EOSC-Life

EOSC-Life: Building a digital space for the life sciences

D7.2 – Final EOSC-LS Cloud Observatory Report

WP7 – Cloud Deployment Services
Lead Beneficiary: EMBL-EBI and Charité
WP leaders: Steven Newhouse and Harald Wagener
Contributing partner(s): EMBL-EBI, Charité, ERINHA, Masaryk University, CSC, and CNR

Authors of this deliverable: **S. Newhouse, H. Wagener, M. Dieckmann, A. Niewielska, J. Clark-Casey, S. Insua**

Contractual delivery date: **28 February 2022**
Actual delivery date: **08 June 2022**
H2020-INFRAEOSC-2018-2

Grant agreement no. 824087
Horizon 2020
Type of action: RIA

Table of Contents

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Executive Summary

This report details the evolution of the support mechanisms available to the Biological and Medical ESFRI research infrastructures (BMS RIs) and how the allocated cloud resources have been used at this stage of the EOSC-Life project by the BMS Host Institutions (BMS HIs). We analyse the activities carried on to date within the Work Package 7 (WP7) tasks:

- T.7.1 Cloud Resource Planning and Integration
- T.7.2 Cloud Training Programme
- T.7.3 Sensitive Data Hosting and Processing Policy Framework
- T.7.4 Managed Platforms for Data Analysis
- T.7.5 Operations and Support Helpdesk

and provide our point of view on how the WP7 work will evolve during the remaining months of the project. In general this deliverable does not repeat reporting of activities already reported in the initial version of the Cloud Observatory Report (D7.1[1]), except where is needed to provide context.

Our observations within WP7 from the first 3 years of the project are:

- There is a considerable need for upskilling. Cloud technologies are still complex and require substantial understanding of IT infrastructure basics to set up and operate.
- A recent shift to container technologies unifies and simplifies distribution of applications and services, but requires additional skills from operators and scientists. A similar effect can be seen in the shift between traditional file and directory based storage and object storage for modern, cloud-native applications and services.
- An alternate approach is the use of services and tools that abstract away the need to understand how the underlying infrastructure works.
- This bears a concrete risk that Life Scientists and their Institutions lose the ability to develop, run, and operate their own infrastructures because there will be a lack of organic specialisation from within Life Sciences domain to facilitate future development of infrastructure and workflow management development
- All approaches still face challenges in the appropriate storage and handling of sensitive data. While service-based workflow offerings allow ease of access, they make reasoning about (transient) data storage more difficult.

Project Objectives

The goal of WP7 is to provide a set of integrated cloud resources to support the cloud based FAIR RI data resources (WP1), workflows (WP2) and science demonstrators (WP3) that are being supported within the project. Through the provided cloud resources, this WP also promotes the implementation and adoption of cloud interoperability standards to ensure that EOSC cloud providers are compatible with life-science services and with each other.

[1] https://doi.org/10.5281/zenodo.4010420

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

With this deliverable, the WP7 has contributed to the following objectives:

a. Provide an integrated set of public sector community cloud providers appearing in the EOSC Service Catalogue with access managed using the LS AAI (WP5) - Task 7.1
b. Procure commercial cloud providers using a cloud procurement framework (e.g. HNSciCloud or GEANT) that are integrated with the LS AAI and appear in the EOSC Service Catalogue. - Task 7.1
c. Allocate resources from the supported EOSC-Life cloud providers in the EOSC Service Catalogue to the selected workflows and demonstrators from the EOSC-Life project, and compensate the cloud providers for the usage made at the agreed rates and mechanisms approved by European (H2020) projects. - Task 7.1
d. Provide a technical support and consultancy function that will work with the selected demonstrators and workflows to support their deployment onto the allocated cloud resources to eventually establish a sustainable ResOps "Community of Practice" for cloud-enabled workflows. - Task 7.2 and 7.5
e. Establish a set of quantifiable capabilities needed for the secure hosting of sensitive data and identify which of the supported cloud providers in the EOSC Service Catalogue have these characteristics. - Task 7.3
f. Principles for a secure and trusted cloud will be implemented in cooperation with WP4 which will collect the requirements for the BMS RI so that cloud solutions offered for the storage and computation of human research data comply with the legal requirements for sensitive data. - Task 7.3
g. Data access controls for sensitive data on cloud target systems are enforced in collaboration with WP5, and leverage technologies (e.g. OAuth2 based implementation) from previous projects (CORBEL and EXCELERATE). - Task 7.3
h. Provide deployable platforms for e.g. Galaxy and CWL based workflows (using GA4GH Cloud-WG compatible service). - Task 7.4

Detailed Report on the Deliverable

The report details the support mechanism available to the BMS RIs and how the allocated cloud resources have been used so far by the BMS RIs. To do so, we analyse the activities carried on to date within the WP7 tasks.

Activities on Cloud Resource Planning and Integration

Task 7.1 is led by Steven Newhouse (EMBL-EBI) with contributions from the team at EMBL-EBI (Susheel Varma, Montse Gonzalez, Oihane Fano-Bilbao and Justin Clark-Casey) and feedback from across the Work Package and project.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Background

The purpose of this task is to establish a transparent process by which requests for cloud resources coming from across the project (primarily WP1, WP2 and WP3) can be allocated to a set of cloud service providers internal and external to the project.

Description of the work accomplished

The initial focus during the first year was to establish the EOSC-Life WP7 Resource Allocation Process[2] (RAP) by which cloud requirements from across the project could be received and reviewed, before being allocated to the most appropriate cloud service provider. The RAP team is accessed through eosc-life-wp7-rap@elixir-europe.org and will initially ask requestors to submit their request through a form[3] which captures all relevant information to support the review. Following successful technical review and feedback, the activity will be allocated to a cloud provider.

EOSC-Life is able to access a variety of cloud service providers from within and external to the project. Internally, cloud service providers are expected to appear in the EOSC Service Catalogue[4] under (currently) the ELIXIR and eventually the 'EOSC Life' tag. Collaboration has also been established with the Open Clouds for Research in Europe (OCRE) project which will be providing cloud resources that can be used for payment and some additional support for moving applications to cloud providers.

Since the previous cloud observatory report, three further requests for resources have been received through the WP1 'cloudification' of data resources call and the further demonstrators call from WP3. These have been handled through the RAP. As with the earlier two projects, all have been for relatively minor quantities of resources, in this case between 1K and 10K Euros.

Next steps

The RAP has successfully handled requests from an initial proposal by a project through to their acceptance of resources from a selection of cloud providers. However, the number of requests has been low despite a number of presentations given by WP7 (e.g. in WP1 and WP3 open call outreach sessions) about the resources on offer. Moreover, the requests received have been for a small amount of resources.

From evidence elsewhere (e.g. from non-RAP technical reviews conducted by WP7 for the WP1 and WP3 open calls), it appears that many projects already have a relationship with a cloud provider, for example one commonly used by their institute(s) before they engage in the open call process. This is particularly true of larger projects - commonly their resource requests are for

[2] https://docs.google.com/document/d/1Z0Izwh8vafqMZ8R6dDmIQgvjRxsF3IZMjl65s3m2xL4/edit?usp=sharing
[3] https://docs.google.com/document/d/1lEf3HMbULAiXZ6zLGSCKj3SLs37enq1ud4Do3E5efaM/edit
[4] https://marketplace.eosc-portal.eu/services/c/compute?research_areas-filter=&research_areas-all=&providers-filter=&providers-all=&target_groups-all=&related_platforms-filter=&related_platforms-all=&related_platforms%5B%5D=15&rating=&location=

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

supplementary resources for purposes such as one-off bursts of processing or training rather than sustained requirements.

How and whether this needs to change will be a topic that we need to explore, both in the EOSC-Life context and in the EOSC context generally.

Activities on Cloud Training Programme

Task 7.2 was led by Tony Wildish (EMBL-EBI) and Marius Dieckmann (ELIXIR-DE), with contributions from the EMBL-EBI Cloud Consultants (David Yuan, CD Tiwari, and Soumyadip De) & CSC's Cloud specialists (Shubham Kapoor, Juhani Kataja & Kalle Happonen). In April 2021 Santiago Insua took on the role of Tony Wildish and continued the actions on this task.

Background

The purpose of this task is to provide the scientific community with the knowledge they need to make effective and efficient use of cloud technologies. Many scientists have plenty of experience of HPC computing with large in-house batch clusters, but do not know how to translate that experience into computing in the cloud. This task will draw on existing training materials that exist in the community, with the team filling in the gaps in training material needed specifically for the life science community. Cloud training materials will be made available on-line with the EOSC-Life WP9 (e.g. using ELIXIR Training platform TeSS service[5]).

In D7.1, a 'ResOps course'[6], covering the use of containers built with Docker, continuous integration/deployment with Gitlab, and an introduction to Kubernetes was prepared, alongside an advanced Kubernetes course, as well as IaaS cloud (OpenStack)[7] and PaaS Cloud (Kubernetes/OKD)[8] solutions.

Description of the work accomplished

The ResOps course has been made available as a set of standalone videos and exercises, suitable for study by individuals or groups in a non-classroom setting, or for use in a scheduled course. The on-line course is available via the EBI training webpage[9] and the public repository[10].

The course contents have been reviewed and updated on Jan. 2022.

de.NBI and EMBL-EBI are now also collaborating on an advanced Kubernetes training course, building on from the material of the two basic courses. This will be a one-day remote attendance

[5] https://tess.elixir-europe.org
[6] https://bit.ly/resops-2019
[7] https://www.csc.fi/fi/web/training/-/pouta-cloud-course-2019
[8] https://www.csc.fi/fi/web/training/-/fundamentals-of-container-clouds-with-rahti
[9] https://www.ebi.ac.uk/training/online/courses/resops-cloud-native-tools-and-technology-for-researchers/
[10] https://gitlab.ebi.ac.uk/TSI/tsi-ccdoc

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

event, with all the presentations recorded, so it can be taken offline by future participants. This course has been delivered and the material published at GitHub[11].

Activities on Sensitive Data Hosting and Processing Policy Framework

The task 7.3 is led by Harald Wagener (BIH, Germany) with contributions from WP7 and WP4 members, namely Chris Lawerenz, Willem van Boiten and Christian Ohmann.

Background

The objective of the task is to support users to find cloud providers that are able to meet their sensitive data processing requirements and are able to offer the capabilities expected of research infrastructure and services so there is an assurance to the user and the data owners that the cloud infrastructure and services can be used to host and process sensitive data for research and analytics.

Description of the work accomplished

A questionnaire has been developed to allow cloud providers to go through a self-assessment. The result of that self-assessment aims to provide end users a selection of cloud providers that have filled out the questionnaire.

The current document is available as Google document[12].

The questionnaire and accompanying documents have been circulated with various infrastructure providers who participate in EOSC-Life WP7 and have been shared with other interested third parties (such as WP4, and external projects like HealthyCloud). The current questionnaire can be answered by infrastructure providers with ease (about 2 hours of effort, less so if the infrastructure is certified by one of the many standards the questionnaire references).

The questionnaire has been shared with adjacent projects such as HealthyCloud and is pending integration into the WP4 toolbox, as soon as that becomes available.

Activities on Managed Platforms for Data Analysis

This activity was initially led by Shubham Kapoor (CSC, FInland) and for most of year 1 by Susheel Varma (EMBL-EBI) and now is led by Ania Niewielska (EMBL-EBI) and Justin Clark-Casey (EMBL-EBI). Additional contributions have come from the Software Development & Operations team at EMBL-EBI (Ania Niewielska, Thiago Albuquerque, Cibin Sadasivan Baby and others), Cloud

[11] https://github.com/ag-computational-bio/kubernetes-course

[12] https://docs.google.com/document/d/1751i2qJ0NJZnoDuVXEOLb8mT_8OAiKeMuIaMT2iXBUs/edi

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Specialists at CSC (Álvaro González, Kalle Happonen, Jarno Latinen & others) and ELIXIR-DE (Marius Dieckmann, Sven Twardziok & others).

Background

The purpose of this task is to contribute to the design and development of community-created workflow execution platforms that can be deployed onto EOSC cloud infrastructure and offered to users as per managed *Platform as a Service (PaaS)* or in some cases as per *Software as a Service (SaaS)* models.

These platforms will leverage global standards emerging from the GA4GH community. Users of the Workflow Platform will be able to create portable workflows in languages such as CWL using tools packaged as containers and send those workflows for execution to Workflow Platform installations. The choice of execution environment will be primarily driven by data locality, availability of resources and the cost of compute.

The joint effort around developing GA4GH-Compatible Workflow Platform is further complimented by the ELIXIR Cloud and AAI Project[13]. This aims to establish a federated cloud computing network across European e-infrastructures by enabling the analysis of population-scale genomic and phenotypic data across participating international nodes. GA4GH recognizes this as a Driver Project[14], one that will help guide their development efforts and pilot their tools.

Description of the work accomplished

Development activity dominated the initial phase of the project, the biggest effort being to transform a successful Proof of Concept (PoC) into a production ready system.

It soon became apparent that transforming the loose set of tools and deployments into a production platform could not be done in isolation. A project of this magnitude requires alignment with other academic and commercial third-party efforts, and the platform would not reach maturity without contributions from workflow engine developers and further development of GA4GH standards. Thus, in the later phases we shifted the majority of our effort into connecting with the wider workflow engine community and other GA4GH standards implementers, and developing the GA4GH Cloud Workstream standards according to the community needs.

This development effort continues. Since the previous cloud observatory report the majority of planned enhancement and new features have been implemented or are under active development:

- **S3 compatible object store support** for large data sets has been added in TESK backend component

[13] https://elixir-europe.github.io/cloud/
[14] https://www.ga4gh.org/how-we-work/driver-projects/

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

- **CI/CD** activity continues within the Pubgrade[15] project - a decoupled CI/CD system to roll out new versions to decentralised network of deployments
- The support for **new workflow languages** - Nextflow and Snakemake - has been added to the platform via the WesKit[16] project and implementation of a TES client in the Snakemake engine.
- Introducing intelligent **workflow federation logic** is planned as a project for the upcoming Google Summer of Code.

The current use of the Platform is limited to test installations and demonstration workflow runs. Nevertheless, the presence of the Platform at various events (detailed below) generates considerable interest and the community around the initiative is growing.

The Platform has been consistently present at the following events since 2019:

- ELIXIR AHM
- GA4GH Plenary
- ELIXIR BioHackaton
- Google Summer of Code

The Platform has demonstrated execution of real life scientific workflows namely:

- Rare Diseases Workflow, setup includes execution of workflow on the existing platform with or without FEGA integration.
- Multiple Sequence Alignment workflow for comparing Covid19 surface glycoprotein.
- Marine Metagenomics (MMTESP) workflow.

Finally, Elixir Cloud & AAI project members are actively contributing to the development of GA4GH standards and have been acting as spec champions of TRS and TES standards[17].

Next steps

The effort of project members, who have been co-leading the development of GA4GH Task Execution Service API and guided it through the approval process, has resulted in the first officially approved[18] version of the standard (GA4GH TES v1.0). Development work continues - the next version of TES is expected to include changes requested by the workflow engine community (the Galaxy Project and Nextflow) and the commercial cloud vendors (Microsoft).

Ownership of the very first piece of the GA4GH-Compatible Workflow Platform, a component called TESK that implements the TES standard, has been transferred from EMBL-EBI to the Elixir Cloud & AAI community. This is to broaden participation and increases the likelihood that the project will be sustained beyond EOSC-Life. Active development continues.

TESK developers have been recently contacted by a commercial vendor who used the component in their proprietary large scale workflow execution platform. They have extended it with various

[15] https://github.com/elixir-cloud-aai/Pubgrade
[16] https://gitlab.com/one-touch-pipeline/weskit/documentation
[17] https://doi.org/10.1016/j.xgen.2021.100029
[18] https://www.ga4gh.org/news/ga4gh-tes-api-bringing-compatibility-to-task-execution-across-hpc-systems-the-cloud-and-beyond/

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

features and want to work with the TESK developers to contribute those improvements to the public TESK codebase. This should bring closer the ambitious goal of handling large simultaneous workflows within the platform.

Activities on Operations and Support Help desk

Task 7.5 is led by Montserrat González Ferreiro (EMBL-EBI) with the collaboration of all WP7 and especially Matej Antol (MU) and Alexandr Kolovratnik (MU).

Background

The technical operation for individual cloud resources is provided outside the EOSC-Life project but the operation and support of the EOSC-Life infrastructure as a whole are supported through task 7.5. Operations and Support Helpdesk. The task involves providing a support help desk and a monitoring system for key services.

These are the requirements for the support help desk:

- It should be accessible by end-users from WP1, WP2, WP3, and cloud providers from within the BMS RIs.
- It should bring together experts in cloud deployment and cloud operation to help integrate cloud providers into the EOSC infrastructure and to support end-users and data experts as they deploy their workloads.
- It should provide the demonstrators selected in WP3 access consultancy support as they start planning their cloud deployment. This consultancy support is provided by technical staff participating in WP7.
- It should establish ticket queues in an existing helpdesk technology (e.g. Request Tracker) to support task 7.5 and potentially other WPs within the project.

These are the requirements for the monitoring system:

- Monitor any key services (e.g. data transfer) to ensure their continued operation and performance.

Description of the work accomplished

As reported in D7.1, the helpdesk was set up following an initial survey to understand the requirements, defining a set of procedures for handling the incidents and service requests to meet these requirements, the configuration of Request Tracker (RT) hosted by Masaryk University which can be accessed by eosc-life-helpdesk@elixir-europe.org. A supporting Privacy Notice is provided for the helpdesk. The helpdesk was structured around the available services and software in WP7 captured as a result of a survey across WP7.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Work accomplished	Objective	Date
Survey to monitor key services and gauge their impact to the EOSC-Life community	Collection of monitoring systems, impact measurement and success stories. A report was created and shared within the EOSC-Life community (See Appendix 1).	Jan 2021
Fine-tune the Request Tracker queue at Masaryk University	During 2020 EMBL-EBI and MU discussed and implemented improvements in the triage mechanism. Also kept the access to the relevant people updated (joiners/leavers), and explored the possibility to provide a joint help desk service with WP5 but the offer was declined.	Feb 2021
Communicate the WP7 Helpdesk availability to WP3	Inform new projects and WP3 of the availability of the WP7 Helpdesk via email.	Jun 2021
Present the WP7 Helpdesk and gather feedback from WP3	A presentation in the WP3 monthly meeting provided an introduction to the WP7 Helpdesk.	Sep 2021
Helpdesk operation	EMBL-EBI and MU keep the Request Tracker queue configuration updated and review it periodically to make it as efficient as possible.	Ongoing

Next steps

During the first part of the project, we have focussed on providing the help desk solution to provide a unique contact point to the EOSC-Life service users. During the second half of the project, we focused on ensuring that the help desk adapted to any new project needs, engaged WP3, and monitored the key services provided by WP7. In practice, the helpdesk has seen little adoption and WP1 / WP3 used their own implementation to manage their open and sensitive data calls.

WP1 & WP3 engagement

WP7 have supported WP1 and WP3 calls through various activities: Technical review of call submissions, consultancies with applicants on their intended use of cloud technologies and need for training; offering of training courses (see above), as well as procuring resources through the Resource Allocation Process (see above). In the current phase, WP7 is attending status and

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

update meetings between the funded calls and WP1/WP3 to make sure the projects are on track from a technical point of view and get the training, support, and resources required to make the best use of the offerings by partnering Research Infrastructures.

Conclusion

The deliverables of WP7 have led us through three years of discussion, sometimes intense, as to how we satisfy the needs of life scientists when it comes to the cloud. What regulatory challenges do we need to understand and how do we work through the limitations they can sometimes impose? How do we make the cloud and workflow services that we offer easy for researchers to obtain and consume? How can we best equip scientists and developers with the skills they need to use that infrastructure? And last but not least, what can we do to address the concerns of researchers about the sustainability of services and training?

Our involvement with WP1 and WP3 has shown that most research projects fall into two categories. Either they are staffed with experts who are familiar with a given infrastructure or service offering and have secured access to those resources beforehand, or the project participants have little to no skills when it comes to the cloud, and accordingly their requirements are ill-defined with little access to infrastructure.

For the latter projects, adopting a higher level service such as GALAXY that avoids exposing the user to low-level technical concerns can sometimes be a quick win. But overall, the challenges faced by this group lead to a slow adoption of cloud services, not least because of their concerns about the sustainability of smaller/institution specific cloud infrastructures that are dependent on project-to-project funding.

The current situation bears a concrete risk that Life Scientists and their Institutions lose the ability to develop, run, and operate their own infrastructures. If the trend of centralising cloud infrastructure continues, then there will be fewer organic opportunities to identify talent that can develop, run, maintain, and operate the services needed in the future.

Abbreviations

AAI - Authentication and Authorisation Infrastructure

BMS RI - Biological and Medical Research Infrastructures

BMS HI - Biological and Medical Host Institutions

CI/CD - Continuous Integration and Continuous Delivery/Deployment

CWL - Common Workflow Language

DRS - Data Repository Service

DURI - Data Use & Researcher Identities

EOSC - European Open Science Cloud

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

ESFRI - European Strategy Forum on Research Infrastructures
FAIR - Findable Accessible Interoperable Reusable
FEGA - Federated European Genome-Phenome Archive
FTP - File Transfer Protocol
GA4GH - The Global Alliance for Genomics and Health
HTTP - Hypertext Transfer Protocol
LS - Authentication and Authorisation Infrastructure
PaaS - Platform as a Service
PoC - Proof of Concept
RAP - Resource Allocation Process
RI - Research Infrastructure
RT - Request Tracker
SaaS - Software as a Service
SDO - Software Development and Operations
SLA - Service Level Agreements
TESK - Task Execution Service
TeSS - Training e-Support System
TRS - Tool Registry Service
WDL - Web Development Language
WES - Workflow Execution Service
WP - Work Package
WS - Web Services
WPL - Work Package Leaders

Delivery and Schedule

The delivery is delayed: Yes

This deliverable was delayed by three months due to changes to the WP leadership.

Adjustments

Adjustments made:

None

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Appendices

Appendix 1: WP7 Cloud Deployment Services monitoring & Impact by January 2021

1. CSC/ELIXIR-FI

Service monitoring

cPouta & ePouta are monitored using OpenStack APIs and its performance via tests, hypervisors, router performances, VM flavour statistics, and resource usage. Naturally also central storage services and networks are monitored as well. Nagios is used for data gathering & Opsview for alerts. Also graphite is used for storing numerical time series and Grafana for visualisation of the numerical times series.

Rahti monitors OKD nodes, cluster accessibility, OKD API calls & cluster usage. Nagios & Prometheus are used for data gathering, Opsview for alerts & Grafana for visualisation.

Service impact

The life-science usage of cPouta is 13% and in ePouta over 80%. Both have very high demand and also the container platform is getting more and more popular.

There are several hundreds of users of the life science community, from CSC perspective, who are the IT admins. The IaaS provider does not know how many end-users are using the services since CSC does not administrate the user account on the virtual machines.

Success story

http://www.elixir-finland.org/en/disease-prediction-models-are-becoming-more-accurate-thanks-to-the-computational-methods/

2. EMBL-EBI Embassy Hosted Kubernetes

Service monitoring

Embassy Hosted Kubernetes includes a preinstalled Prometheus Operator which runs as a service in the cluster. This includes Grafana which publishes graphical representations of node metrics on a predetermined port. The link to this is bespoke to the individual cluster. Users can choose to use this, or install their own monitoring application as they have full admin rights to the kubernetes cluster. Internally we monitor the cluster health via Slack alerts derived from the Prometheus service in each cluster and the Openstack API.

Service impact

All users have registered using either the internal EMBL request portal or been commercial users. It is possible some EMBL users are collaborating with EOSC-life.

3. de.NBI Cloud Tübingen

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Service monitoring
IaaS, PaaS, WaaS, SaaS, CaaS is monitored considering their vCPUs, vRAM, storage volume using Zabbix, ELKStack, Prometheus, Tripwire and Grafana. Now two cloud sites following the OpenStack region approach are operating. The monitoring is only accessible internally on admin level but data could be shared for reporting.

Service impact
The number of active users is 258 spread over 84 active projects(05.2022). These are users who have registered for the de.NBI Cloud Tübingen via the de.NBI Cloud portal. Currently we have no means to detect, tag or identify EOSC-life users.

4. de.NBI Cloud Heidelberg

Service monitoring
IaaS, PaaS, WaaS, SaaS, Block storage, Object Storage and practically all kinds of bioinformatics pipelines and their corresponding ecosystems are monitored considering their vCPUs, vRAM, storage, volume and network load using CheckMK, ELKStack and Grafana. The monitoring services are only available internally for de.NBI system administrators.

Service impact
Currently there are 70 users with active projects assigned to our cloud site which have registered via the central de.NBI cloud portal. At the moment there are no means to identify EOSC-life users.

5. de.NBI Cloud Berlin

Service monitoring
IaaS, PaaS, WaaS,SaaS, Block storage, Object Storage, Managed Kubernetes, and practically all kinds of bioinformatics pipelines and their corresponding ecosystems are monitored considering their vCPUs, vRAM, storage, volume and network load using CheckMK, ELKStack and Grafana.

Service impact
Currently there are more than 25 users with active projects assigned to our cloud site which have registered via the central de.NBI cloud portal. At the moment there are no means to identify EOSC-life users.
All de.NBI Cloud locations successfully transitioned from Elixir AAI to LS Login in early 2022 with no negative impact to end users.

6. CNR

Service monitoring
Galaxy VMs with/without encrypted storage, cloud portal for services deployment at IaaS level are monitored with Zabbix. There are also critical problems that could affect the service. Also the PaaS level is monitored with the same approach/tools. Additionally, track is kept of the status of the central services (PaaS Core Services) that are needed to deploy the Galaxy services on the IaaS.
Here you will find an example of monitoring of the services:
http://laniakea-monitor.cloud.ba.infn.it/

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Service impact

The service has been substantially improved recently, making it compatible with the official Ansible recipes to deploy Galaxy, thus allowing a faster alignment of Laniakea with all future Galaxy releases.
The service is currently used by several important Italian research and health institutions that are using Laniakea for their daily work, also outside the context of EOSC-Life, for example:

- Istituto Ortopedico Rizzoli (2 internal Galaxy servers).
- Istituto Zooprofilattico Sperimentale della Puglia e della Basilicata (2 internal Galaxy servers and 1 IRIDA instance).
- Ospedale Pediatrico Giannina Gaslini (public server).
- University of Milan (2 public Galaxy servers, Vinyl and CorGAT, and tools development).
- IBIOM-CNR (public Galaxy server and tools development).
- Istituto Zooprofilattico Sperimentale del Piemonte, Liguria e Valle d'Aosta (private Galaxy cluster)
- University of Turin (training)

While the number of Institutions currently using Laniakea can be easily tracked, we can give only an estimate of the number of users who are using the on-demand Galaxy instances, being these privates. We roughly estimate about 100 active Galaxy users.

7. de.NBI Cloud Freiburg

Service monitoring

The services that the European Galaxy server offers (IaaS, PaaS, TIaaS, SaaS, block storage, object storage) and practically all Galaxy related information (number of workflows and users, queuing states, load on different services, routing statistics, ELIXIR-AAI logins, etc) are monitored using the ELK-Stack and Grafana.
Everything is exposed under https://stats.galaxyproject.eu and related sub-sites.

Service impact

Latest numbers about users, AAI and executed workflows can be found here:
https://stats.galaxyproject.eu/d/000000012/galaxy-user-statistics

At the time of writing:

- 23.000 users
- 300 using LS Login
- 19.000 workflows
- 110.000 workflow executions
- ~ 500.000 jobs a month

Success stories

Success stories can be found at https://galaxyproject.eu/news, especially under the tag TIaaS, paper or COVID-19. Moreover, a list with more than 100 scientific publications citing the European Galaxy server can be found at the project website.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.