EGI-Engage

Infrastructure tests and best usage practices for life science service providers

D6.10

Date	28 Jun 2016
Activity	WP6
Lead Partner	CSC
Document Status	FINAL
Document Link	https://documents.egi.eu/document/2802

Abstract

The ELIXIR Competence Centre (CC) aims to bring EGI resources, especially the EGI Federated Cloud, to be available to the ELIXIR user community. The document briefly describes ELIXIR's plans for using services and technologies from the EGI Federated Cloud in the ELIXIR Compute Platform. The document provides guidelines and usage practices for cloud providers on how to join the EGI Federated Cloud as a service provider. The document also sums up the experiences and plans of the ELIXIR CC members regarding their participation in the ELIXIR Compute Platform.

This material by Parties of the EGI-Engage Consortium is licensed under a Creative Commons Attribution 4.0 International License.

The EGI-Engage project is co-funded by the European Union (EU) Horizon 2020 program under Grant number 654142 http://go.egi.eu/eng

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:	Kimmo Mattila	CSC / SA2	3/Jun/2016
Reviewed by	Lars Ailo Bongo Vera Yvonne Hansper Tiziana Ferrari	Uni. of Tromso / ELIXIR Compute group CSC / ELIXIR Compute group EGI.eu / EGI-Engage project director	10/Jun/2016 11/Jun/2016 27/Jun/2016
Approved by:	AMB and PMB		29/Jun/2016

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.6-19	13/May/2016	Adding content to various sections (See list of contributors in next section)	K. Mattila / CSC
v.20	30/May/2016	Merged CNRS contribution, cleaned CSC section, wrote Summary and Next steps	G. Sipos / EGI.eu
v.21	3/June/2016	Added JetStream section; Updated CSC section	K. Mattila / CSC
v.22	3/Jun/2016	Finalised for external review	G. Sipos / EGI.eu
v.23	15/Jun/2016	Updated based on external reviewers' feedback	K. Mattila / CSC G. Sipos / EGI.eu
v.24	28/Jun/2016	Updated based on project director's feedback	G. Sipos / EGI.eu

TERMINOLOGY

A complete project glossary is provided at the following page: https://wiki.egi.eu/wiki/Glossary

Contents

List of contributors

Section	Contributor
The ELIXIR Compute Platform; Role of service providers	Steven Newhouse (EMBL-EBI) Gergely Sipos (EGI.eu)
Integration guidelines for service providers	Vincenzo Spinoso (EGI.eu) Enol Fernandez (EGI.eu, OpenStack) Boris Parak (CESNET, OpenNebula)
Integration status and plans	CSC: Kimmo Mattila, Vera Hansper CESNET: CNRS: Christophe Blanchet EMBL-EBI: Steven Newhouse, Charles Short GRNET: Kostas Koumantaros JetStream: Rob Quick (Uni. of Indiana)
Report on AAI integration	Christos Kanellopoulos (GRNET) Peter Solagna (EGI.eu) Marios Chatziangelou (IASA, AppDB) David Meredith (STFC, GOCDB)

Executive summary

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 the responsibilities of participating service providers: In the near future providers are expected to operate Infrastructure as a Service (IaaS) cloud resources, and in the mid-term complement these with domain or application-specific services (such as GA4GH Beacon service; Certain datasets).
(2) Presents integration guidelines for cloud providers who are wishing to participate in this infrastructure. OpenStack, OpenNebula or Synnefo based IaaS systems are supported at the moment.
(3) Provides status update about the completed, ongoing and planned integration of cloud resources into the platform by CC members: CESNET, EMBL-EBI and GRNET have cloud sites with the required federation capabilities and are ready to join the ELIXIR Compute Platform. Further discussions and technology/policy analysis are needed with CSC, CNRS and SURFsara to decide about joining the ELIXIR Compute Platform.

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.

The core elements of the platform are now available. Two cloud sites (CESNET and GRNET) are already fully integrated into it; one site is in the piloting phase (EMBL-EBI); one is in preparation (JetStream) and one is considering joining (CSC). Instructions for integrating new resources (OpenStack, OpenNebula, Synnefo) to EGI Federated Cloud are summed up and the experiences from the integration process are discussed.

Extra attention is given in the document to the recent work related to the integration of the EGI and ELIXIR Authentication and Authorization Infrastructures (AAI). The integrated AAI will play a key role in the seamless access of EGI services by life science communities.

Based on the cloud and AAI integration achievements that are documented in this document, the consortium – in partnership with the ELIXIR-EXCELERATE project – will move ahead with broadening the scale and scope of the ELIXIR Compute Platform by adding additional cloud sites and new types of services; integrate the use cases that have been brought together by the CC members in M6.3[1], and will define the rules and policies for accessing the deployed capabilities and for monitoring this use and the impact it generates.

[1] https://documents.egi.eu/public/ShowDocument?docid=2675

1 Introduction

ELIXIR[2] 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[3] 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. Exchange information to make these communities better aware of current and emerging services and needs.
2. Identify life science use cases that could benefit from EGI services and could make big impact on the ELIXIR and EGI communities. Analyse the e-infrastructure requirements of the use cases. (These have been documented in M6.3 document[4])
3. Implement the use cases as demonstrators based on EGI's e-infrastructure services. Collaborate during implementation with relevant EGI and ELIXIR partners, such as EUDAT[5] to establish a generic infrastructure, the 'ELIXIR Compute Platform', that can underpin demonstrators and production services 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.

[2] http://www.elixir-europe.org/
[3] http://www.egi.eu/
[4] https://documents.egi.eu/document/2675
[5] http://www.eudat.eu/

2 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 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] that formed the basis of the Tiered model that was initially adopted by the 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 will 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 based analysis services 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 available 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, the role of ECP is to define a minimal 'neck' of an hourglass that ELIXIR Researchers and

[6] It is worth noting that the MONARC report is now considered out-dated 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.

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] for 'ELIXIR infrastructure services' (all other services). ELIXIR has been working with EGI to connect the ELIXIR AAI and EGI AAI services to offer seamless access to EGI services for life science users. The basic ELIXIR AAI environment has been available since the end of 2015 and further developments and refinements are coming during 2016. Section 5.1 provides a summary of this work.
- 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 resources 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]. 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 3 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]. The use of the GOCDB service registry together (through the ELIXIR AAI) is the high priority integration activity. This would enable simple and reliable discovery of the integrated infrastructure services discoverable by life science users and applications. Section 5.2 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 these high level services. 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] of AppDB in the ELIXIR Compute Platform. Section 5.3 of this report summarises the outcome of this work.

[7] ELIXIR AAI – Requirements and Design: https://docs.google.com/document/d/1CMY1np3GyvPD8LcKvXljXcRO04V2zu3n_Jcg19jgNOw/edit
[8] https://www.egi.eu/solutions/fed-cloud/index.html
[9] https://www.egi.eu/solutions/fed-ops/index.html
[10] https://appdb.egi.eu/browse/cloud

3 Integration guidelines for service providers

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

3.1 Generic concepts

The EGI Federated Cloud integrates public and community clouds into a scalable computing platform for data and/or compute driven applications and services. Each cloud can provide 'Infrastructure as a Service' features for users, i.e. CPU/GPU computing, storage and network configuration. The architecture of the EGI Federated Cloud is based on the extension of the Cloud Management Frameworks deployed at the resource centres to provide a set of agreed uniform interfaces towards users and towards infrastructure operators. The federation allows resource centers to keep their autonomy in terms of ownership of the exposed cloud resources. EGI/ELIXIR 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. Connectors currently exist to federate OpenNebula, OpenStack and Synnefo clouds. Connectors to federate other types of clouds (e.g. StratusLab) can be developed as a joint effort of EGI and ELIXIR (or other communities). These clouds can be accessed in the federation through OCCI, OpenStack Nova interfaces, and through high level tools and orchestrators that are compatible with these (e.g. Occopus, Terraform, etc.)

There are several 'Virtual Organisations' (VOs) formed within the EGI Federated Cloud. Each VO of the cloud federation is a resource allocation for a specific community (for example ELIXIR). The cloud providers who support a given community join the VO dedicated to that community. Users and their applications can consume cloud resources from the VO after becoming members of the same VO. Within a VO a community can create multiple membership groups and allocate different privilege levels and/or resource quotas to these. Some communities go even further and setup multiple VOs, typically one for each application area, or sub-discipline within the main community discipline. Having multiple VOs allow stronger user isolation, and richer statistics on infrastructure usage per application area.

[11] OpenStack, OpenNebula and Synnefo are those cloud management frameworks for which integration components already exists in EGI. Integrator components can be developed for additional cloud system if needed.

The ELIXIR community recently established its first VO in the EGI Federated cloud. The VO is called vo.elixir-europe.eu[12]. The exact usage policies of this VO (for life science researchers and use cases) are yet to be defined by ELIXIR, mainly in WP4 of the EXCELERATE project. ELIXIR will also consider the setup of additional VOs, e.g. one for each of the four scientific application area that the EXCELERATE project is focussed one.

The EGI cloud resource provider installation manual[13] provides all the steps to deploy and configure the software components to support the federation on the supported Cloud Management Frameworks (CMF), and to join VOs. Whenever possible these software components are designed and developed to not interfere with 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 (e.g. the ELIXIR VO members). Integration with the new EGI AAI is currently under development (for OpenStack, OpenNebula, Synnefo clouds).
- **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**. The 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 procedures to register and certify a Resource Centre (RC)[14], which makes the EGI infrastructure aware of the resources that the new provider offers. RA also 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. ELIXIR is

[12] The 'ID card' of the ELIXIR VO in the EGI Operations Portal: http://operations-portal.egi.eu/vo/view/voname/vo.elixir-europe.org
[13] https://wiki.egi.eu/wiki/MAN10
[14] https://wiki.egi.eu/wiki/PROC09_Resource_Centre_Registration_and_Certification

currently reviewing these certification procedures to understand which elements are needed to the support ELIXIR Compute Platform.

Cloud Resource Provider Installation Manual currently exist for OpenStack and OpenNebula cloud providers at https://wiki.egi.eu/wiki/MAN10. At the time of writing these guides are known to work with OpenStack Havana, Icehouse, Juno, Kilo and Liberty releases, and with OpenNebula v4.4.x, v4.6.x, v4.8.x, v4.10.x, v4.12.x, v4.14.x. One cloud site (GRNET), based on the Synnefo Cloud Management Framework is also operating in the EGI Federated Cloud, and if requested the operators can help other sites to deploy this technology.

4 Integration status and plans

This section provides status information about the sites that participate in the ELIXIR Competence Centre. The section captures the current status of these sites, and plans for providing cloud services in the ELIXIR Compute Platform. The section covers:

- CSC from Finland
- CESNET from Czech Republic
- CNRS from France
- EMBL-EBI from UK
- GRNET from Greece
- JetStream from the University of Indiana, USA

4.1 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. Access is granted by the CSC resource allocation board, but cPouta also serves international collaboration projects, including some ELIXIR use cases. There are also some commercial users, who buy the cPouta capacity directly from CSC.

As a member of EGI, CSC is interested in being part of the EGI Federated Cloud to offer resources for ELIXIR and possibly also for other communities. However, as cPouta IaaS cloud is a production service that primarily supports Finnish researchers, offering this to international communities through the EGI Federated cloud should not affect CSC's service level. In particular, ELIXIR and EGI usage policies must not have a restricting impact on providing the CSC’s core services to national users.

The recent development in the EGI Federated Cloud environment should make it easier for CSC to start acting as a resource provider for the EGI federated cloud. For example, features that previous caused Nova/Keystone compatibility issues are no longer used in the EGI Federated Cloud. However, to utilize these improvements CSC will first need to upgrade the local OpenStack several version steps (Kilo or Liberty) which will require several months to be implemented.

After this upgrade both the technical and policy issues related to quotas, accounting and billing integration will need to be readdressed and solved. The final decision on the actual integration has yet to be made, and therefore there are no commitments on the timetable of the technical

integration. However, if the decision is positive, it should be technically possible to start the integration work after these changes have been made.

Operating an OpenStack cloud is already a resource intensive task, it is important that the instruments allowing the federated operations of distributed OpenStack clouds are easy to deploy and operate. The current cloud federated integration approaches are being tested against the requirements from CSC with the aim of further improving them.

4.2 CESNET

CESNET operates a sizable national HPC infrastructure in the Czech Republic providing resources to local academic communities, including cloud-based resources, various types of storage, and identity management services. On an international level, CESNET is a member of EGI (NGI_CZ), as a resource provider and a technology provider, specifically in the context of the EGI Federated Cloud.

As a resource provider, CESNET is fully integrated and offers the following cloud services compliant with the EGI Federated Cloud:

- Virtual machine management via OCCI
- Accounting via APEL/SSM
- Information discovery via BDII
- Virtual machine image management via HEPiX vmcatcher/vmcaster
- AAI

As a technology provider, CESNET develops and maintains the following EGI Federated Cloud integration components and tools:

- OCCI components for virtual machine management (rOCCI-{core, api, cli, server})
- OpenNebula APEL connector (oneacct-export)
- OpenNebula Perun connector (fctf-perun)
- OpenNebula vmcatcher handler (itchy, nifty)
- OCCI monitoring probes for Nagios

CESNET can support the appliances and virtual organizations required by the ELIXIR Compute platform. (Currently there is one ELIXIR virtual organisation, but more are expected in the future.)

4.3 CNRS

The French Institute of Bioinformatics (IFB) is a national service infrastructure in bioinformatics and the French node of ELIXIR. The IFB's principal mission is to provide core bioinformatics resources to the French life science community (academic and private partners) coupled to the required computing and storage capacity in a national bioinformatics cloud. To address the most

common needs, a selection of major scientific software tools was made and they were installed in pre-configured virtual images (cloud appliances), ready to run on the IFB's cloud. To date, 36 appliances cover different domains of the life sciences, for example users can deploy a cloud virtual pipeline for microbial genomes analysis. The long-term objective is to create a federation of clouds that rely on the interconnected IT infrastructures of the IFB's platforms, providing personalized services to analyze life science data.

Currently, the IFB cloud is based on the StratusLab cloud middleware and comprises 200 computing cores, two terabytes (TB) of RAM and 50 TB of storage. We plan to increase these resources to 5,000 cores with 500 TB of storage by the end of 2016. The IFB's IT infrastructure is hosted at IDRIS, one of the major national high-performance computing centers (GENCI). This ensures the stability of the IFB's infrastructure by providing a reliable power supply, cooling system and large network bandwidth commonly found in such environments.

CNRS IFB cloud cannot be yet easily integrated with the EGI Federated Cloud technologies that are proposed for use in the ELIXIR Compute Platform. This is due the facts that (i) it is in production status for the French life science community, and (ii) it is currently relying on StratusLab middleware which does not currently have connectors to participate the EGI FedCloud.

A way of providing resources to the CC-ELIXIR may be through

1. Those French NGI sites offer resources in the ELIXIR Compute Platform that are already participate in the EGI Federated Cloud (therefore have the necessary technologies already deployed).
2. If the development effort is justified, then develop the necessary connectors for StratusLab to participate in the EGI technology-based ELIXIR Compute Platform by those French sites that are affiliated with ELIXIR-FR and run StratusLab.

4.4 EMBL-EBI

The EMBL-EBI Embassy Cloud (http://www.embassycloud.org/) is a 1200 core OpenStack platform co-located with EMBL-EBI's services and data resources. Access to the Embassy Cloud is available for researchers outside EMBL if they have collaboration with staff at EMBL.

This resource is now being integrated into the ELIXIR Operations Centre within EGI's Grid Operations Centre Database (https://goc.egi.eu), which is now also integrated with the ELIXIR AAI through the EGI AAI gateway (Section 5.2). This work by EGI has allowed the EMBL-EBI site to appear as part of the ELIXIR Operations Centre. Other ELIXIR Compute Platform cloud sites that are being integrated using the EGI model can also be added into this domain if they do not wish or are not able to benefit from the operational services of their national e-Infrastructure.

ELIXIR cloud sites that are already part of an NGI can be given the 'elixir' scope. This allows these NGI sites, and the sites within the ELIXIR Operations Centre to be identified and to be extracted to appear in other EGI tools. This integration is now taking place by EGI within the site monitoring tool ARGO – http://argo.egi.eu.

Work on the integration of the EMBL-EBI site completely into the EGI Federated Cloud continues with the integration of accounting, information services, and with AppDB for virtual machine integration continuing.

There are however, current security concerns which include:

1. Nova API v2.1 (which is at the time of writing the default in OpenStack Liberty), no longer allows granular authorization within a project. So in effect users within the same project are not protected from each other. Requests to change this behaviour have been sent to the OpenStack Nova project (not only from EGI, but from various other communities and companies)[15]. The code changes required for the 'original behaviour' are now specified[16], and are waiting for implementation and integration, which is expected to happen in one of the upcoming Nova releases. (Next Nova releases are expected in October 2016 and April 2017.) In the meantime ELIXIR should plan and use user access policies that are compatible with the current Nova behaviour.
2. The lack of CRL (Certificate Revocation List) checking by VOMS library/Apache in OpenStack is a concern as revoked user certificates will still be allowed to access the cloud resource. The fix for this problem is now under finalisation in EGI, and is expected to be made available during the summer of 2016.

4.5 GRNET

GRNET operates Infrastructure as a Service ~okeanos via large datacenters (84 racks, 1200+ servers, 10000 Virtual Machines active, 5 PB 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 virtual organisations and virtual machine images required by the ELIXIR Competence Centre. The plans for the next year are to update snf-occi to support OCCI 1.2 specification and to extend it capabilities.

[15] https://bugs.launchpad.net/nova/+bug/1539351
[16] https://review.openstack.org/#/c/324068/

4.6 Jetstream

Jetstream, led by the Indiana University Pervasive Technology Institute (PTI), will add cloud-based computation to the US national cyberinfrastructure[17]. Researchers will be able to create virtual machines on the remote resource that look and feel like their lab workstation or home machine, but are able to harness thousands of times the computing power. Jetstream will provide the following core capabilities:

1. Use Virtual Machines interactively
2. Researchers and students can move data to and from Jetstream using Globus Transfer
3. Use virtual desktops.
4. Publish VMs with a DOI.

Jetstream will be attractive to communities who have not been users of traditional HPC systems, but who would benefit from advanced computational capabilities. Among those groups are researchers not only in biology, but also atmospheric science, observational astronomy, and the social sciences.

The Jetstream system received official acceptance after review by the National Science Foundation at the beginning of May 2016. Since this was announced an ELIXIR start up allocation has been approved by XSEDE. An initial meeting with the Jetstream administrators was conducted on May 18th to discuss the goals of the Jetstream interoperability project and the use of the EGI FedCloud integration appliance. While there were some policy concerns surrounding the XSEDE allocation process, no immediate technical concerns were expressed by the administrators. An initial attempt to run the EGI FedCloud integration appliance will happen during the summer with results reported to the ELIXIR Competence Centre in September. Once the integration appliance has successfully been installed ELIXIR application testing can occur.

[17] http://jetstream-cloud.org/index.php

5 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 EGI services with the ELIXIR Authentication-Authorisation Infrastructure (AAI). The AAI integration was implemented through the recently released EGI AAI proxy service[18]. The GOCDB service registry, and the AppDB Virtual Machine Image catalogue were selected as priority services to conduct the integration. Besides these OpenStack was also integrated with the EGI AAI proxy.

These integrations enable members of the ELIXIR Community to interact with GOCDB, AppDB and OpenStack resources using their ELIXIR user identities:

- In GOCDB these ELIXIR service providers can register and manage the registration of basic infrastructure resources (cloud and storage).
- In AppDB ELIXIR application developers can register Virtual Machine Images and Virtual Appliances for publishing these on the ELIXIR cloud sites and for sharing these with the broader life science community.
- In OpenStack authenticated ELIXIR users can instantiate applications from the pre-deployed Virtual Machine Images and Virtual Appliances.

The next subsections provide details on how the integration was achieved and what are the next steps.

5.1 Integration of the ELIXIR AAI with the EGI AAI proxy

The 'EGI AAI proxy' is a new service in the EGI service portfolio and it enables access to EGI services and resources with federated authentication mechanisms. Specifically, the proxy service is operated as a central hub between federated Identity Providers (IdPs) and the EGI Service Providers (SPs). The main advantage of this design principle is that all entities only need to establish and maintain technical and trust relations to a single entity, the EGI AAI proxy, instead of managing N-to-M relationships. The proxy acts as a Service Provider towards the Identity Providers and as an Identity Provider towards the Service Providers.

Using the EGI AAI proxy, users are able to authenticate with the credentials provided by the IdP of their Home Organisation through eduGAIN, as well as using social identity providers, eGOV IDs, or other selected external identity providers. To achieve this, the EGI AAI has built-in support for SAML, OpenID Connect and OAuth2 providers and already enables user logins through Facebook, Google, LinkedIn, and ORCID. In addition to serving as an authentication proxy, the EGI AAI provides a central Discovery Service (Where Are You From – WAYF) for users to select their preferred IdP.

[18] 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

The EGI AAI proxy is also responsible for aggregating user attributes originating from various authoritative sources (IdPs and attribute provider services) and delivering them to the connected EGI service providers in a harmonised and transparent way. Service Providers can use the received attributes for authorisation purposes, i.e. deciding what the user is allowed to use/do with the service. The EGI services require a minimum set of attributes from the EGI AAI proxy about the user to grant access[19].

Within ELIXIR the ELIXIR AAI service operates as both an IdP and attribute provider service. It's assumed that this service will manage user accounts and personal attributes for every ELIXIR user. During the last months the EGI AAI proxy has been integrated with the ELXIR AAI with the goal to enable members of the ELIXIR Community to access EGI services. Two EGI services – the GOCDB service registry and the AppDB Virtual Machine marketplace – were chosen as priority services for the ELIXIR integration. The integration work was achieved by connecting the SP proxy element of the EGI AAI proxy with the IdP proxy of the ELIXIR AAI by exchanging their SAML metadata in XML format.

The typical Single Sign-On (SSO) flow in the integrated system begins with the ELIXIR user accessing an EGI application (GOCDB or AppDB) through a web browser (SP-initiated SSO). Since the user is not logged in at the SP yet, the SP sends the user to the EGI IdP proxy to authenticate. The user is then redirected to the EGI central Discovery Service page where he/she is able to select to authenticate at ELIXIR. This results in a SAML authentication request from the EGI AAI proxy to the ELIXIR AAI (it's IdP proxy component). Compliance of the EGI SP proxy with the GÉANT Data Protection Code of Conduct (CoCo) and the REFEDS Research and Scholarship (R&S) allows the ELIXIR AAI IdP proxy to release the minimum set of attributes required for a user to make use of the EGI service.

The released attribute set includes an opaque (eduPersonalUniqueId), as well as a username-based (eduPersonPrincipalName) user identifier scoped at elixir-europe.org. The first name (givenName), surname (sn) and email (mail) of the user are also provided. To indicate the set of rights to specific EGI resources, the ELIXIR IdP proxy releases an eduPersonEntitlement attribute. This is a multi-valued attribute, with each value formatted as a URI. The structure of the value itself however, is based on an agreement between ELXIR and EGI, since there are currently no commonly accepted standards or recommendations regarding the representation of entitlements. The agreed eduPersonEntitlement content is documented at https://wiki.egi.eu/wiki/EGI_AAI_integration_with_ELIXIR_AAI.

The ELIXIR IdP builds an assertion containing the attributes above, which is sent to the SP proxy component of the EGI AAI proxy. Based on the authentication method selected by the user, the EGI proxy assigns a Level of Assurance (LoA)[20], which is conveyed to the SP through the

[19] The required attributes are eduPersonalUniqueId and the username-based eduPersonPrincipalName user identifier (scoped at elixir-europe.org in the case of ELIXIR).

[20] LoA expresses the ability to determine, with some level of certainty, that the electronic credential representing an entity (human or a machine) with which it interacts to effect a transaction, can be trusted to actually belong to the entity.

eduPersonAssurance attribute and the Authentication Context Class of the SAML authentication response. EGI AAI currently distinguishes between three LoA levels, namely, Low, Substantial and High. Some EGI SPs have been configured to provide limited access (or not to accept at all) credentials with the Low LoA. Details regarding LoA levels in case of GOCDB and AppDB are reported in the next subsections.

5.2 Integration of GOCDB with the EGI AAI proxy

GOCDB has been integrated with the EGI AAI proxy. This allows users without a client x509 certificate to access GOCDB provided the user authenticates to the AAI proxy using an authentication scheme that provides an adequate level of assurance (LoA). As a central configuration management database, GOCDB requires the highest level of assurance. For example, if the user authenticates to the ProxyIdP using a scheme that provides a low or medium level of assurance such as social-media/Facebook, then access to GOCDB is denied. Conversely, if a project integrates a trusted identity provider with the ProxyIdP where the user authenticates using a scheme that provides a high LoA (e.g. institutional credentials), access is granted. During the login process, the LoA category that is assigned to the user's authenticated session (e.g. low, substantial, high) is communicated to the GOCDB. The ELIXIR AAI IdP provides the highest level of assurance to GOCDB.

Based on the provided assurance level, GOCDB can make its own internal authorisation decisions. If a user is granted access, they use the existing GOCDB role mechanisms to request roles and permissions. There is no automatic mapping between ELIXIR roles and GOCDB roles.

1. User logs in to GOCDB service registry with an ELIXIR account (http://goc.egi.eu)
2. User requests a role. The request is passed to the existing users who already own the necessary roles to approve or reject the role request. Currently Steven Newhouse from EMBL-EBI and Miroslav Ruda from CESNET have roles over the ELIXR group. (Once this initial community admin group is established within GOCDB by the GOCDB admin, the group can subsequently self-manage their user memberships.)
3. User is granted with the approved role at next login, and can perform authorised operations.

5.3 Integration of AppDB with the EGI AAI proxy

The EGI AppDB is planned to be used as a marketplace of Virtual Machine Images (VMIs) within the ELIXIR Compute platform. A user can have three roles when accessing the EGI AppDB marketplace:

- Visitor: Can browse publicly visible VMIs, can download them for local use. Visitors do not have to login.
- VMI developer: Any user with a valid account can register new VMIs and VMI versions in the marketplace and (optionally) can submit these to the community coordinator for inclusion in the community image list.

- Coordinator of a scientific community: Can add VMIs to the community image list to trigger the replication of these VMIs to the cloud sites that support the community. The community image list includes VMs that are of high relevance to the scientific community. Community coordinators have to login to AppDB and must have attributes that express affiliation to a community and coordinator role within that community.

The goal of the AppDB–ELIXIR AAI integration was to enable authorized access to the marketplace for members of the ELIXIR community - i.e. for those possessing ELIXIR accounts.

The integration of AppDB with the EGI ProxyIdP has been completed, so AppDB can login and authorise users coming through the EGI AAI proxy. Authorization of user actions is based on relevant entitlements returned by the EGI AAI proxy. More detailed information about authorization process can be found at https://wiki.egi.eu/wiki/EGI_AAI_integration_with_ELIXIR_AAI.

Since early June these new developments are available in the AppDB production instance (http://appdb.egi.eu) so ELIXIR users can login into the system with their ELIXIR accounts. Their role (if any) is recognized by AppDB automatically and they are granted with the appropriate permissions within the system.

5.4 Integration of OpenStack with the EGI AAI proxy

Until recently OpenStack resources of the EGI Federated Cloud were able to authenticate and authorize users solely through X.509 proxy certificates. This prevented users without a client X.509 certificate to interact with the providers. The latest versions of the OpenStack identity service (Keystone) now includes support for federated users: Keystone can act as a Service Provider (SP) that consumes identity properties issued by an external Identity Provider, such as SAML assertions or OpenID Connect claims. The OpenStack web dashboard (Horizon) has also introduced support for this kind of authentication and authorisation. Based on this new OpenStack feature EGI has connected OpenStack with the EGI AAI proxy. The integration is achieved with a new configuration of the Keystone federation that uses EGI AAI as a SAML IdP. Detailed instructions are available in the OpenStack guides[21].

The screenshot below shows the SAML-enabled OpenStack dashboard displaying a drop down menu of available authentication providers. When selecting the egi.eu IdP proxy option, the user will be redirected to the EGI AAI proxy and can follow the process described in section 5.1 to authenticate with an ELIXIR account.

[21] http://docs.openstack.org/developer/keystone/federation/federated_identity.html

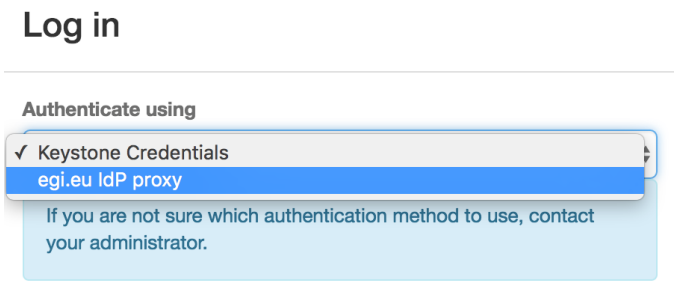

Figure 1. EGI AAI IdP proxy configured in OpenStack

Site administrators need to set up the rules that map the attributes released from the EGI AAI proxy to local OpenStack identity objects. There are many ways to setup and combine these rules. For the EGI integration the following objects are used:

- eduPersonalUniqueId, is mapped to the user id in Keystone.
- eduPersonEntitlement defines the OpenStack groups the user is allowed to access,
- Level of Assurance may restrict grant access to users with the desired LoA

Once the eduPersonEntitlement contents are agreed, recommended mappings will be provided for administrators to easily integrate their resources. Users authenticated through this mechanism can be managed by the OpenStack administrator as any other user.

6 Summary and next steps

Based on the joint work of the ELIXIR and EGI communities, the basic building blocks of the ELIXIR compute platform have been established:

1. The ELIXIR AAI and EGI AAI systems have been connected, and ELIXIR users can login with ELIXIR accounts to the EGI GOCDB and AppDB services. Community managers, cloud resource managers are recognised with distinguished roles within these systems.
2. Installation guidelines and tools have been improved for OpenStack and OpenNebula cloud providers to participate in the EGI Federated Cloud, and particularly in its ELIXIR Virtual Organisation.
3. 3 cloud providers (CESNET, EMBL-EBI and GRNET) are deployed and are ready to join the ELIXIR Compute Platform that is represented by the vo.elixir-europe.eu virtual organisation.

The next step for the CC is to finalise the integration of the reported building blocks, and then start implementing the cloud use cases from M6.3 on top of it. Particularly the CC members – with the support of the broader ELIXIR and EGI communities – must:

1. Join the ELIXIR Virtual Organisation with the three existing sites (CESNET, EMBL-EBI, GRNET).
2. Work with other cloud providers in the CC to eliminate the issues that are blocking them to join the ELIXIR Compute Platform (CSC, CNRS, SURFsara).
3. Define the policies and protocols for ELIXIR users and use cases about getting access to resources from the ELIXIR Compute Platform. (Accounting and monitoring of resources; Policies and quotas on resource use.)
4. Discuss, specify and implement the involvement of ELIXIR community-specific services (e.g. Beacon service[22] by the Global Alliance for Genomics & Health)
5. Create the VMs that are required for the M6.3 scientific use cases; Roll these out to the ELIXIR VO through the AppDB VM catalogue. Document the experiences in D6.15 (due in January 2017).

Besides the above tasks the following actions can help even further in expanding the uptake of the ELIXIR Compute Platform. These can mostly run based on the experiences gained from the CC use cases, but some can already start and run in parallel with those:

1. Decide on how to structure the ELIXIR Compute Platform. How many, and what kind of Virtual Organisations to create? What user roles to use? How much capacity/service limitations to associate to those roles?
2. Finalise the ELIXIR access to OpenStack clouds through the ELIXIR-EGI AAI integration. This will lower the barrier even further for ELIXIR users to access resources in the ELIXIR Compute Platform. Depending on the experiences introduce similar ‘direct’ connection to other types of clouds that form the ELIXIR Compute Platform.

[22] https://genomicsandhealth.org/work-products-demonstration-projects/beacon-project-0

3. Based on the operational experiences gained from the use cases in the CC, create documentations for the broader ELIXIR community on how to join and operate resources in the ELIXIR Compute Platform.