



ELIXIR competence center – Three months remaining

Kimmo Mattila / CSC

www.elixir-europe.org



ELIXIR competence center

- **ELIXIR:** Research infrastructure focused on life science data. Supported by EMBL-EBI and 20 European countries



- The **ELIXIR competence center** (part of EGI-Engage project) aims to facilitate and promote usage of EGI resources among ELIXIR and molecular biology research community.

ELIXIR CC members

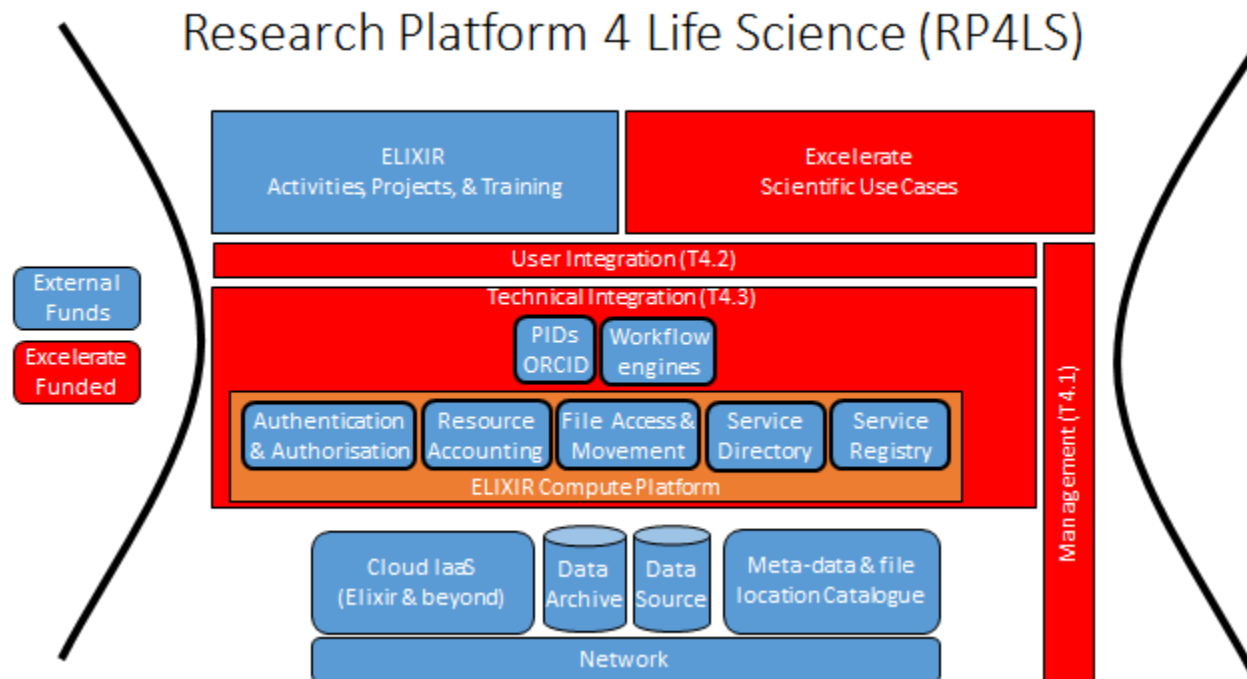


- CESNET (<https://www.cesnet.cz/>)
- CNRS IFB-core (<https://www.france-bioinformatique.fr>)
- CSC (<http://www.csc.fi>)
- EMBL-EBI (<http://www.ebi.ac.uk/>)
- GRNET (<https://grnet.gr/en/>)
- SurfSARA (<https://www.surf.nl/over-surf/werkmaatschappijen/surfsara>)
- Other collaborators: (Open Science Grid, other ELIXIR nodes)



The ELIXIR Compute Platform

A common compute platform to support ELIXIR-EXCELERATE H2020 use cases. Later on other users too.



The ELIXIR Compute Platform

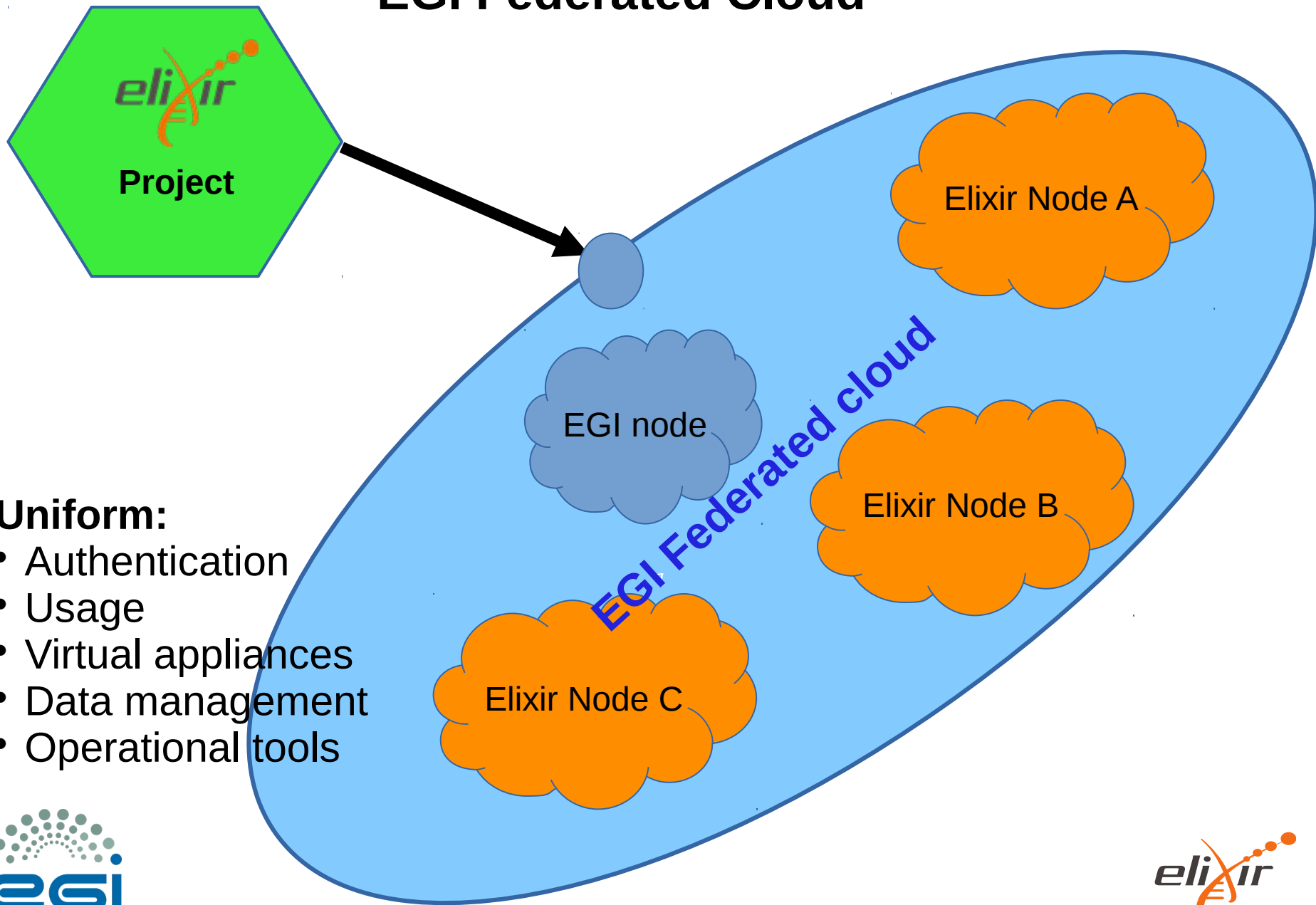
A coherent collection of services and technologies that allow effective production bioninformatics services

- Computing (**cloud** & HTC)
- Storage and Data transfer
- **Authentication and authorization**
- **Applications and tools**

EGI, EUDAT, PRACE, and **GEANT** are among potential service providers.



EGI Federated Cloud



Uniform:

- Authentication
- Usage
- Virtual appliances
- Data management
- Operational tools

ELIXIR CC - so far

- Launch: September 2015
- 5 Use cases selected
- Providing services to EGI Federated Cloud:
 - Best practices and experiences report
- Authentication and authorization:
 - Interoperability between ELIXIR AAI and EGI AAI
 - ELIXIR VO established



ELIXIR CC - Technics

- EGI Federated Cloud resource integration process and has been tested and improved
- ELIXIR VO and ELIXIR compatible AAI available in EGI Federated Cloud
- CILogon can be used to create proxy certificates
- AppDB is used for distributing virtual appliances
- rOCCI, jOCCI, OpenStack Cli and Terraform can be used manage the VMs



ELIXIR CC - Technics

- Still only 3 (CESNET,EMBL-EBI, GRNET) out of 6 CC members are providing resources for ELIXIR VO.
- **Local issues slow down the service integration**
 - Technical issues
 - User policies and resource allocation
 - gain/loss ratio for service provider

Scientific Use cases

Three out of five scientific use cases have effectively used EGI Federated Cloud:

- Marine Metagenomics
- Insyght comparative genomics
- PhenoMeNal

Less active use cases:

- cBioPortal
- JetStream collaboration



Marine Metagenomics 1.

METAPipe integraton

A tool to automatically launch a spark-cluster with metagenome annotation server.

METAPipe environment development:

- University of Tromsø

Virtual spark cluster implementation in OpenStack

- University of Tromsø (Aleksander Agafonov)
- CSC

Terraform based EGI Federated Cloud implementation

- CESNET (Cuong Duong Tuan)

Meta-Pipe Analysis work-flow in EGI Federated Cloud

Metapipe User interface

Meta-pipe 2.0 (beta) (version 0.1.0)

input:

Cutoff for assembled contigs:
500
Remove any contigs that are shorter than this length

Database search tools:

 Blast+Uniref50
 InterProScan5
 PPIAM

Enable developer options (not supported):
yes (NOT recommended, NOT supported, DO AT YOUR OWN RISK!)
This is for internal development use only.

Meta-pipe job tag:

MMG-Cluster setup tool
<https://github.com/cduongt/mmg-cluster-setup-CESNET>
Terraform + rOCCI + proxy certificate
Resource definitions



EGI Federated Cloud

Front node

Spark master

Spark worker

[METAPIPE Job List Admin] || 11 jobs || Status: 8 job

Filter by: JobID State UserID GroupID

Job ID	State	Submitted	UserID	GroupID	Tag	Latest Attempt State	Latest Attempt ExecutorID
gt_IH9OLZJGSX	SUBMITTED	14/03/2017 08:35:55.898	system_gateway	cesnet-metapipe	RUNNING	test-executor	
gt_TAGQONQFAB	DELAYED	15/03/2017 13:58:43.141	system_gateway	cesnet-metapipe	EXECUTOR_FAILED	test-executor	
gt...	DELAYED	13/05/2017 10:39:30.300	system_gateway	cesnet-metapipe	WORKFLOW_FAILED	test-executor	
gt_JUMZFBRYAK	SUBMITTED	09/03/2017 14:45:21.550	system_gateway	CESNET	QUEUED_LOCALLY	default	

Meta-Pipe Job manager

Marine Metagenomics 2.

EBI Marine Metagenomics pipeline

- Goal: pipeline pulls the data from the public ENA database, carries out the processing and send the results back to a private FTP endpoint.
- Terraform based implementation
- EBI has been testing the pipeline in different cloud environments
- Test in EGI Federated cloud is just starting

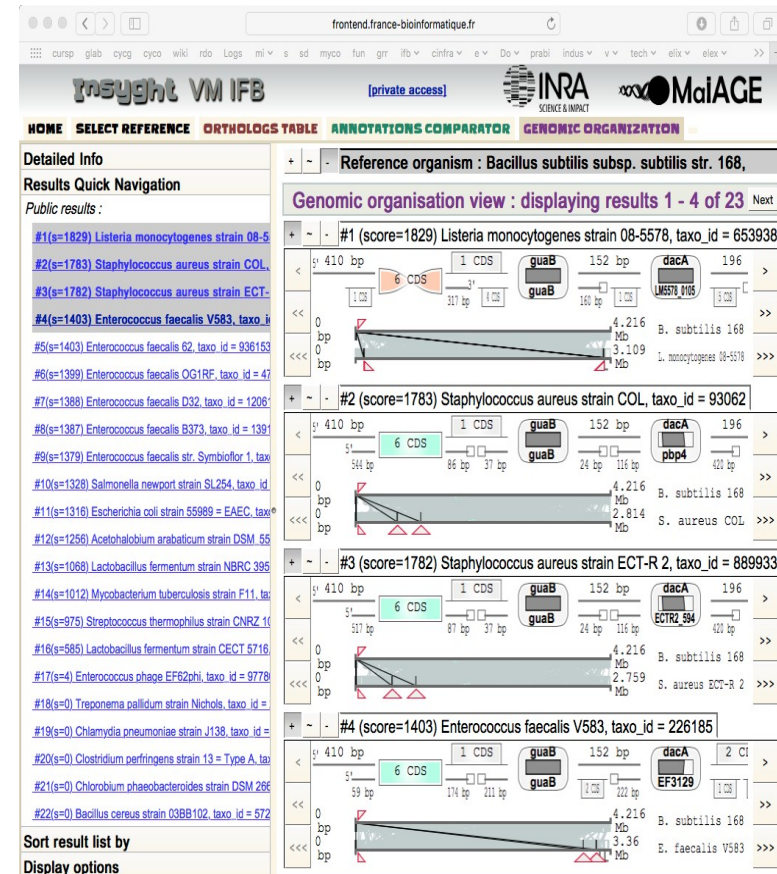
PhenoMenal

- Deployment available for OpenStack, AWS and Google Compute platform.
- PhenoMeNal frontend utilises the EMBL-EBI's Cloud Portal REST API to carry out deployment.
- Testing FedCloud usage with the new **Terraform OpenStack plugin developed by EGI** for the EGI FedCloud will soon be carried out.

Insyght, Comparative genomics



- Scientific field : microbial genomics
- Usage : analysis / navigation among large sets of data (homologies, syntenies, etc.)
- Domains of application : comparative genomics
 - Finds niche-specific genes, infers protein functions, explore phylogenetic profiles, analyzes evolutionary events
- Contact: T. Lacroix (IFB-MIGALE)



Insyght in EGI Federated Cloud (ELIXIR-CC)



Complex application, it requires

- web interface
- DBMS
- computing cluster with a master and several computing nodes.

Deployed on the EGI Fed Cloud

- Deployment done manually with scripts and the use of the jocci tool.
- Users should be able to deploy themselves a whole infrastructure



Insyght in EGI Federated Cloud (ELIXIR-CC)



Foreseen improvements

- Evaluate the new AppDB Dashboard for the deployment
- Use of an orchestrator (Insyght is currently deployed on the French life science clouds with the SixSq-SlipStream orchestrator)



Last months of ELIXIR CC

- Increasing EGI Federated Cloud compatibility
- Implementing use cases
- Piloting usage
- Reporting

After EGI-Engage

- Use cases will continue evolving
- ELIXIR VO offers one possibility for resource management
- AAI development and testing will continue
- Scientific User cases may be linked EOSC



Lessons learned

All components, that are required to utilize EGI Federated Cloud by the ELIXIR community

Now exists, but:

- End-users are not accustomed to certificates, scripts and OCCI
- Support from EGI Federated Cloud team is essential, both for service developers and for end users.
- Capacity providers should benefit more for providing resources through EGI Federated Cloud.
- End user dissemination requires mature services with sufficient computing/storage resources