

**EGI-Engage**

**ELIXIR Competence Centre**

Life science requirements analysis and driver
use case(s) with implementation roadmap

M6.3

|  |  |
| --- | --- |
| **Date** | Xx xx xxxx |
| **Activity** | SA2 |
| **Lead Partner** | EGI.eu |
| **Document Status** | DRAFT |
| **Document Link** | <https://documents.egi.eu/document/2675>  |

Abstract

The ELIXIR Competence Centre of the EGI-Engage project facilitates collaboration between EGI and ELIXIR service developers and providers by collecting, analysing and comparing life science community needs with EGI technical offerings. The process is driven by scientific use cases that are selected for the 18 month long work. This document is the first milestone of the Competence Centre effort: a selection of those science cases from the ELIXIR community that could benefit from EGI services; an initial analysis of the requirements derived from these science cases; and a roadmap to implement these science cases with the use of EGI services, within the ELIXIR Compute Platform. The document also provides a roadmap for the integration of EGI services into this platform.

**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** | 12/Jan/2016 | ToC with initial text for ELIXIR-CC | G. Sipos / EGI.eu-SZTAKI |
| **v.2** | 31/Jan/2016 | Text added about EGI developments for ECP | G. Sipos / EGI.eu-SZTAKI |
| **...** |  |  |  |
| **v.n** |  |  |  |

**TERMINOLOGY**

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

**Contents**

[1 Introduction 5](#_Toc441960992)

[2 Scientific use cases 6](#_Toc441960993)

[2.1 Marine metagenomics use case - Kimmo 6](#_Toc441960994)

[2.1.1 Introduction 6](#_Toc441960995)

[2.1.2 Scientific use case description 6](#_Toc441960996)

[2.1.3 E-infrastructure requirements 7](#_Toc441960997)

[2.1.4 Impact 7](#_Toc441960998)

[2.2 Phenomenal project use case - Steven, Enol 8](#_Toc441960999)

[2.2.1 Introduction 8](#_Toc441961000)

[2.2.2 Scientific use case description 8](#_Toc441961001)

[2.2.3 E-infrastructure requirements 8](#_Toc441961002)

[2.2.4 Impact 9](#_Toc441961003)

[2.3 cBioPortal replication use case - Miroslav, Michal 10](#_Toc441961004)

[2.3.1 Scientific use case description 10](#_Toc441961005)

[2.3.2 Scientific use case description 10](#_Toc441961006)

[2.3.3 E-infrastructure requirements 10](#_Toc441961007)

[2.3.4 Impact 11](#_Toc441961008)

[2.4 SurfSara use case? - Irene 11](#_Toc441961009)

[3 Implementation roadmap 12](#_Toc441961010)

[3.1 Introduction 12](#_Toc441961011)

[3.2 The ELIXIR Compute Platform – Steven 12](#_Toc441961012)

[3.2.1 EGI services in the ELIXIR Compute Platform – Gergely 12](#_Toc441961013)

[3.2.2 New EGI AAI proxy service 14](#_Toc441961014)

[3.3 Marine metagenomics use case - Kimmo 16](#_Toc441961015)

[3.4 Phenomenal project use case – Steven, Enol 16](#_Toc441961016)

[3.5 cBioPortal replication use case - Miroslav, Michal 16](#_Toc441961017)

[3.6 SurfSara use case - Irene 16](#_Toc441961018)

**Executive summary**

Aaaa

# Introduction

ELIXIR is a pan-European research infrastructure in agreement between 15 European governments to build a sustainable European infrastructure for biological information, supporting life science research and its translation to medicine, agriculture, bioindustries and society.

Life science is a fast moving field. For the EGI services to become relevant and help keep European Life Sciences globally competitive, 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 of the EGI-Engage project evaluates, adopts and promotes technologies and resources from the EGI Federated Cloud towards 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’s Federated Cloud collaboration.
2. Identify life science use cases which could benefit from the EGI Federated Cloud services and could make big impact on ELIXIR and EGI communities. Analyse their e-infrastructure requirements, especially with respect to the use of services and technologies adopted within the EGI Federated Cloud.
3. Implement selected life science use cases based on EGI. Collaborate on the implementation with EGI’s and ELIXIR’s partner e-infrastructures, primarily EUDAT.
4. Evaluate the implementations and disseminate the experiences gained with the use cases and with the EGI services towards ELIXIR, EGI and other relevant communities.

This document is a milestone after stage 2 of this process. The document was written by life science and e-infrastructure experts from ELIXIR and EGI, who were brought together within the competence centre. The document captures scientific use cases, derived requirements and envisaged implementation roadmap based on EGI services. Contributors of the report were:

|  |  |  |
| --- | --- | --- |
| Name | Role in ELIXIR/EGI | Contribution to the report |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Scientific use cases

This section provides information about the use cases that have been identified by the Competence Centre. These use cases represent scientific workflows that can be ported to the EGI Federated Cloud. The search for use cases was restricted to workflows that require only ‘non-sensitive’ data, because this simplifies complexity, and also make the ELIXIR Competence Centre effort complementary to the BBMRI Competence Centre activities (task SA6.4 of EGI-Engage), where the focus is on handling sensitive data with EGI services. Each of the use cases are described from three perspectives:

1. Scientific
2. E-infrastructure
3. Impact

These aspects together provide a comprehensive view on the use cases and help the Competence Centre focus its limited effort on those cases that would offer the best value vs. implementation and operational cost.

## Marine metagenomics use case - Kimmo

### Introduction

Aaaaa

### Scientific use case description

|  |  |  |
| --- | --- | --- |
| User Story | *Describe here your use case in term of user story:** *Step1: …*
* *Step2: …*
* *…*
 |  |
| (Potential) User base | *Describe the current and potential user base in term of size (number of users), type (from which scientific domain?), characteristics (academic, private company, students, etc…).**Describe how the potential user base could be caught.* |  |

### E-infrastructure requirements

|  |  |  |
| --- | --- | --- |
| HW Resources | *List the amount of HTC and/or cloud resource you need.**List any other HW requirements.* |  |
| SW Resources | *Describe the external SW products you need and the related licenses.**List the technology providers for each of the above SW product.* |  |
| Cost of delivery | *Describe what needs to be developed/integrated and how much this would cost in terms of effort (estimated PMs)* |  |
| Operational aspect | *Describe who would operate the services related to this use case.**Assess the effort needed to maintain these services up & running beyond the lifetime of the CC.**List the NGIs interested on supporting these services.* |  |

### Impact

|  |  |  |
| --- | --- | --- |
| Business plan 1 | *Define one or more business plan taking into the account the (potential) user base, the cost to delivery and maintain the service.**Identify the full market potential (all user types / categories and size) and most likely uptake possible.**Identify what revenue types will the provider obtain in return for the investment described above and possible estimates e.g. direct payment(s); funding; in-kind contribution**What are the organisational, technical, financial, market and/or legal risks associated to the service provider e.g. inability to scale to demand* |  |
| Business plan 2 | *…* |  |

## Phenomenal project use case - Steven, Enol

### Introduction

Aaaaaa

### Scientific use case description

|  |  |  |
| --- | --- | --- |
| User Story | *Describe here your use case in term of user story:** *Step1: …*
* *Step2: …*
* *…*
 |  |
| (Potential) User base | *Describe the current and potential user base in term of size (number of users), type (from which scientific domain?), characteristics (academic, private company, students, etc…).**Describe how the potential user base could be caught.* |  |

### E-infrastructure requirements

|  |  |  |
| --- | --- | --- |
| HW Resources | *List the amount of HTC and/or cloud resource you need.**List any other HW requirements.* |  |
| SW Resources | *Describe the external SW products you need and the related licenses.**List the technology providers for each of the above SW product.* |  |
| Cost of delivery | *Describe what needs to be developed/integrated and how much this would cost in terms of effort (estimated PMs)* |  |
| Operational aspect | *Describe who would operate the services related to this use case.**Assess the effort needed to maintain these services up & running beyond the lifetime of the CC.**List the NGIs interested on supporting these services.* |  |

### Impact

|  |  |  |
| --- | --- | --- |
| Business plan 1 | *Define one or more business plan taking into the account the (potential) user base, the cost to delivery and maintain the service.**Identify the full market potential (all user types / categories and size) and most likely uptake possible.**Identify what revenue types will the provider obtain in return for the investment described above and possible estimates e.g. direct payment(s); funding; in-kind contribution**What are the organisational, technical, financial, market and/or legal risks associated to the service provider e.g. inability to scale to demand* |  |
| Business plan 2 | *…* |  |

## cBioPortal replication use case - Miroslav, Michal

### Scientific use case description

Aaaaaa

### Scientific use case description

|  |  |  |
| --- | --- | --- |
| User Story | *Describe here your use case in term of user story:** *Step1: …*
* *Step2: …*
* *…*
 |  |
| (Potential) User base | *Describe the current and potential user base in term of size (number of users), type (from which scientific domain?), characteristics (academic, private company, students, etc…).**Describe how the potential user base could be caught.* |  |

### E-infrastructure requirements

|  |  |  |
| --- | --- | --- |
| HW Resources | *List the amount of HTC and/or cloud resource you need.**List any other HW requirements.* |  |
| SW Resources | *Describe the external SW products you need and the related licenses.**List the technology providers for each of the above SW product.* |  |
| Cost of delivery | *Describe what needs to be developed/integrated and how much this would cost in terms of effort (estimated PMs)* |  |
| Operational aspect | *Describe who would operate the services related to this use case.**Assess the effort needed to maintain these services up & running beyond the lifetime of the CC.**List the NGIs interested on supporting these services.* |  |

### Impact

|  |  |  |
| --- | --- | --- |
| Business plan 1 | *Define one or more business plan taking into the account the (potential) user base, the cost to delivery and maintain the service.**Identify the full market potential (all user types / categories and size) and most likely uptake possible.**Identify what revenue types will the provider obtain in return for the investment described above and possible estimates e.g. direct payment(s); funding; in-kind contribution**What are the organisational, technical, financial, market and/or legal risks associated to the service provider e.g. inability to scale to demand* |  |
| Business plan 2 | *…* |  |

## SurfSara use case? - Irene

# Implementation roadmap

## Introduction

During 2015 the ELIXIR community – in collaboration with various e-infrastructures and other service providers – initiated the development of a 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 as well as use cases from other biomedical sciences Research Infrastructures. Because of the fundamental role of the ECP within ELIXIR, EGI-Engage Competence Centre use cases should also consider the ELIXIR ECP as a foundational infrastructure, and support EGI’s effort in integrating EGI services into this platform. The next subsections describe the architecture of the ECP, the ongoing activities in integrating EGI services into the ECP, and the implementation roadmap for the previously described use cases on top of the ECP.

## The ELIXIR Compute Platform – Steven

The capabilities and functional building blocks of the ECP have been discussed and identified in 2015 through a series of workshops and online consultation meetings. ….

### EGI services in the ELIXIR Compute Platform – Gergely

This section provides details about the development activities that are required and are ongoing/planned in EGI to customise services for inclusion in the ECP. These requirements are based on the EGI-related recommendations of the ‘The ELIXIR Compute Platform: A Reference Technical Services Architecture for supporting Life Science Research’ document, and additional recommendations that emerged for EGI from already ongoing ECP-related developments.

There are three key aspects of the ECP that influence the EGI service developments:

1. The ECP will include a federated cloud Virtual Organisation from EGI. This VO together with the included cloud sites and Virtual Machine images will be maintained by representatives of the ELIXIR community (VO managers, site managers, VM developers).
2. ELIXIR operates an Identity Provider (IdP) at CESNET. Every ELIXIR user will have a user account and user attributes at this IdP. ECP services – including those contributed by EGI – must be able to authenticate and authorise users by their ELIXIR accounts and attributes.
3. The ELIXIR IdP provides only the hashed eppn for SPs that are not committing to the GEANT Code of Conduct (CoCo)[[1]](#footnote-1). The ELIXIR IDP provides the following set of attributes for Service Providers that are committing to the CoCo:
	1. eduPersonPrincipalName - ELIXIR ID
	2. displayName
	3. email

The following table provides a listing of the use case requirements that were derived for EGI services from the above three points, and indicates the status of ongoing developments. The table is an updated version of the ‘ELIXIR Compute Platform Timeline of EGI developments’ document[[2]](#footnote-2).

|  |  |  |  |
| --- | --- | --- | --- |
| **EGI service** | **Requirement** | **Priority** | **Status of development** |
| VO Membership Management tool (PERUN) | ELIXIR members should be able to join the ELIXIR VO with their ELIXIR ID. | High | Completed and deployment is in place for the vo.elixir-europe.eu VO. |
| Applications Database Virtual Machine Marketplace (AppDB) | ELIXIR users should be able to login to AppDB with ELIXIR IDs. AppDB should recognise ELIXIR VM admins and allow them to add manage the ELIXIR VO VM image list. | High | Direct integration with the ELIXIR IdP is completed and deployed in the AppDB development instance[[3]](#footnote-3). The setup will be reassessed in 2016 Q1, when the EGI AAI proxy[[4]](#footnote-4) will come online.  |
| Operations Portal | ELIXIR cloud providers and VO admins should be able to login to the Operations Portal with their ELIXIR identity to send service downtime broadcast messages to members of the ELIXIR VO.  | High | Integration of the Operations Portal with the ELIXIR IDP will be achieved through the EGI AAI proxy in 2016 Q2. |
| Service monitoring (ARGO) | Create a site listing feature under 'Site status reports'. (Currently it's only a search function so the user must know what to look for.) | Medium | Requires only service configuration. This will be done after there is at least one ELIXIR cloud site in the ELIXIR VO. (EBI cloud site is currently under setup in the VO.) |
| Add the first set of ELIXIR sites to the site list: EBI, CSC, CESNET, SURFsara. |
| In the Availabilities/Reliabilities menu introduce a new subcategory: ELIXIR REPORT[[5]](#footnote-5). The ELIXIR sites should be listed in this subcategory. |
| Service registry (GOCDB) | ELIXIR cloud providers should be able to login to GOCDB to register/update information about their site. | Low | Direct integration with the ELIXIR IdP is completed and deployed in the GOCDB development instance[[6]](#footnote-6). The setup will be reassessed in 2016 Q1, when the EGI AAI proxy will come online. |
| Per-user subproxy system (PUSP) of the EGI long-tail of science platform | When the PSUP system with some gateways becomes fully operational in the long-tail platform, then give a demo to ELIXIR about this.  | Low | The PSUP within the long-tail platform is delivered at the end of February 2016 as D4. By EGI-Engage. A demo van be given at one of the upcoming ELIXIR CC teleconferences. |
| Accounting system (APEL) | The EXCELERATE project will define the ELIXIR Metrics database and portal. APEL should gather and send accounting data from the ELIXIR sites to this ELIXIR metrics database.  | No | Development can start approx. in spring 2016, when the ELIXIR metrics database and portal are specified.  |

### New EGI AAI

Several European Research Infrastructures recently decided to operate their own AAI (Authentication and Authorisation Infrastructure) to provide user identity and user attributes for their community members. These RI-specific AAIs simplify and harmonise access to online services across institutional and national borders. ELIXIR recently also established its own AAI, hosted at CESNET from late 2015.

The trend of dedicated RI AAIs made EGI reassess its own AAI architecture. The community concluded that it needs to evolve its own AAI architecture to allow coordinated linkage of EGI services with externally operated RI AAIs. The EGI community – within the JRA1.1 task[[7]](#footnote-7) of the EGI-Engage project – started the design and development of the new EGI AAI in March 2015. The work aimed at a system that

1. simplifies the process of connecting EGI services (e.g. AppDB, Operations Portal, GOCDB, etc.) with AAI architectures operated by external infrastructures, such as the ELIXIR.

AND

1. harmonises the integration of EGI services across multiple, externally operated RI AAIs. (e.g. AppDB would be connected to the ELIXIR AAI, the DARIAH AAI, the EPOS AAI in a harmonised way).

The design of the new EGI AAI system has finished in 2015, and the first development release will be available for early adopter use cases in XX/XX. In the heart of the new AAI there is an ‘IdP/SP Proxy’ component based on SAML technology (See Figure 1). This component acts as a Service Provider (SP) for the supported identity federation (e.g. the ELIXIR IdP), while at the same time, it will act as an Identity Provider (IdP) for the EGI services (e.g. AppDB, Operations Portal, GOCDB, etc.). The IdP/SP Proxy will be responsible for mapping an external user identity to an ‘EGI identifier’ which will be used for the same user across all the EGI services. The IdP/SP Proxy will be able to import attributes from external attribute authorities (e.g. from ELIXIR IdP) and assign these to the internal EGI user identifier. Based on the imported attributes the EGI services can authorise users across the whole EGI network in a coherent way. (e.g. an ELIXIR site manager will be recognised in both GOCDB and Operations Portal).



Figure . Architecture of the new EGI AAI.

During 2015 both AppDB and GOCDB were directly connected to the ELIXIR IdP. As soon as the new EGI AAI becomes available these connections will be reassessed and should be changed to connections through the IdP/SP Proxy. The benefit of this would be a coherent integration among EGI and ELIXIR AAI which would lower the cost of maintenance and further development of the collaborative ELIXIR – EGI setup. The Operations Portal have not been connected to the ELIXIR AAI yet. The connection will be realised in xxxx through the IdP/SP Proxy.

## Marine metagenomics use case - Kimmo

Aaaaaa

## Phenomenal project use case – Steven, Enol

Aaaaaa

## cBioPortal replication use case - Miroslav, Michal

Aaaaaa

## SurfSara use case - Irene

Aaaaaa

1. <https://wiki.edugain.org/Data_Protection_Code_of_Conduct_Cookbook> [↑](#footnote-ref-1)
2. <https://docs.google.com/document/d/1J3XPAvX0jVhJ_pFex5gXWYBazRweKNKbqRfTUJF56M0/edit#heading=h.idg60lt9pvvi> [↑](#footnote-ref-2)
3. <https://appdb-dev.marie.hellasgrid.gr> [↑](#footnote-ref-3)
4. The EGI AAI proxy will provide harmonised AAI integration facility between EGI services and community-specific AAI systems, such as the ELIXIR IdP. [↑](#footnote-ref-4)
5. ELIXIR would like to have its own identity on the portal even if it reuses sites from EGI and EUDAT at the infrastructure level. [↑](#footnote-ref-5)
6. <David to add> [↑](#footnote-ref-6)
7. [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-7)