chipster servers to be able to use this platform. Alternatively, researcher can set up your his own Chipster server.

Chipster is available as a Virtual machine image and thus utilizing this VM in EGI Federated Cloud could provide an easy to use solution for a user willing to set up her own Chipster server.

### 4.1.2 Fitting Chipster to EGI Federated Cloud

In order to use the publicly available Chipster VM in EGI federated cloud, minor changes had to done to the VM image (including. XXXX and XXXX). The required modifications were applied to the VM building process and they are now permanently included to the VM building process.

The most problematic feature in setting up a Chipster server, is setting up the bioinfomatics tools, used by the Chipster server. This so called “tools directory” consists of large number of bioinformatics applications, R modules and reference data sets that together require nearly 200 GB of storage space. To make the Chipster setup in EGI Federated Cloud fast to set up to save disk space, the Chipster server process was tuned so that in stead of downloading the tools set for each VM separately we wanted to set up the tools directory that several Chipster VM:s could use the same tools directory.

A cluster specific read-only NFS mount was tested first, but finally CVMFS based remote mounting of the tools directory was used. The benefit of CVMFS is that same installation can be used in any fedCould server............

### 4.1.3 Pros and Cons

To be added

## 4.2 BioShaDock

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://wiki.egi.eu/wiki/MAN10 [↑](#footnote-ref-10)
11. https://wiki.egi.eu/wiki/PROC09\_Resource\_Centre\_Registration\_and\_Certification [↑](#footnote-ref-11)