# Laniakea@ReCaS: an ELIXIR-ITALY Galaxy on-demand cloud service

Tangaro M.A.<sup>1</sup>, Donvito G.<sup>2</sup>, Antonacci M.<sup>2</sup>, Chiara M.<sup>3</sup>, Mandreoli P.<sup>3</sup>, Pesole G.<sup>1,4</sup>, Zambelli F.<sup>1,3</sup>

1. Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies - National Research Council (IBIOM-CNR), Bari, Italy 2. National Institute for Nuclear Physics, Bari Section, Italy 3. Dept. of Biosciences, University of Milan, Italy 4. Dept. of Biosciences, Biotechnologies and Pharmacological Sciences, University of Bari "Aldo Moro", Italy

# ELIXIR All Hands 2019, 17-20 July, Lisbon, Portugal

Although several Galaxy public services are available, a private Galaxy instance is still mandatory or preferable for several use cases including heavy workloads, data privacy concerns or particular customization needs. Cloud computing technologies provide a viable way to deploy Galaxy private instances, freeing users from the onerous deployment and maintenance of local IT infrastructures. In the last few years, ELIXIR-IT led the development of Laniakea, a software framework that facilitates the provisioning of on-demand Galaxy instances as a cloud service over einfrastructures. The user interacts with a Laniakea service through a web front-end that allows to configure and launch a production-grade Galaxy instance in a straightforward way. Through the interface, the user can deploy Galaxy instances over single VMs or virtual clusters, link them to shared reference data volumes and plain or encrypted volumes for storing data. A selection of "flavours", that is Galaxy instances pre-configured with sets of tools for specific tasks, is also available. When the users is satisfied, Laniakea takes oved and deploys the desired Galaxy instance over

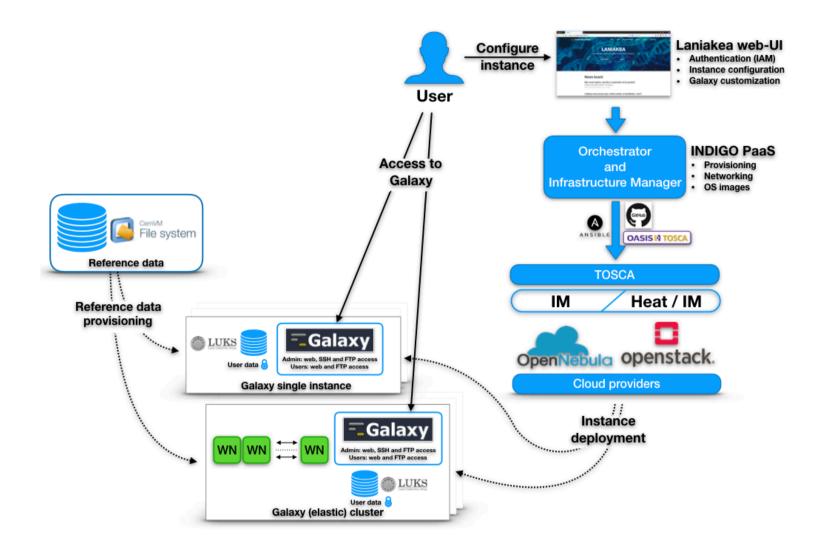
# LANIAKEA

Galaxy is currently the prevailing workflow manager for bioinformatics thanks to its many useful features and a user-friendly interface. While several Galaxy public services are available to researchers, either general purpose or dedicated to specific research domains, there are still many scenarios where a private Galaxy instance is necessary or preferable, including for example heavy data analysis workloads, data privacy concerns or specific customization needs.

the cloud, providing a public IP and full administrative privileges over the new instance.

In Dec-2018, we launched the beta-test phase of the first Laniakea-based Galaxy on-demand ELIXIR-IT service: Laniakea@ReCaS. After six months of helpful testing, we are now ready to announce the production phase of this service.

Access to the service will be provided on a per-project basis through an open-ended call defining terms and conditions, project proposals will be evaluated by a scientific and technical board. Accepted proposals will be granted a package of computational resources for running on-demand Galaxy instances for a duration compatible with the project requirements.



Laniakea is a software framework that facilitates the provisioning of on-demand Galaxy instances as a cloud service over e-infrastructures, by leveraging on the open source software catalogue developed by the INDIGO-DataCloud H2020 project, which aimed to make cloud e-infrastructures more accessible by scientific communities.

End-users interact with Laniakea through a web front-end allowing for a general setup of a Galaxy instance. The deployment of the virtual hardware and of the Galaxy software ecosystem is subsequently performed by the INDIGO Platform as a Service layer. At the end of the process, the user gains access to a private, production-grade, fully customizable, Galaxy virtual instance. Laniakea features the deployment of a stand-alone or cluster backed Galaxy instances, shared reference data volumes, encrypted data volumes and rapid development of novel Galaxy flavours for specific tasks.

#### **GALAXY PRODUCTION ENVIRONMENT**

Galaxy is deployed for a multi-user production environment, i.e. there are some additional auxiliary application needed for the best performance (the basic Galaxy installation is suitable for development by a single user):

- PostgreSQL as database
- NGINX as web server (+ upload module)
- uWSGI link between the service and the web server
- Proftpd as FTP server

Laniakea supports also virtual clusters deployment, allowing to instantiate Galaxy with SLURM as resource manager and to customize the cluster virtual hardware.

### **TOOLS AND REFERENCE DATA**

Galaxy flavors available: Galaxy instances are deployable with several sets of topic-specific (e.g. RNA-Seq, variant calling,...) pre-installed tools.

Reference data available: each instance comes with reference datasets (e.g. genomic sequences) already available for many species, that are shared across Galaxy instances through the CERN-VM File System (cernvm.cern.ch) technology, thus avoiding unnecessary and costly data duplication.

#### **NEW DASHBOARD**

The web front-end provides different tabs for configuring the virtual hardware and the Galaxy instance (e.g administrator credentials).

In the latest development iteration of Laniakea, a novel and more flexible web interface is going to be introduced, facilitating a straightforward customisation of the user experience through human readable YAML configuration files.

Select TOSCA Template: galaxyyaml * Select TOSCA Template: galaxyyaml @ galaxyyaml & galaxyyaml @ galaxyyaml & galaxyyamy & galaxyyam
Configure       Configure         Template: galaxy.yaml       Template: galaxy.yaml         Description: Deploy Galaxy on a single Virtual Machine installing it from scratch (SLOW). The basic configuration includes CentOS 7, the selected Galaxy flavour, companion software and reference data. Configure, click on the "Submit" button and wait for the confirmation e-mail(s) with instructions on how to provide your passphrase (if encryption is enabled) and log in to your new Galaxy instance. If after some
Template: galaxy.yaml       Template: galaxy.yaml         Description: Deploy Galaxy on a single Virtual Machine Installing it from scratch (SLOW). The basic configuration includes CentOS 7, the selected Galaxy flavour, companion software and reference data. Configure, click on the "Submit" button and wait for the confirmation e-mail(s) with instructions on how to provide your passphrase (if encryption is enabled) and log in to your new Galaxy instance. If after some       Description: Deploy Galaxy on a single Virtual Machine Installing it from scratch (SLOW). The basic configureation includes CentOS 7, the selected Galaxy flavour, companion software and reference data. Configure, click on the "Submit" button and wait for the confirmation e-mail(s) with instructions on how to provide your passphrase (if encryption is enabled) and log in to your new Galaxy instance. If after some
Description: Deploy Galaxy on a single Virtual Machine installing it from scratch (SLOW). The basic configuration includes CentOS 7, the selected Galaxy flavour, companion software and reference data. Configure, click on the "Submit" button and wait for the confirmation e- mail(s) with instructions on how to provide your passphrase (if encryption is enabled) and log in to your new Galaxy instance. If after some mail(s) with instructions on how to provide your passphrase (if encryption is enabled) and log in to your new Galaxy instance. If after some
selected Galaxy flavour, companion software and reference data. Configure, click on the "Submit" button and wait for the confirmation e- mail(s) with instructions on how to provide your passphrase (if encryption is enabled) and log in to your new Galaxy instance. If after some mail(s) with instructions on how to provide your passphrase (if encryption is enabled) and log in to your new Galaxy instance. If after some
hours you do not receive any e-mail please be sure to check your SPAM BOX. hours you do not receive any e-mail please be sure to check your SPAM BOX.
Virtual hardware Galaxy Advanced Galaxy Advanced
Instance flavour Galaxy version
Small (1 cpu, 2 GB RAM, 20 GB dsk)         *         Galaxy release 19.01         *
CPUs, memory size (RAM), root disk size Galaxy release 18.05 recommended
Galaxy instance SSH public key Instance description
ELIXIR-ITALY
Paste here your SSH public key Set Galaxy Brand
Enable encryption Galaxy administrator e-mail
e Plain mode
© Encryption
ncrypt storage
Storage volume size Galaxy flavours
50 GB   Galaxy minimal  Galaxy tools preset
zeieci zouade aise
Submit Reference data repository
ELIXIR-IT Galaxy reference data CVMFS repository * Select reference data repository
Submit

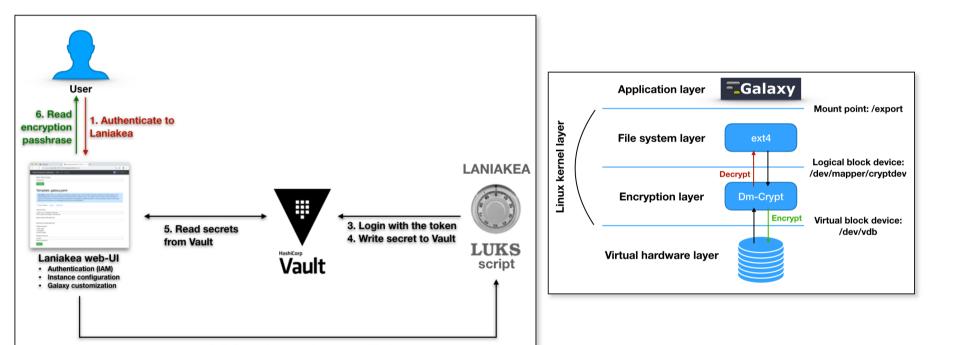
## **AUTOMATIC ELASTICITY**

#### **REVISED DATA ENCRYPTION PROCEDURE**

Data privacy is granted through LUKS storage encryption as a service: data are isolated from any other instance on the same platform and from the cloud service administrators, opening to the adoption of Galaxy based cloud solutions even within clinical environments.

The encryption procedure has been completely re-worked and automated, allowing the user to encrypt storage on-demand, with a strong random alphanumerical passphrase without hassle.

In fact, Laniakea now leverages the key management system Hashicorp Vault (vaultproject.io) to store encryption keys, which are shown only if explicitly requested by the user.



Dynamic cluster resources scaling is supported, deploying and powering-on new working nodes depending on the cluster workload and powering-off them when no longer needed.

2. Deliver token to write secret to Val

# The ReCaS datacenter

The Bari ReCaS DataCenter has been built by the University of Bari "Aldo Moro" and the National Institute of Nuclear Physics (INFN) in the framework of the ReCaS project (PON Research and Competitiveness 2007-2013 Notice 254 / Ric ).

It was completed in July 2015 and inaugurated on July 9, 2015 and hosts ~12000 CPU cores (~ 4 GB RAM per core) and ~6 PB of storage capacity.

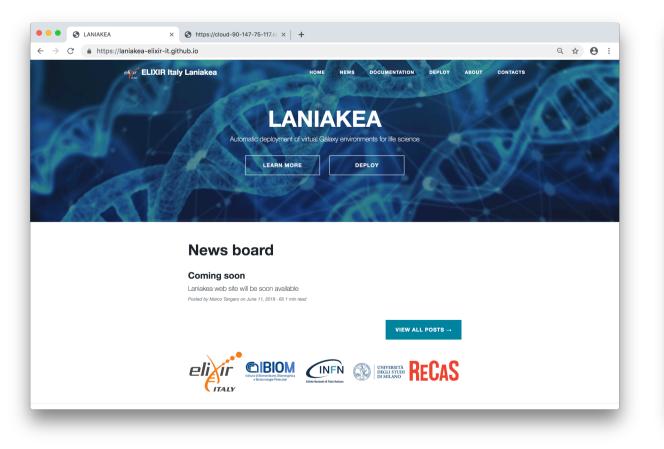
The ReCaS cloud platform that hosts Galaxy onthe demand service is based on OpenStack (Mitaka), with about 1700 CPU cores, 6.7 TB of RAM and 270 TB of storage (replica 3).



## LANIAKEA@ReCaS

In December 2018, Laniakea @ReCaS announced its closed beta program to test the maturity of the service. The program involved the participation of users from several research institutions and universities with different scientific backgrounds. The users were asked to stress-test Laniakea@ReCaS by deploying, deleting and extensively using one or more virtual Galaxy instance for their daily research activities. During the beta program, we gathered as much information from our users as possible, worked in fixing the juvenile issues of the service and prioritized a list of items for future developments.

The production phase of the ELIXIR-IT Laniakea@ReCaS service will start in the second half of 2019. Access to the service will be offered on a per-project basis through an open-ended call defining terms and conditions of the service. Each project proposal will be evaluated by a scientific committee and a technical board. Successful proposals will be granted a standard package of computational resources for running Galaxy instances with Laniakea for an amount of time compatible with the project requirements.



→ C	160.cloud.ba.infn.it:8443						Ŷ	Θ :
aS Orchestrator Dashboard Hom	ne SLAs Settings l	Jsers					Marco Tai	ngaro <del>-</del>
y deployments:								
w 10 💠 entries 🖉 Refresh 🕂	New deployment					Search:		
id	.↓↑ Status	↑ Description ↓↑	Created At	Updated At	Deployed At	t ↓†		
e98856-3313-eecc-9665-ae3b2967fc04	CREATE_COMPLETE		2019-06-06 12:33:00	2019-06-06 12:25:00	provider-RECAS-BARI	-	۹	
e98856-15aa-a2a5-9665-ae3b2967fc04	CREATE_COMPLETE		2019-06-06 12:32:00	2019-06-06 12:24:00	provider-RECAS-BARI	-	۹	
e9882a-d55a-c393-9665-ae3b2967fc04	CREATE_COMPLETE		2019-06-06 07:23:00	2019-06-06 07:15:00	provider-RECAS-BARI	-	۹	W.

$\leftarrow$ $\rightarrow$ C (i) Non sicuro	90.147.75.40/galaxy?tool_id=testtoolshed.g2.bx.psu.edu%2Frepos%2Felixir-it%2Fcovacs_select_filtration%2Fcovacs_Select_Filtration%2F3.8&vers	ion=3.8&_identifer 🟠	0:
👖 App 🔲 Login - OpenStac	Vault		
🔁 Galaxy / Covacs	Analyze Data Workflow Visualize + Shared Data + Admin Help + User + 🇱	Usir	ng 131.5 G
Tools	covacs, Select, filtration SelectVariants VariantFiltration wrapper for covacs, use in case of not enough snp or indels error in v Options	History	2 <b>0</b> []
search tools	covacs_VariantRecalibrator (Galaxy Version 3.8)	search datasets	C
Get Data	input VCF	scanongtftest	
Send Data	C & Nothing selected	12 shown, 15 deleted	
Collection Operations	Using reference genome	821.48 MB	
<u>Lift-Over</u>			
Text Manipulation	hg19_CATK_bundle   Select genome from the list	27: SCanonByGtf (in doc ker) on data 2 and data	• 🖋 🗙
Convert Formats	TYPE	1 :logfile	
Filter and Sort	SNP -	Die SCanan BerCat (in das	
Join, Subtract and Group	246	26: SCanonByGtf (in doc ker) on data 2 and data	• / ×
Fetch Alignments/Sequences	✓ Execute	1 :filtered statistic cvs	
Operate on Genomic Intervals		25: SCanonByGtf (in doc	• / ×
<u>Statistics</u> Graph/Display Data	IMPORTANT to get the wrapper ready to start the admin user have to download gatk GATK version 3.8 from the broadinstitute site	ker) on data 2 and data	
Phenotype Association	https://software.broadinstitute.org/gatk/download/archive and then move it in the conda_prefix folder, the path of the conda_prefix is written in the	1 :filtered annotated csv	
BED Tools	galaxy.ini(or .ymi) file		● / ×
pre-alignment	more informations at <a href="https://software.broadinstitute.org/gatk/documentation/tooldocs/3.8-">https://software.broadinstitute.org/gatk/documentation/tooldocs/3.8-</a>	ker) on data 2 and data	
DNA-alignment	0/org_broadinstitute_gatk_tools_walkers_variantutils_SelectVariants.php_https://software.broadinstitute.org/gatk/documentation/tooldocs/3.8-	1 :annotated csv	
picard-tools	0/org_broadinstitute_gatk_tools_walkers_filters_VariantFiltration.php		• / ×
vcf-tools	Implemented options SelectVariants & VariantFiltration	ker) on data 2 and data 1 : pdf output	
COVACS	-R Reference sequence file		
covacs Select filtration	-w verification sequence me	22: logfile	• # ×
SelectVariants VariantFiltration	other options are fixed based on covacs pipeline	21: filtered statistic cvs	• 🖋 🗙
wrapper for covacs, use in case of not enough snp or indels	description this step can be used if covacs_indel_snp have not enough data to create the model, it calls SelectVariants and VariantFiltration applying the filter	20: filtered annotated cs	• / ×
error in	expressionfilterExpression IDP<8    QD <.20    FS > 200.0    Read/ObsRanKSum <-20.0° to filter the variants both snp and indels	<u>v</u>	
covacs_VariantRecalibrator		19: annotated csv	
covacs ApplyRecalibrator gatk	Citations 🕼 Show BibTeX	15. amotated Csv	• / ×
ApplyRecalibrator wrapper	Chiara, Matteo and Gioiosa, Silvia and Chillemi, Giovanni and Dac Antonio, Mattia and Flati, Tiziano and Picardi, Ernesto and Zambelli, Federico and	18: pdf output	• 🖋 🗙
covacs VariantRecalibrator	Horner, David Stephen and Pesole, Graziano and CastrignanÄ <sup>2</sup> , Tiziana (2018). CoVaCS: a consensus variant calling system. In BMC Genomics, 19 (1).	2: Mus_musculus.GRCm3	• / ×
GATK VariantRecalibrator wrapper Version = 3.8	[doi:10.1186/s12864-018-4508-1][Link]	8.92.gtf	
covacs HaplotypeCaller GATK		1: Buettner.csv	• / ×

#### CONCLUSIONS

After several months of beta testing, Laniakea ReCaS is reaching its production phase and will be available to researchers through an ELIXIR-IT call.

We expect that this service will offer an useful platform to several classes of end-users, e.g., small research groups, health-operators interested in precision and personalized medicine, SMEs and trainers, by providing custom Galaxy instances over the cloud. We will continue to build on this service and the underlying Laniakea platform in the framework of the EOSC-Life and EOSC-Pillar H2020 projects, working on adding new functionalities and integrating them with ECP and ELIXIR services.

Useful links Laniakea: an open solution to provide Galaxy "on-demand" instances over heterogeneous cloud infrastructures. doi: https://doi.org/10.1101/472464 Documentation: http://laniakea.readthedocs.io GitHub: https://github.com/Laniakea-elixir-it

New Portal: https://laniakea-elixir-it.github.io



Download the Laniakea paper from biorxiv.

# Contact

- Marco Antonio Tangaro ma.tangaro@ibiom.cnr.it
- Federico Zambelli (ELIXIR-ITA technical coordinator) f.zambelli @ibiom.cnr.it
- Giacinto Donvito (INFN Bari Section) giacinto.donvito@ba.infn.it
- Graziano Pesole (head of ELIXIR-ITA Node) g.pesole@ibiom.cnr.it

