



*Secure deployments of Galaxy Servers for analysing personal and Health data leveraging the Laniakea service*



Tangaro MA, Foggetti N, Donvito G, Antonacci M and Zambelli F



International Symposium on Grids & Clouds (ISGC) 2023  
March 19-31, 2023

# Outline



- Legal issues concerning genomic and genetic data
- Tools, data, compute and ... GDPR
- Galaxy
- Laniakea
- Dashboard
- Encryption
- VPN
- Conclusions

# Legal issues concerning genomic and genetic data in Laniakea



- Application of GDPR: art. 9: Particular category of Data (Genetic, genomic and other typology of data)
- The Convention on Human Rights and Biomedicine (article 10), the Universal Declaration on the Human Genome and Human Rights (article 5, letter c), and UNESCO's International Declaration on Human Genetic Data (article 10)
- ELSI (Ethical, Legal, Social implication document)
- Authorization n. 8/2016 – General authorization concerning genetic data processing 15 December 2016:
  - Pseudonymization
  - Use of SSH protocol for sharing data
  - Specific measures of identification
  - High level of security measures in order to protect and to prevent cyber attack

# Use case application

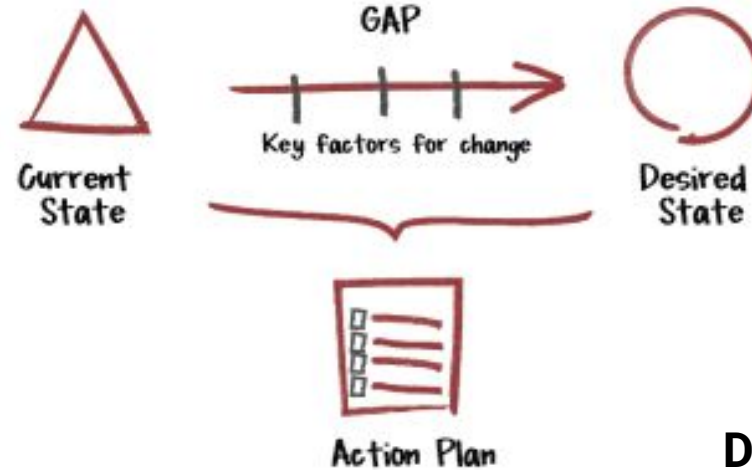
The use case of Lanikea has been analysed also in INFN Cloud



- Definition of Policy for services (mandatory in EU)
- definition of ToU and AUP harmonized
- Possible application of ISO standard in order to improve the legal compliance



# Gap Analysis



**Fragmentation** of applicable law and jurisdiction

**IPR Exception and Limitations:**  
blinding list and not harmonized

**Differences between different law at EU Level concerning specific sectors**  
(e.g., health, genetics)

Harmonization of **pseudonymization e anonymization procedures**

**Data protection:**  
Transnational data processing:  
conflict of laws and sectoral law

**Discrimination defined by the national application of EU Law**

**Common standards and interoperability**

# Guidelines

## Objectives

To Increase the level of awareness in the context of processing data and resources protected by IPR



To Identify the legal framework applicable in the field of data protection and IPR

To manage DMP/OA/OS

Target: Reserarchers

# The Checklist

- See “Legal Compliance Guidelines for Researchers: a Checklist”
  - Both digital (with interactive checkboxes, <https://doi.org/10.5281/zenodo.632766>), and printable (<https://doi.org/10.5281/zenodo.6327691>)
- To guide researchers in management of research outputs vis-à-vis IP and data
- protection laws
- To promote best practices to achieve FAIR ecosystem, removing unnecessary restrictions to reuse and access + facilitating convergences of national solutions

# The Checklist

## RESEARCH PROPOSAL 1

Background information,  
IPR, Exploitation, TTO,  
FAIR, DPIA.

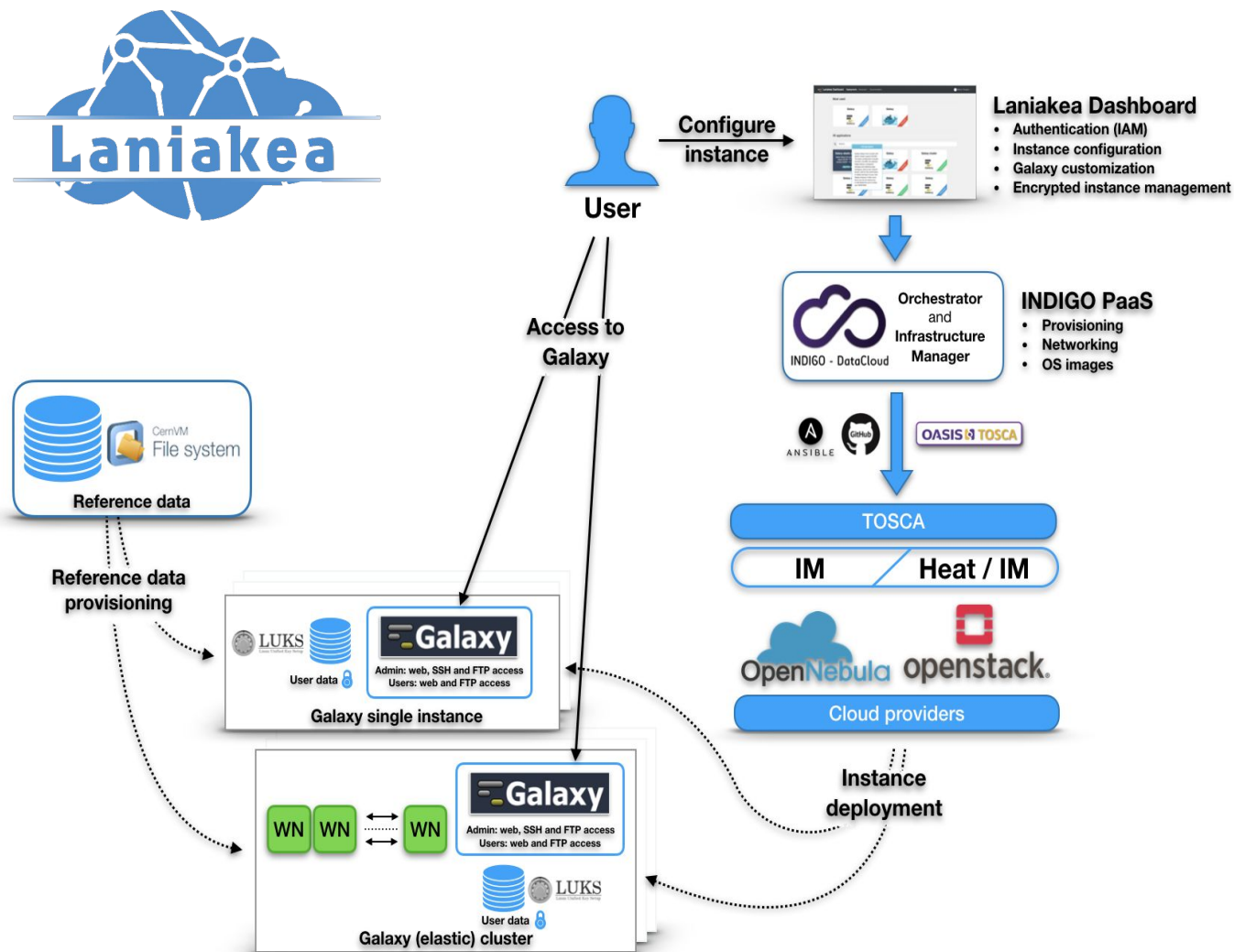
## RESEARCH IMPLEMENTATION 2

IP Management Plan, IP  
FLEXIBILITY, GDPR, DMP,  
Findable, Accessible,  
Re-Usable, Interoperable.

## RESEARCH REVIEW 3

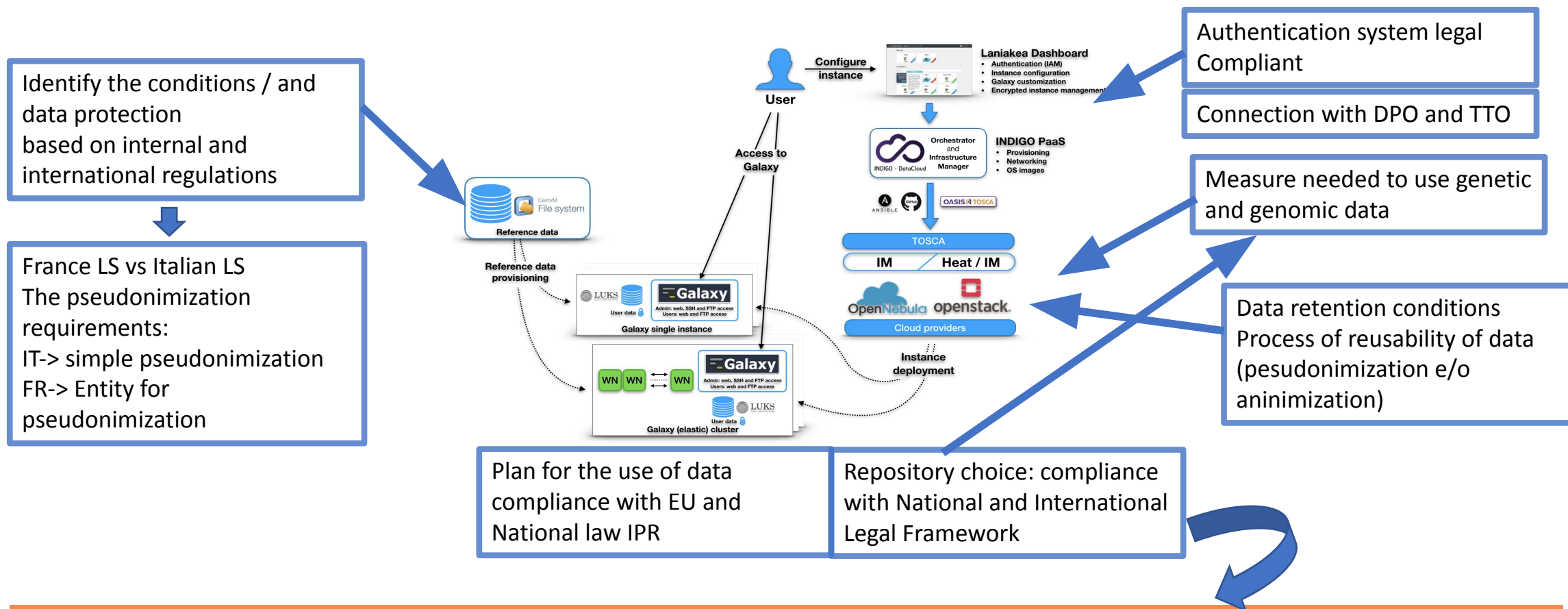
IP Management Plan,  
Licences and  
FAIR/OS/OA, Re-use of  
Data.

# Laniakea architecture



- **Dashboard** - User friendly access to configuration and and launch of a Galaxy instance.
- **IAM** - Authentication and Authorization system.
- **INDIGO PaaS** - Galaxy automatic deployment.
- **Cloud Providers** - (INFN) ReCaS-Bari and others.
- **Persistent storage** - With/without encryption.
- **Reference data availability** - With CERN-VM FileSystem.
- **CLUES** - Elasticity manager.

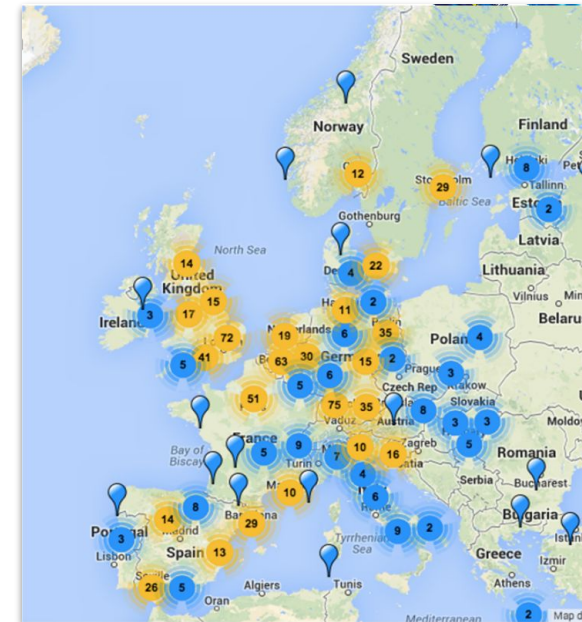
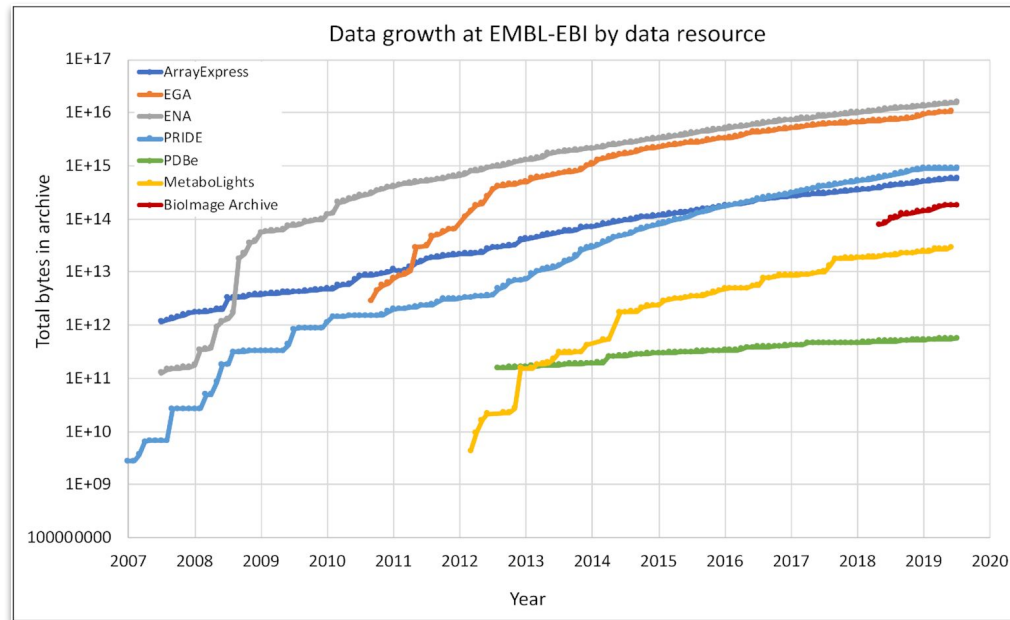
# Legal issue and the application of Checklist



Legal Framework for the use and re-use of health data for scientific purposes

DOI: 10.5281/zenodo.6334878

# Data



Genomic data are distributed across several sequencing centres and/or IT infrastructures

Data volume growing not only in quantity but also on variety!

Data growth at EMBL-EBI Source: Charles E. Cook et al. Nucl. Acids Res. 2020; Volume 48, Issue D1, Pages D17-D23

Discipline	Data size	# devices
HEP-LHC	15PB/year	1
Astronomy	15PB/year	several
Genomics	0.4TB/genome	>1000







# GDPR

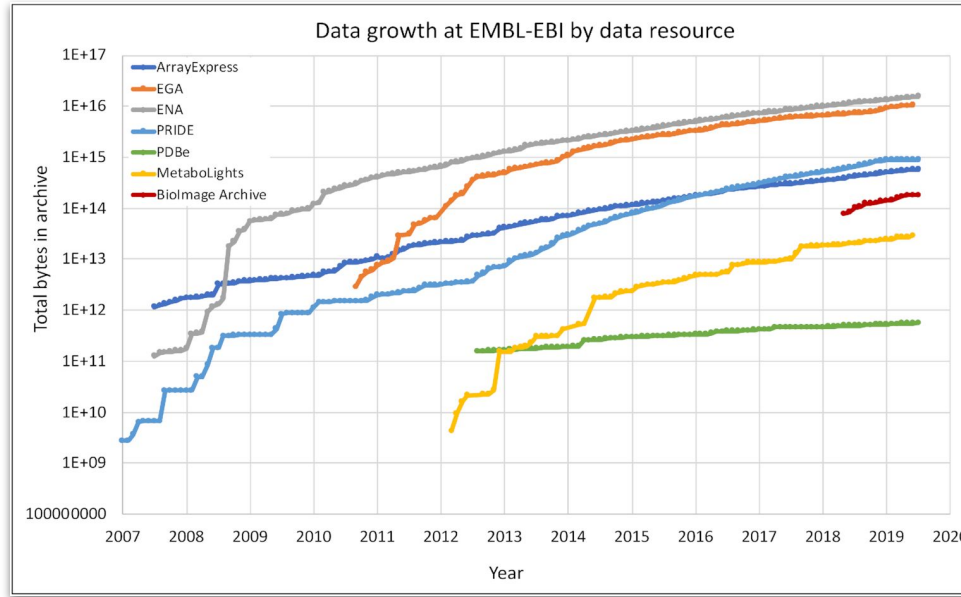
The GDPR explicitly recognizes genetic data as “Special Categories of Personal data”.

Sensitive genetic data processing for scientific research purposes is possible, provided this is allowed by EU or Member States law framework and appropriate safeguards measures are in place.



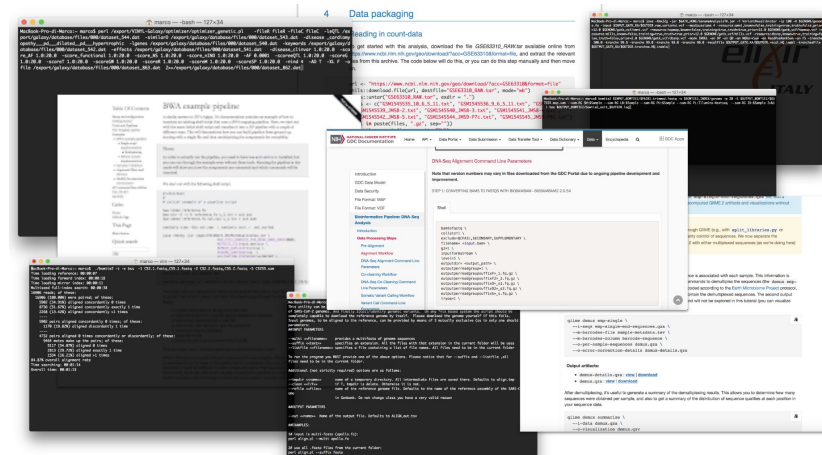
# Motivation

## DATA SOURCES



## DATA STORAGE

## DATA ANALYSIS TOOLS



## DATA PROTECTION (GDPR)



The Galaxy Project logo, consisting of a stylized icon of three horizontal bars (two dark grey, one yellow) to the left of the word "Galaxy" in a large, bold, dark grey font, with the word "PROJECT" in a smaller, spaced-out, dark grey font below it.

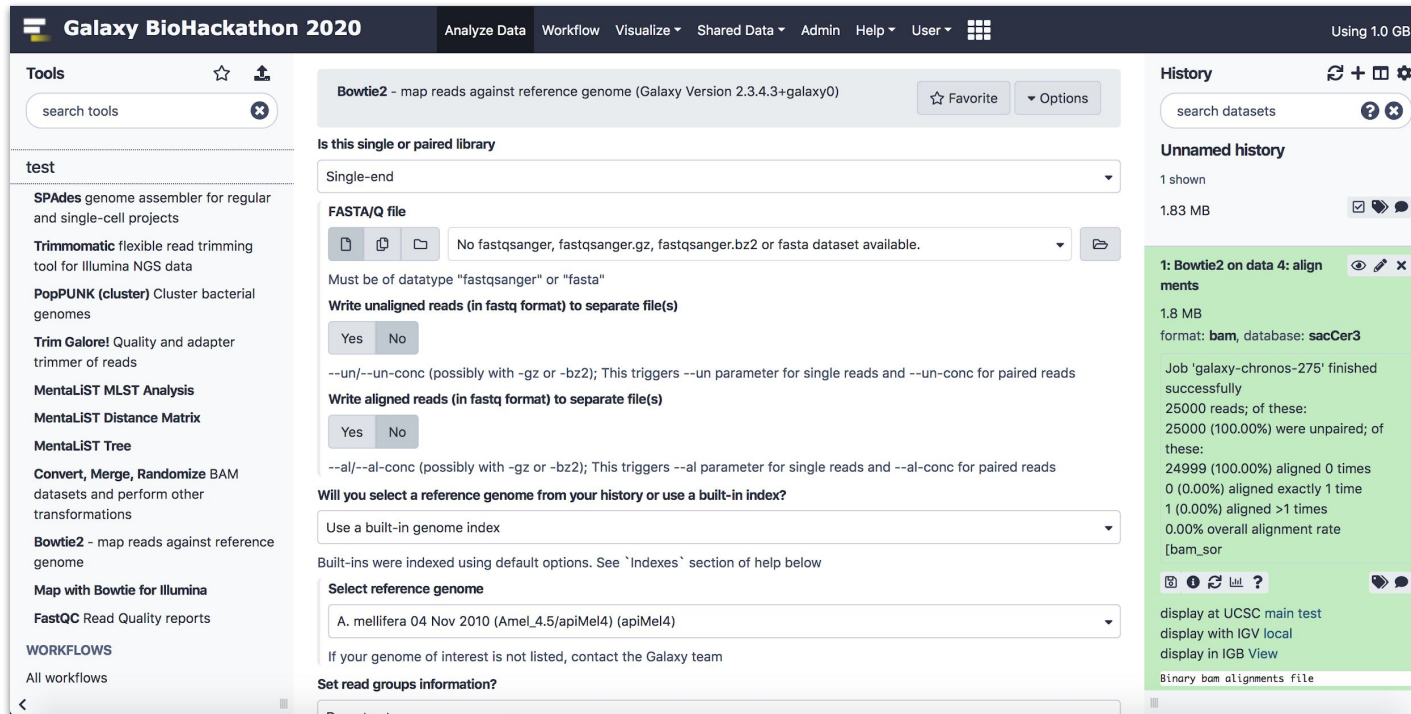
# Galaxy

## PROJECT

**Galaxy is a workflow manager** adopted in many life science research environments in order to facilitate the interaction with bioinformatics tools and the handling of large quantities of biological data.

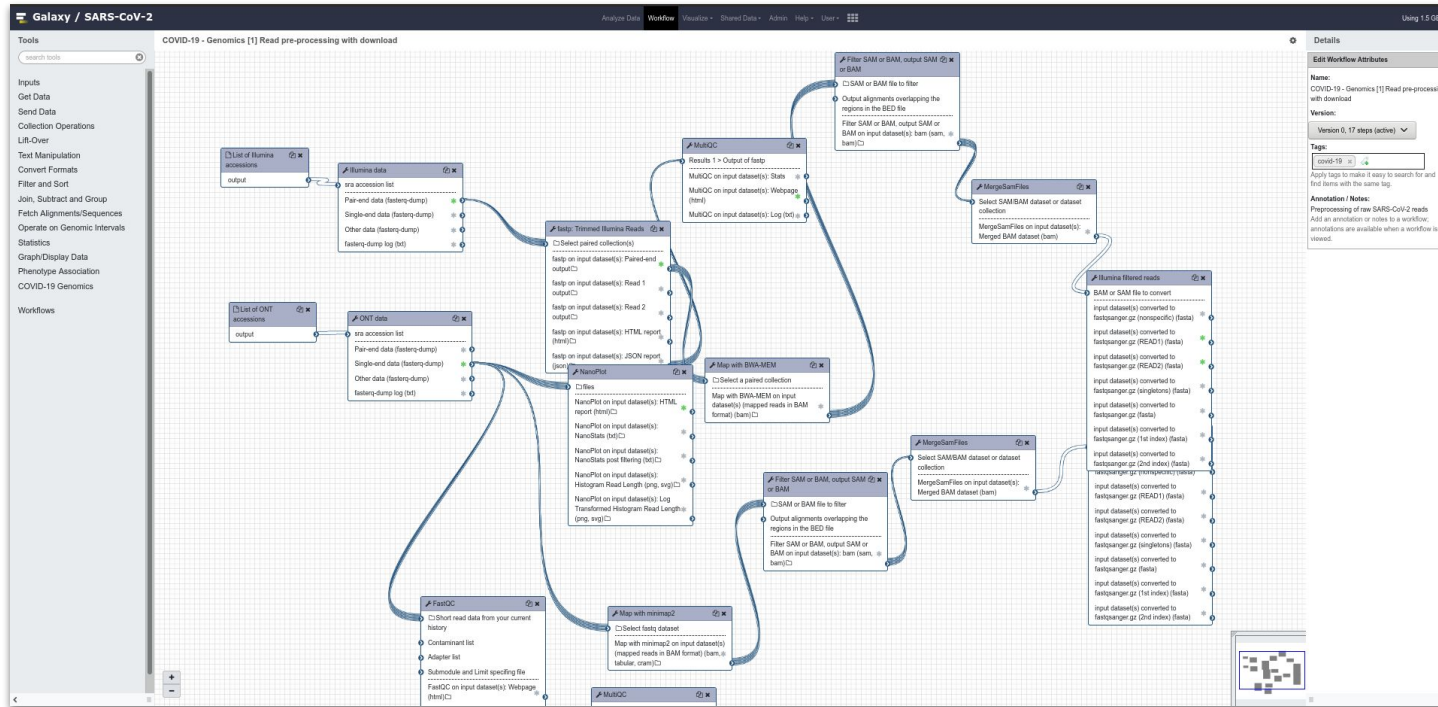
Through a coherent work environment and an **user-friendly web interface** it organizes data, tools and workflows providing **reproducibility, transparency** and **simple data sharing** functionalities to users.

[galaxyproject.org](http://galaxyproject.org)



The screenshot shows the Galaxy BioHackathon 2020 interface. The main tool is Bowtie2, configured for mapping reads against a reference genome. The configuration includes a dropdown for 'Single-end' library type, a 'FASTA/Q file' input field, and options for writing unaligned and aligned reads. A job history panel on the right shows a completed job '1: Bowtie2 on data 4: alignments' with a size of 1.8 MB. The job details indicate that 25,000 reads were unpaired, and 24,999 reads (100.00%) were aligned 0 times, with 1 read aligned exactly 1 time and 1 read aligned more than 1 time. The overall alignment rate is 0.00%.

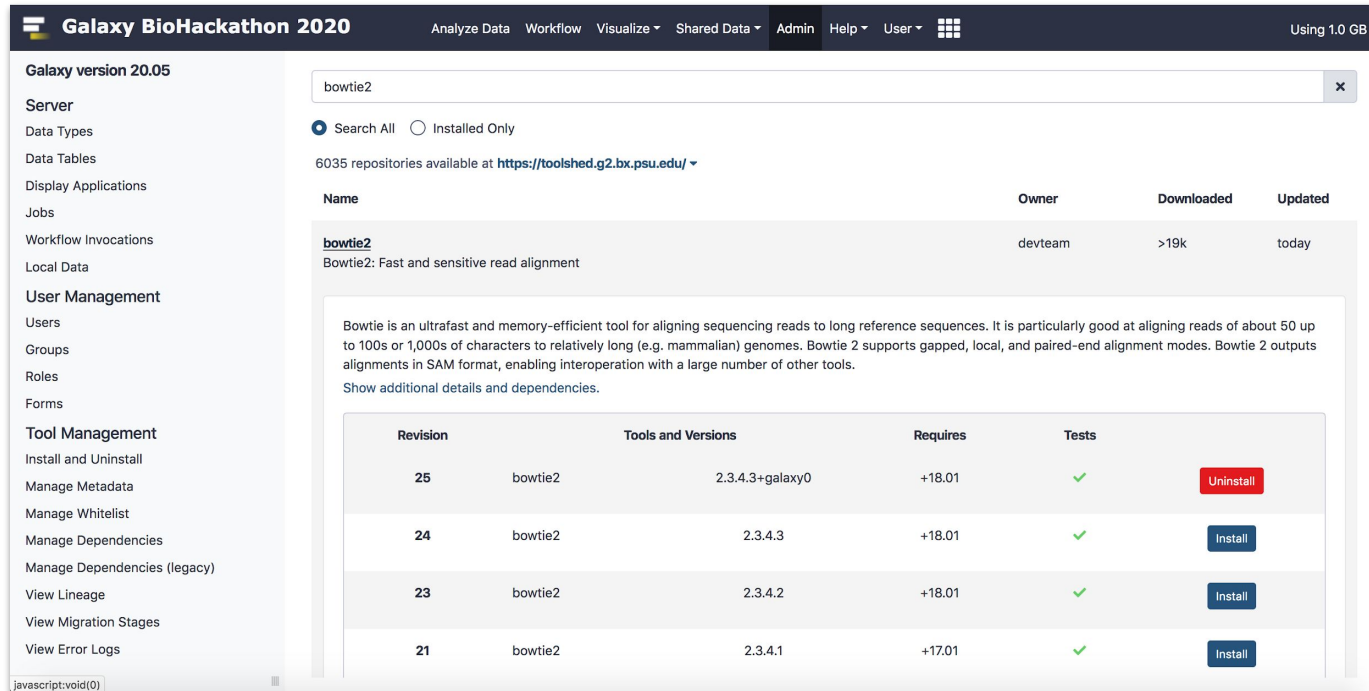
- Tools graphical user interface.
- Input and output data management.
- Output visualization.
- Data and analysis parameters sharing.
- Used tools and parameters configuration always available -> **analysis reproducibility.**
- Reference data already available for many tools.



## Galaxy Workflow Editor

Graphical user interface to easily add, connect and configure tools for composing workflows.





Galaxy BioHackathon 2020 | Analyze Data | Workflow | Visualize | Shared Data | Admin | Help | User | Using 1.0 GB

Galaxy version 20.05

Server

Data Types

Data Tables

Display Applications

Jobs

Workflow Invocations

Local Data

User Management

Users

Groups

Roles

Forms

Tool Management

Install and Uninstall

Manage Metadata

Manage Whitelist

Manage Dependencies

Manage Dependencies (legacy)

View Lineage

View Migration Stages

View Error Logs

Search: bowtie2

Search All  Installed Only

6035 repositories available at <https://toolshed.g2.bx.psu.edu/>

Name	Owner	Downloaded	Updated
<b>bowtie2</b> Bowtie2: Fast and sensitive read alignment	devteam	>19k	today

Bowtie is an ultrafast and memory-efficient tool for aligning sequencing reads to long reference sequences. It is particularly good at aligning reads of about 50 up to 100s or 1,000s of characters to relatively long (e.g. mammalian) genomes. Bowtie 2 supports gapped, local, and paired-end alignment modes. Bowtie 2 outputs alignments in SAM format, enabling interoperability with a large number of other tools.

Show additional details and dependencies.

Revision	Tools and Versions	Requires	Tests	
25	bowtie2 2.3.4.3+galaxy0	+18.01	✓	Uninstall
24	bowtie2 2.3.4.3	+18.01	✓	Install
23	bowtie2 2.3.4.2	+18.01	✓	Install
21	bowtie2 2.3.4.1	+17.01	✓	Install

## Galaxy ToolShed

Serves as an "app store" to all Galaxies worldwide.

It is a **free service** Galaxy developers to share tools.

Galaxy Administrator can install tools on their instances.

Tools dependencies automatically solved.

**All Galaxy users can access to the tools available on a server.**

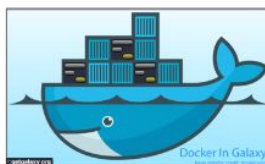
# Laniakea

LANIAKEA IS A CLOUD BASED GALAXY INSTANCE PROVIDER

<https://laniakea-elixir-it.github.io/>

- Laniakea relies on commonly used Life Science Open Source tools, e.g. Galaxy, RStudio, Jupyter, HashiCorp Vault, LUKS and SLURM.
- Laniakea is European Open Science Cloud service provider.

Recommended for scenarios where users need full administrative control over a private Galaxy instance.

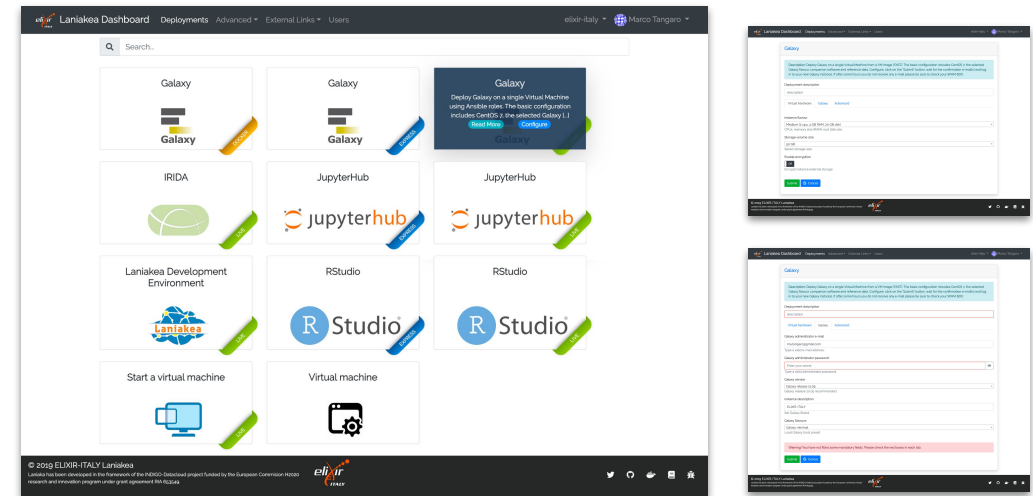


# Laniakea main features

**Dashboard** - By hiding the technical complexity behind a user-friendly web front-end, Laniakea allows its users to configure and deploy “on-demand” Galaxy instances with a handful of clicks.

No need for the end user to know the underlying infrastructure.

No need for maintenance of the hardware and software infrastructure.



## Different deployment strategies:

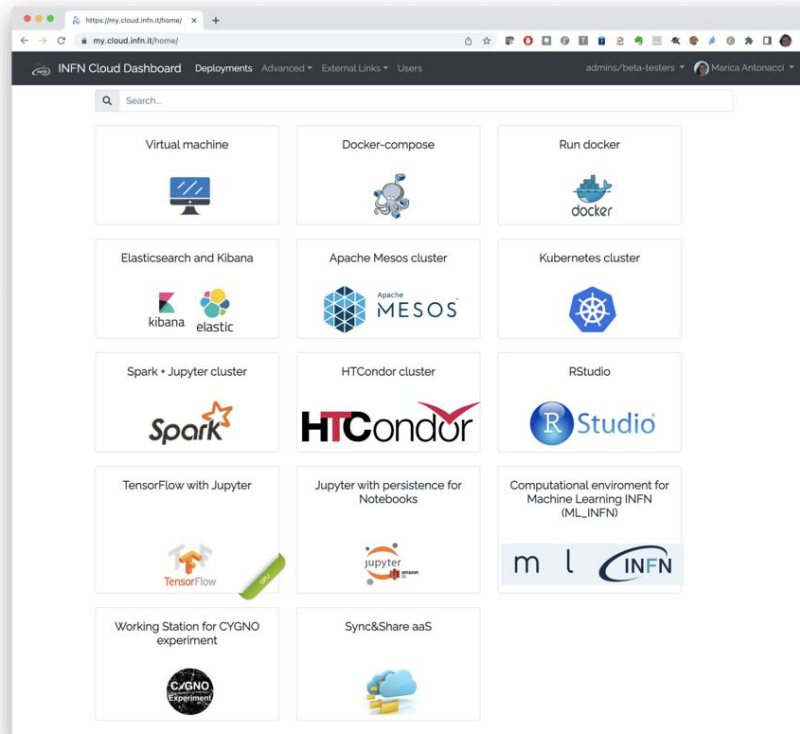
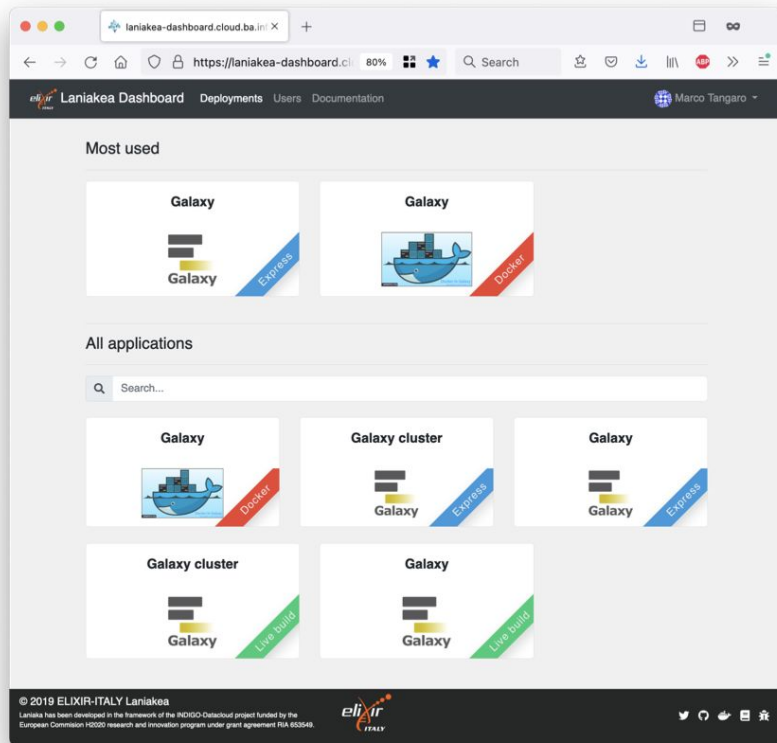
**Live Build**: build Galaxy from scratch -> always up-to-date (deployment time depending by the tools number).

**Express**: pre-built Galaxy images -> fast deployment, but tools not always at the last available version.

**Docker**: fast development of new flavours.



# Laniakea main features



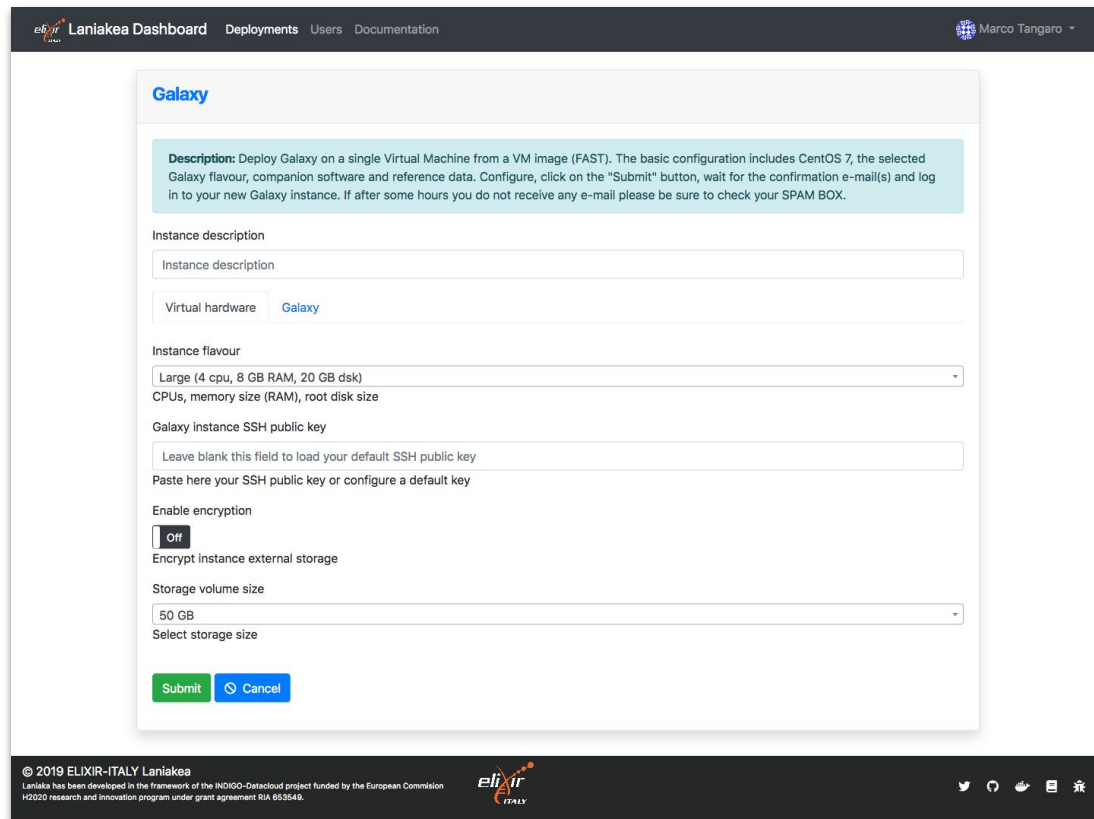
Flask web micro-framework  
([flask.pocoo.org/](https://flask.pocoo.org/)),

Jinja2 template engine  
([jinja.pocoo.org/](https://jinja.pocoo.org/))

Bootstrap 4 toolkit  
([getbootstrap.com/](https://getbootstrap.com/)).

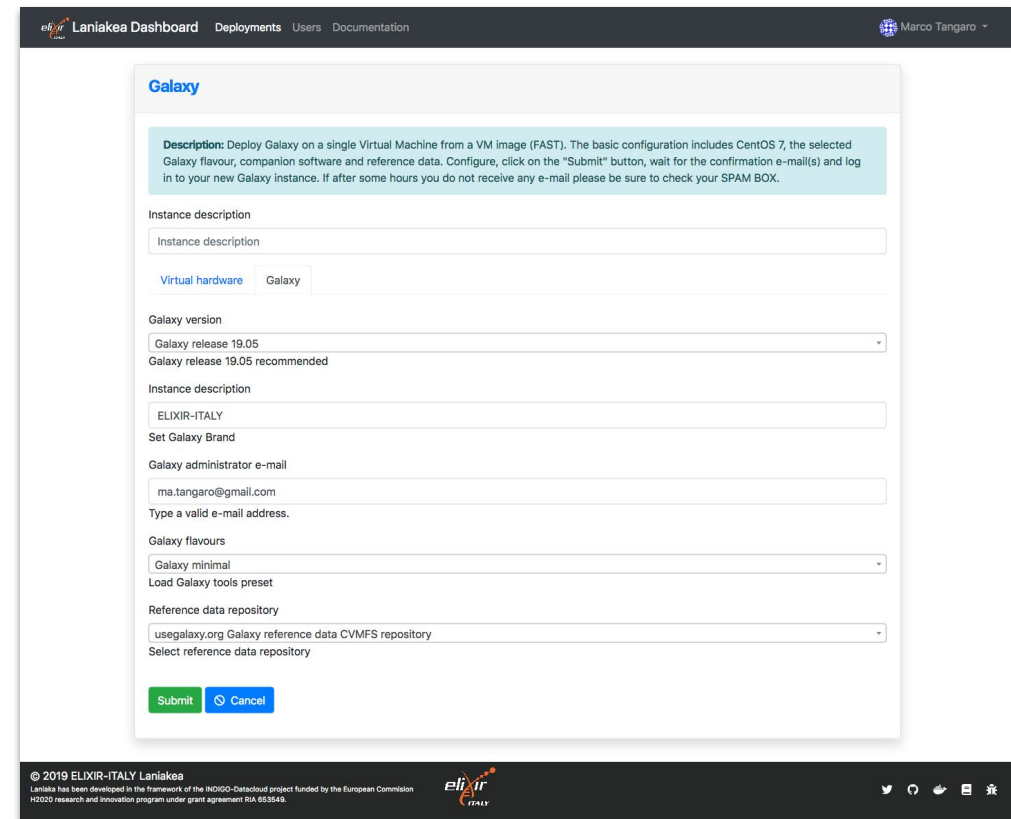
Integrated with Hashicorp  
Vault for user secrets  
management.

# Laniakea main features



The screenshot shows the 'Galaxy' configuration page in the Laniakea Dashboard. It includes a description box, an 'Instance description' text field, a 'Virtual hardware' dropdown menu set to 'Galaxy', an 'Instance flavour' dropdown menu set to 'Large (4 cpu, 8 GB RAM, 20 GB dsk)', an SSH public key text area, an 'Enable encryption' toggle set to 'Off', an 'Encrypt instance external storage' checkbox, a 'Storage volume size' dropdown menu set to '50 GB', and a 'Select storage size' dropdown menu. At the bottom are 'Submit' and 'Cancel' buttons.

Virtual hardware: CPU,  
RAM and Storage



The screenshot shows the 'Galaxy' configuration page in the Laniakea Dashboard, continuing from the previous view. It includes a 'Galaxy version' dropdown menu set to 'Galaxy release 19.05', an 'Instance description' text field containing 'ELIXIR-ITALY', a 'Set Galaxy Brand' text field, a 'Galaxy administrator e-mail' text field containing 'ma.tangaro@gmail.com', a 'Galaxy flavours' dropdown menu set to 'Galaxy minimal', a 'Load Galaxy tools preset' dropdown menu, and a 'Reference data repository' dropdown menu set to 'usegalaxy.org Galaxy reference data CVMFS repository'. At the bottom are 'Submit' and 'Cancel' buttons.

Galaxy software: version,  
credentials, flavor and  
reference data.

# Laniakea main features

**Galaxy flavors** - Deploy Galaxy with sets of tested, validated and pre installed tools, named Galaxy flavors.

Current available tools presets: Galaxy Minimal, Galaxy CoVaCS, Galaxy GDC Somatic Variant, RNA Workbench, Galaxy Epigen, Covid-19.

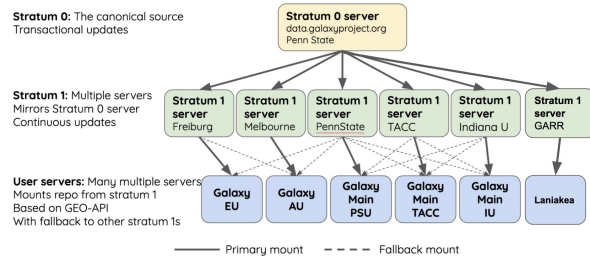


**More Applications** - No more limited to Galaxy. Jupyter Notebooks, RStudio and IRIDA available.

Environment with NextFlow, CWLtool and other development tools available.



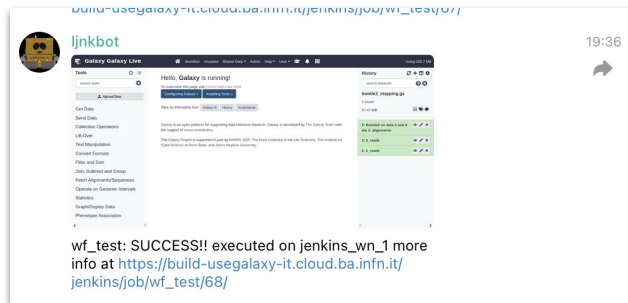
# Laniakea main features



**Shared reference data** - Each instance comes with reference data (e.g. genomic sequences) already available for many species, shared among all the instances through the CERN-VM FileSystem .



**Galaxy with cluster** - allowing to instantiate Galaxy with dedicated Resource Manager, allowing to customize the number of the virtual nodes to be created and their configuration in terms of number CPU and RAM.



**Continuous Integration** - Cloud images creation and service testing are automated using Jenkins and Github.

Cloud images automatically updated when there are changes on github  
Services are tested regularly every week.

# Secured environment features

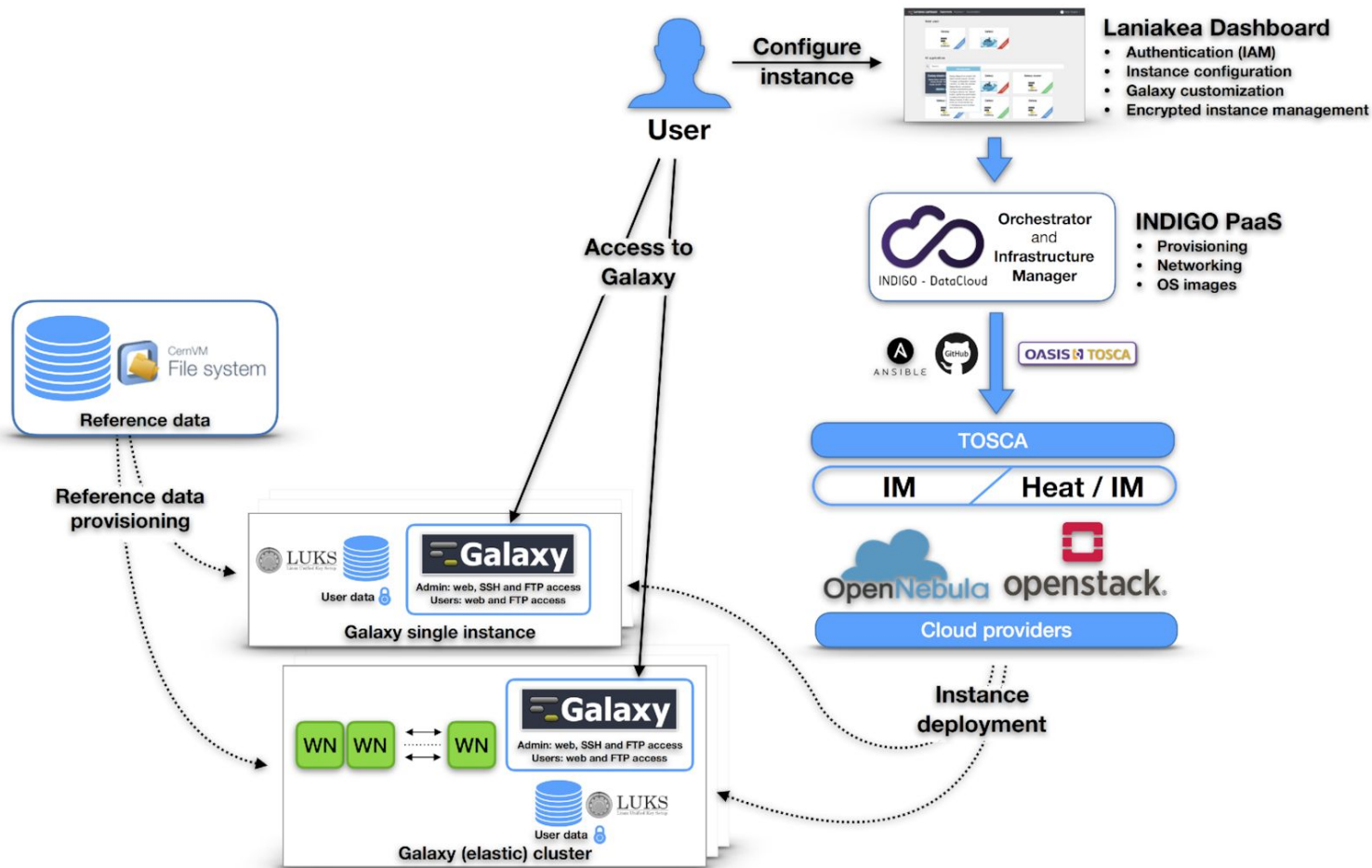
**Storage Encryption** - Data privacy is provided through encryption "on-demand".



**OPENVPN**

**Deployments under Private Network** - Automatic deployments of virtual environments on private networks.

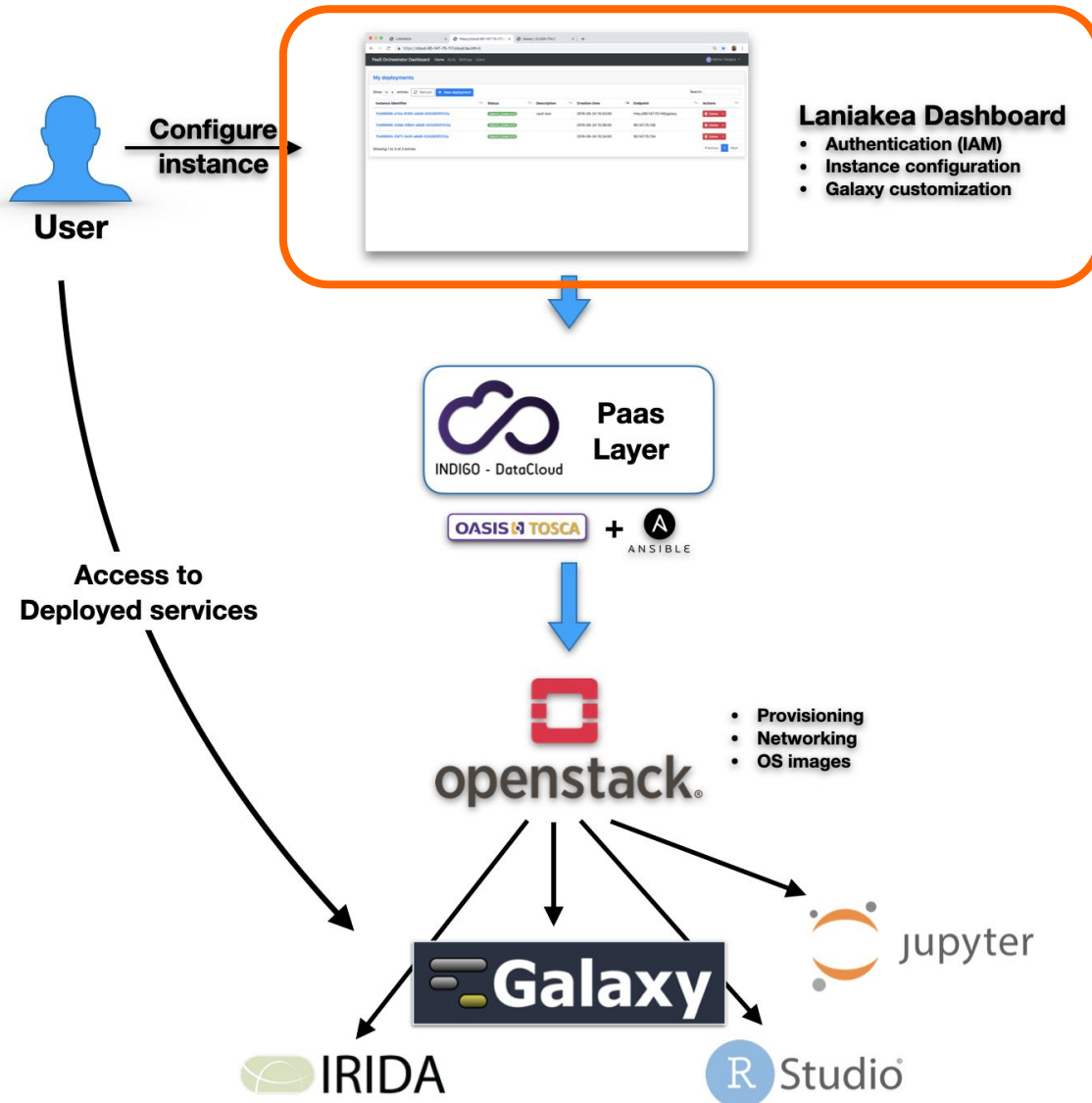
# Laniakea architecture



- **Dashboard** - User friendly access to configuration and launch of a Galaxy instance
- **INDIGO-IAM** - Authentication and Authorization system
- **INDIGO-PaaS** - PaaS layer for Galaxy deployment
- Cloud Provider - ReCaS Bari
- **Persistent storage** with/without encryption
- **Hashicorp Vault** - secrets management
- **Reference data** availability with CERN-VM FS

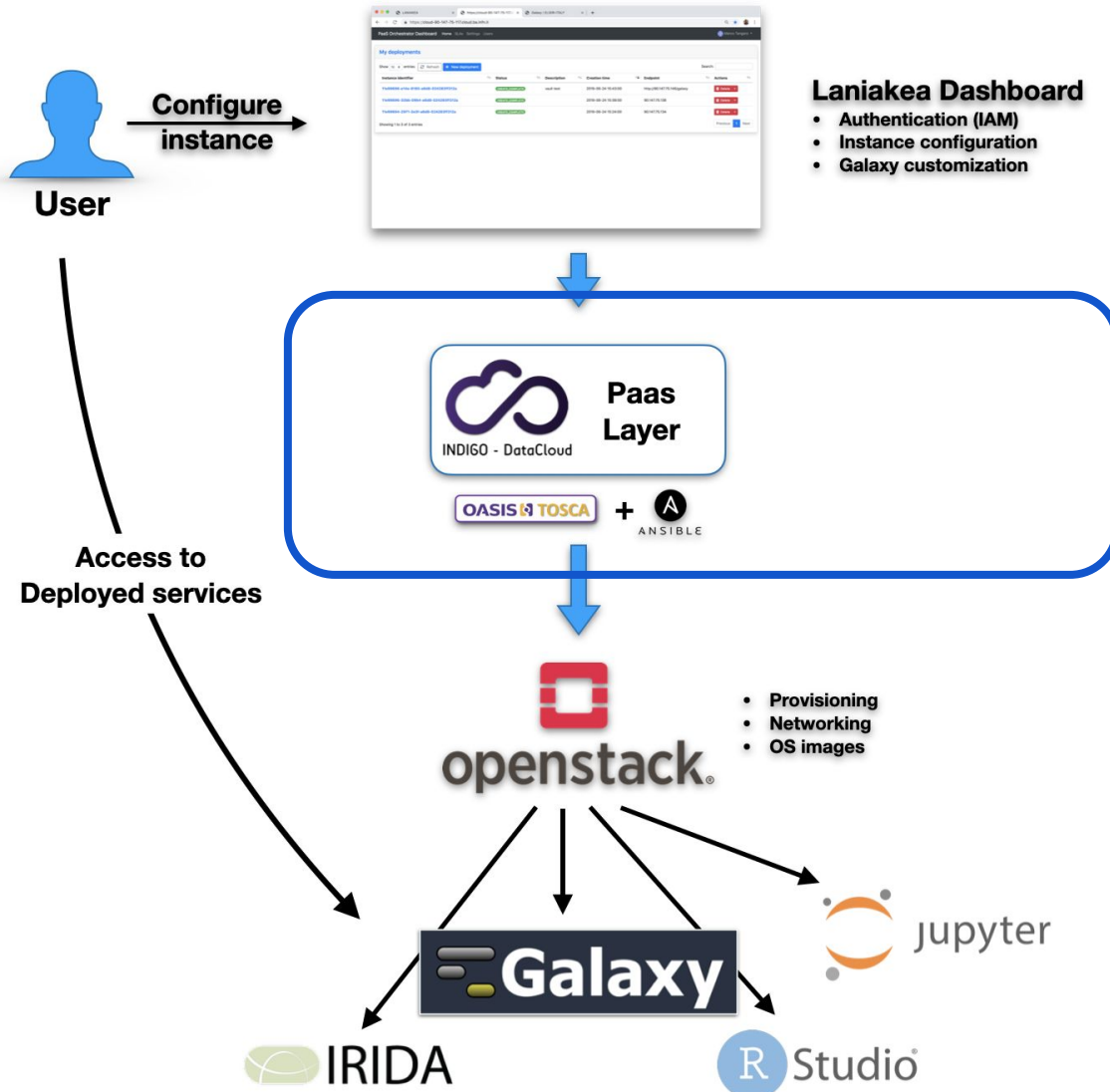


# Laniakea architecture



- **Dashboard** - User friendly access to configure and launch a Galaxy instance
- **INDIGO PaaS** - Galaxy automatic deployment
- **Cloud Providers** - ReCaS-Bari

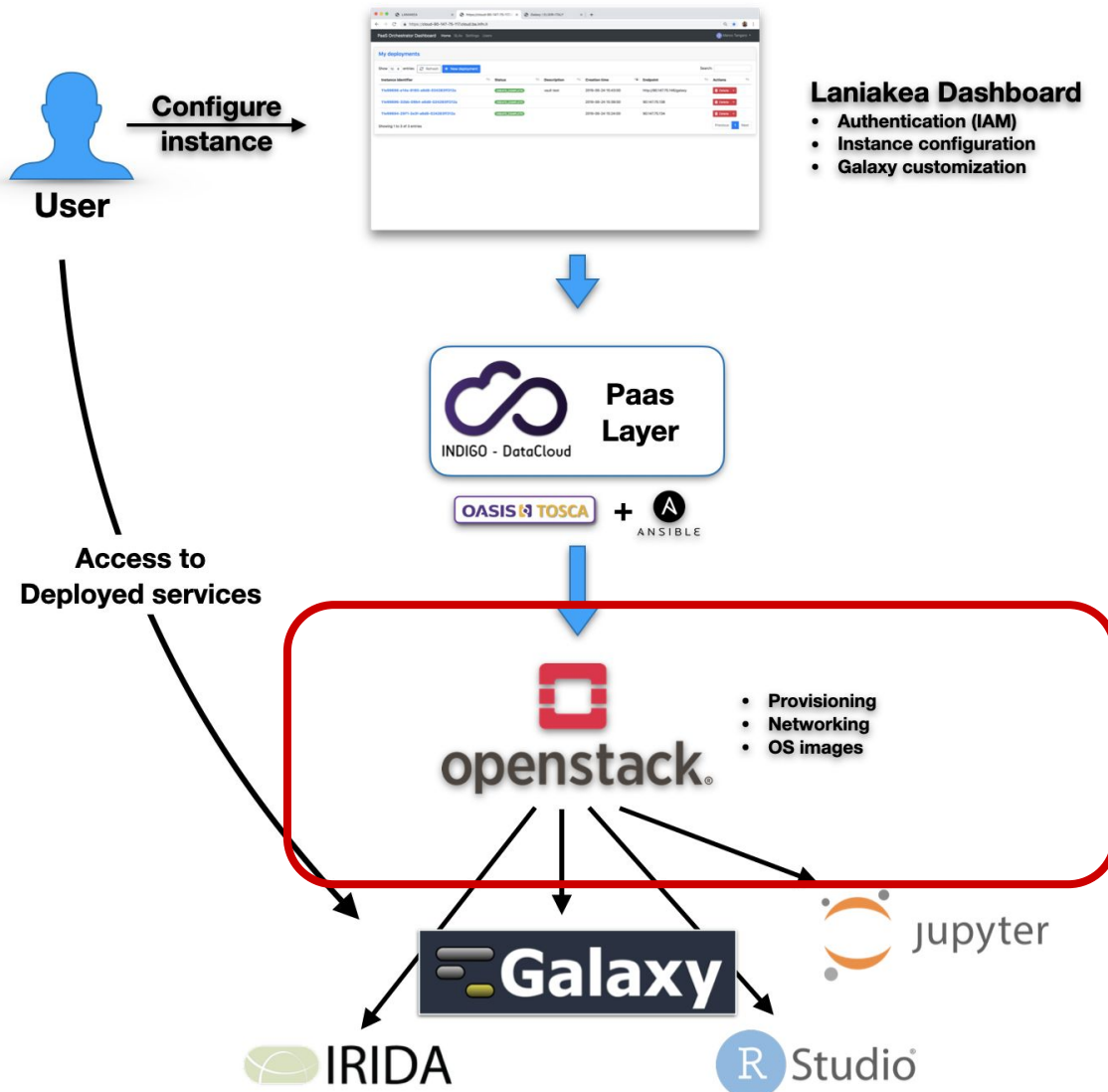
# Laniakea architecture



- **Dashboard** - User friendly access to configure and launch a Galaxy instance
- **INDIGO PaaS** - Galaxy automatic deployment
- **Cloud Providers** - ReCaS-Bari



# Laniakea architecture



- **Dashboard** - User friendly access to configure and launch a Galaxy instance
- **INDIGO PaaS** - Galaxy automatic deployment
- **Cloud Providers** - ReCaS-Bari

# *Laniakea encryption*

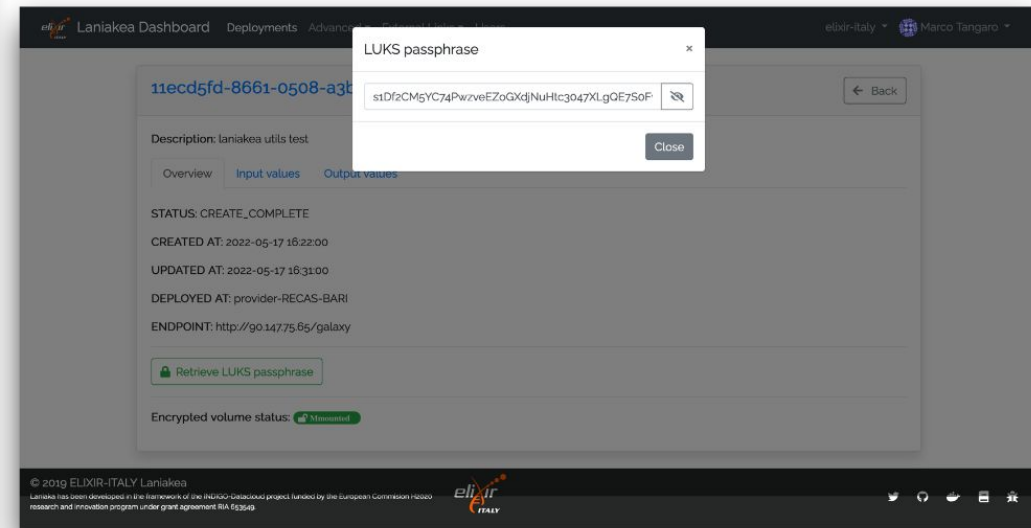
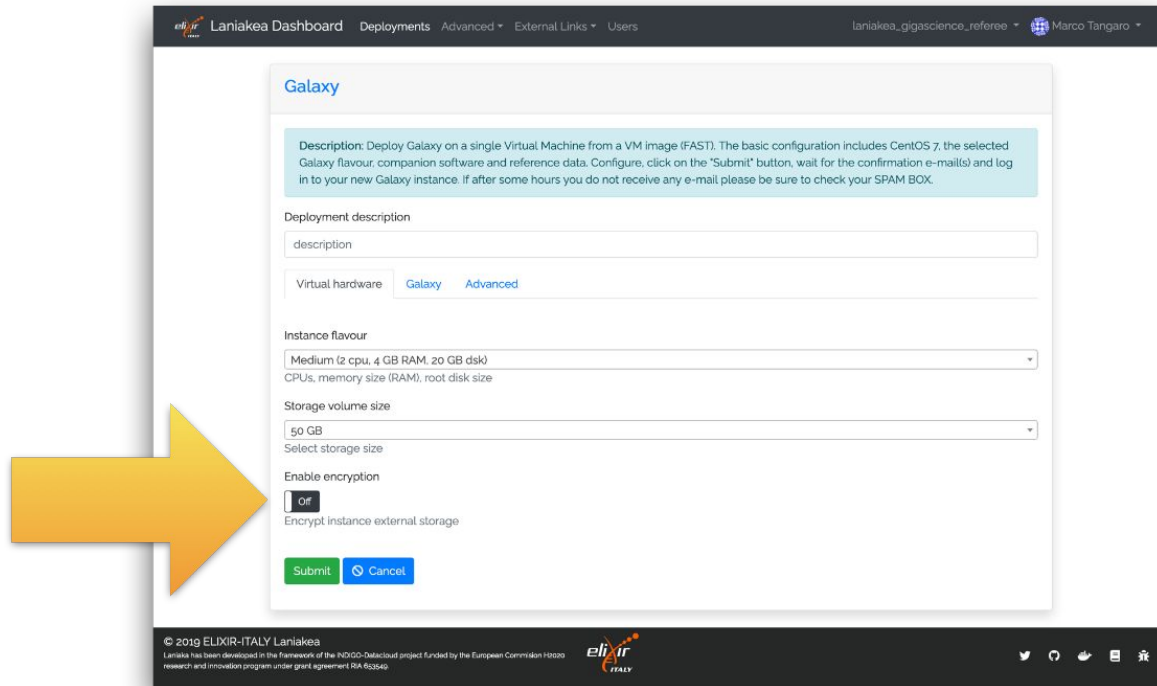
# *Laniakea encryption*



The user data privacy is granted through **LUKS** storage encryption as a service: the encryption procedure is automated in order to simplify the user experience, each user can encrypt storage on-demand, using a strong random alphanumerical passphrase.

This has been achieved integrating the Dashboard and the key management system **Hashicorp Vault** ([vaultproject.io](https://vaultproject.io)) to store encryption keys, which are shown in the Laniakea Dashboard only if explicitly requested by the user.

# User perspective



The user can enable the storage encryption using a switch toggle in the Instance “Virtual hardware” configuration tab.

The procedure is completely automated.

The storage is encrypted and the User can retrieve his random passphrase from the Instance overview page.

# The underlying infrastructure

## LUKS - Linux Unified Kernel Setup

A python package (pyLUKS) is used to encrypt the storage using a random passphrase and then store it on **Hashicorp Vault**.

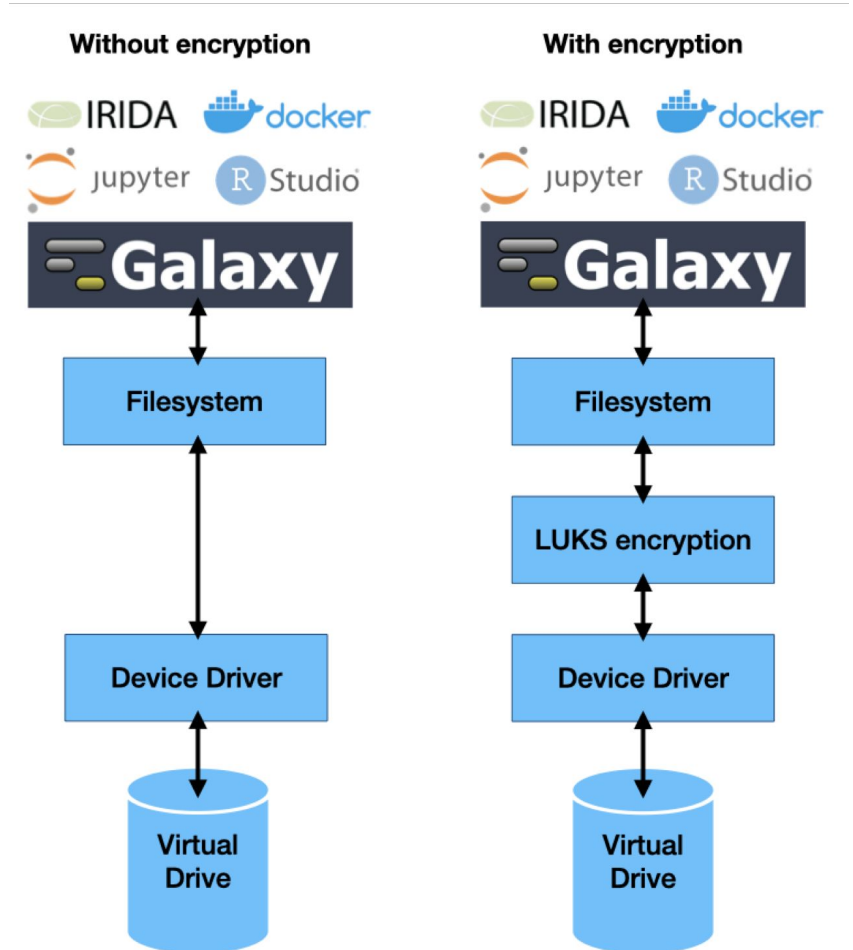
The encryption layer sits between the physical disk and the file system.

Galaxy, or any other application, is unaware of storage encryption.

Galaxy exploits a specific mount point in order to store and retrieve files. Files are encrypted when stored to disk and decrypted when read.

Default encryption algorithm:

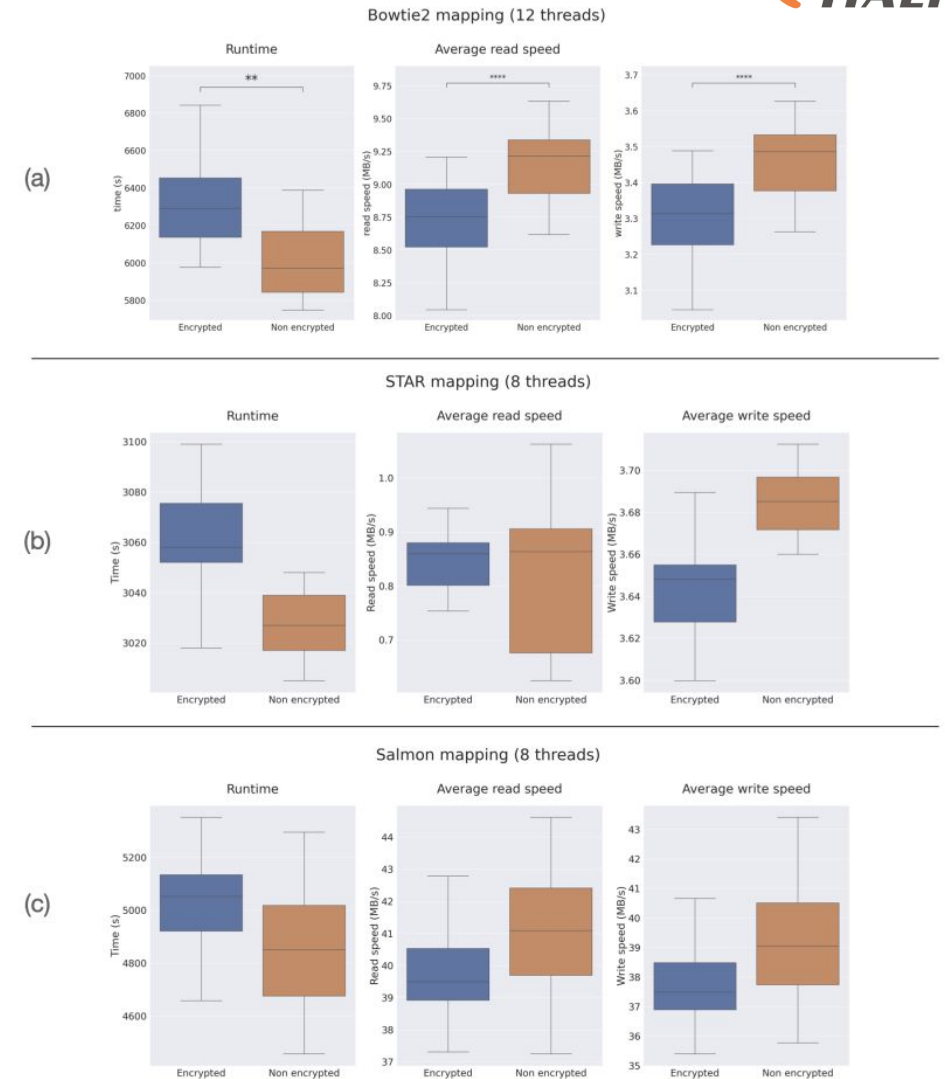
- aes-xts-plain64 encryption
- 256 bit key
- sha256 as hash algorithm used for key derivation.



# The underlying infrastructure

To evaluate the impact of the storage encryption layer on the performance of the main application supported by Laniakea, i.e., Galaxy, we measured jobs runtime and read/write speed on Virtual Machines generated by the Laniakea@ReCaS data center with and without storage encryption.

The impact on the performance of using the encryption layer, as measured in all our tests, is limited to ~5% or less across all the measured parameters and conditions.



# *The underlying infrastructure*

## Key management: Vault introduction

Vault is a tool for securely accessing “secrets”.

A secret is everything you want to tightly control access to, such as encryption passphrases.

Vault main concepts:

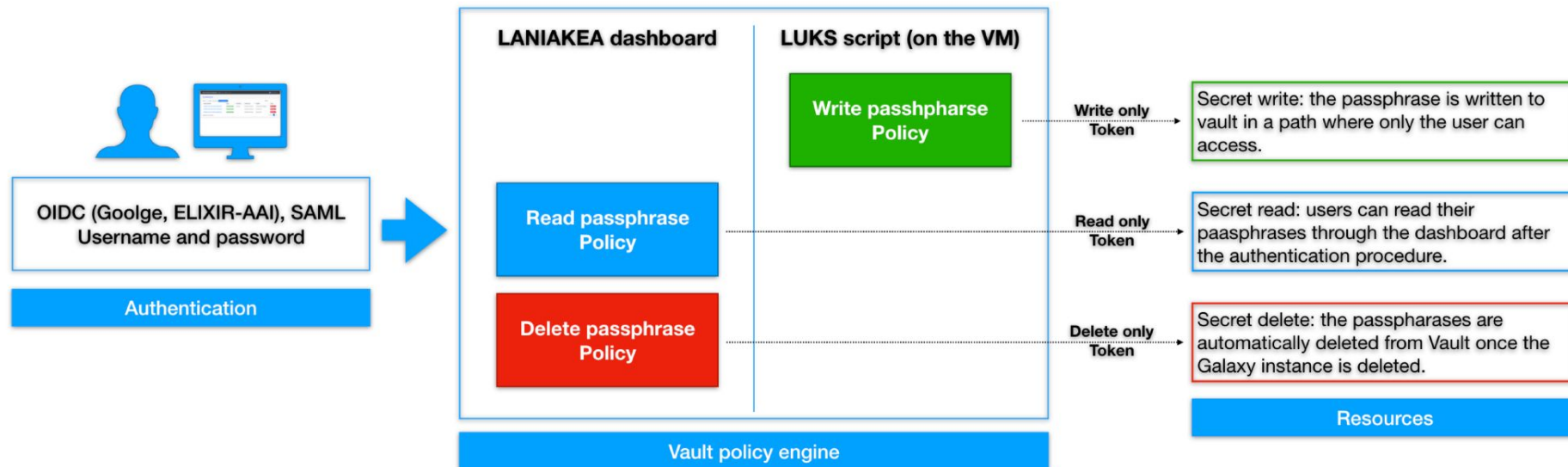
- Everything in Vault is path based: users are able to write their secrets on a specific path, **depending on their Identity**.
- Tokens are the core method for authentication within Vault. After the authentication on the Laniakea Dashboard, **tokens are dynamically generated based on user identity**.
- Policies provide a declarative way to grant or forbid access to certain path and operations, **controlling what the token holder is allowed to do within Vault**.

**A token generated with a specific policy allows to write/read/update a secret in a specific path!**

# The underlying infrastructure

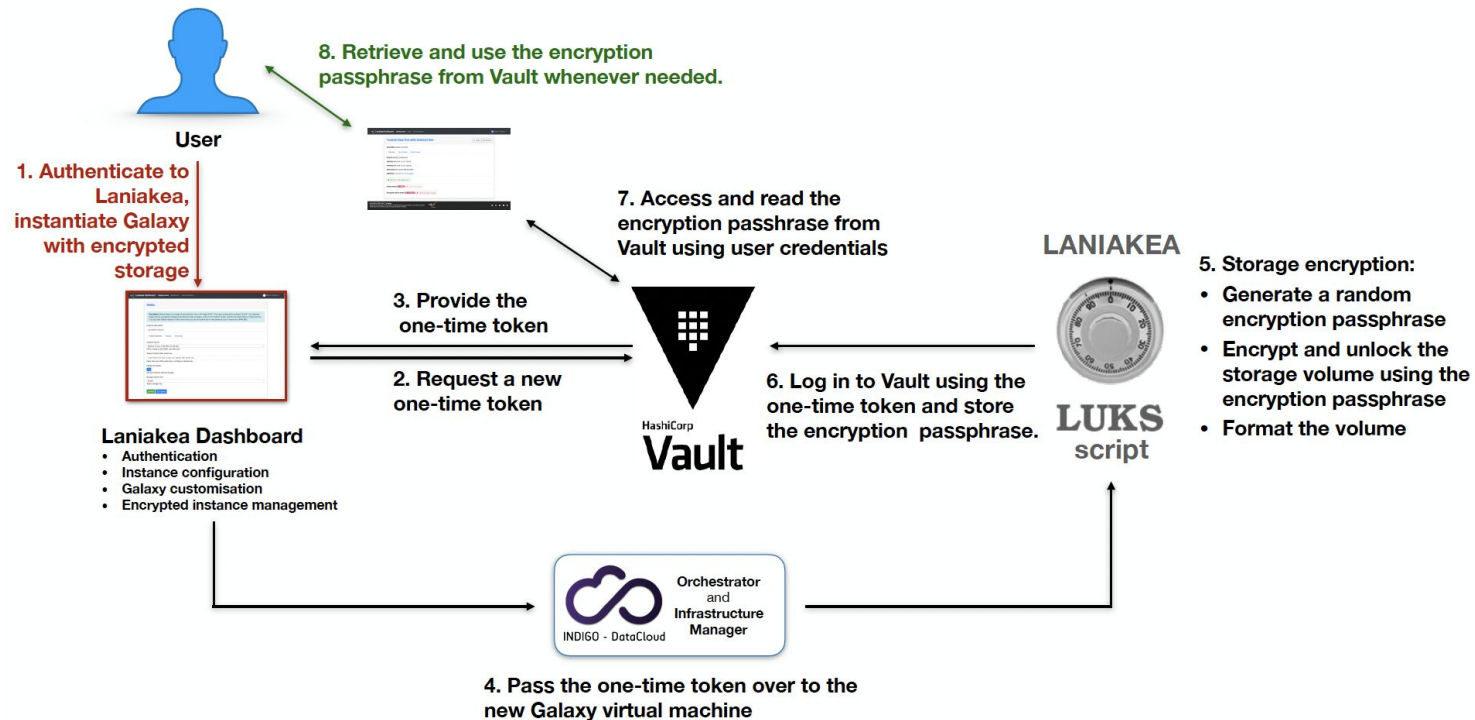
## Key management: Vault authentication and authorization flow

- The “**write only**” token is exploited by LUKS script to store passphrases on Vault.
- The Laniakea Dashboard can **Read**, if required by the user, after the authentication, the passphrase from Vault.
- The Laniakea Dashboard **Deletes** the passphrase from Vault, once the deployment is deleted.





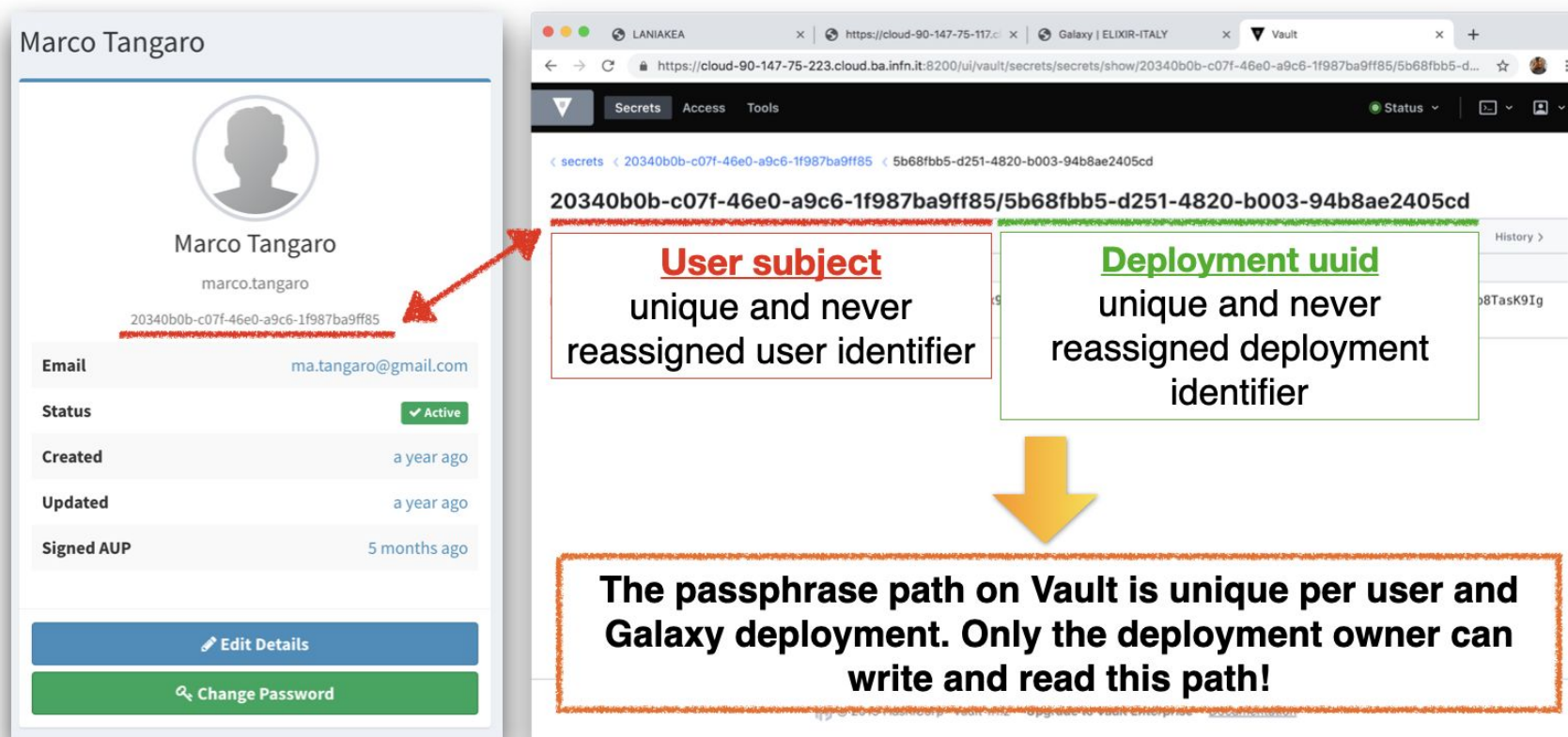
# The underlying infrastructure



1. User Authentication.
2. **A short lived, write only token, usable only once**, is delivered to the Laniakea encryption script on the VM. There's no update policy: this token can't overwrite other passphrases for security reasons.
3. The Storage volume is encrypted by Laniakea pyLUKS package.
4. The passphrase is sent to Vault by Laniakea pyLUKS.
5. After the instance has been successfully deployed the user can retrieve his password through the Dashboard.
6. The user reads the password on the Dashboard.

# The underlying infrastructure

## Key management: passphrase path on Vault



The image shows two side-by-side screenshots. The left screenshot is an IAM user profile for 'Marco Tangaro' with a unique ID '20340b0b-c07f-46e0-a9c6-1f987ba9ff85' highlighted in red. A red arrow points from this ID to the 'User subject' box in the right screenshot. The right screenshot shows a Vault secrets page with a path '20340b0b-c07f-46e0-a9c6-1f987ba9ff85/5b68fbb5-d251-4820-b003-94b8ae2405cd'. A green box highlights the 'Deployment uuid' '5b68fbb5-d251-4820-b003-94b8ae2405cd'. A yellow arrow points from this box to a summary box at the bottom.

**User subject**  
unique and never  
reassigned user identifier

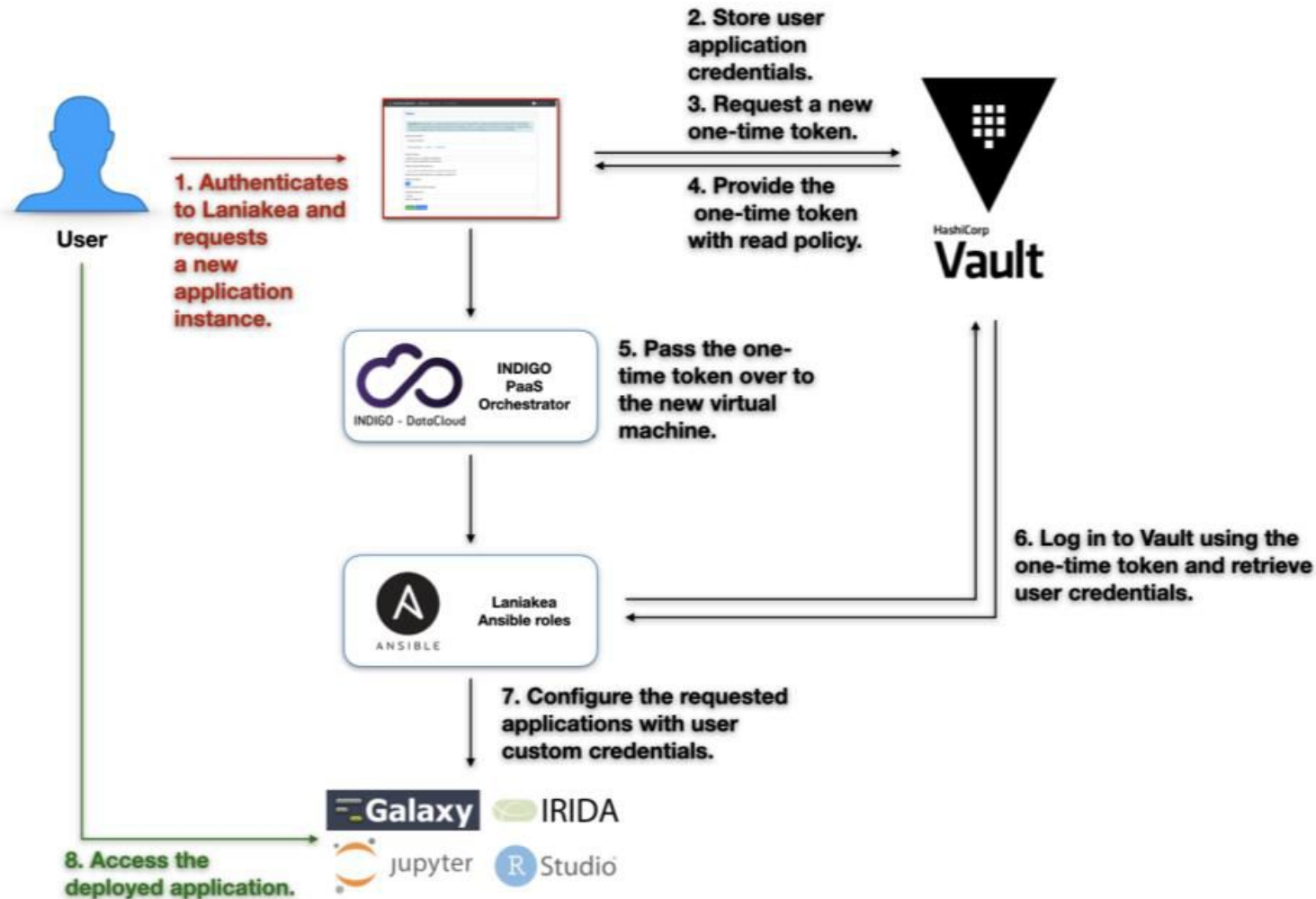
**Deployment uuid**  
unique and never  
reassigned deployment  
identifier

**The passphrase path on Vault is unique per user and Galaxy deployment. Only the deployment owner can write and read this path!**

User identity in IAM

Passphrase path on Vault

# Applications credentials



The storage encryption procedure has been extended to allow also users' credentials customisation for many applications.

In this case Ansible is responsible to retrieve credentials from Vault and correctly configure the application.

# *Deployments under VPN*

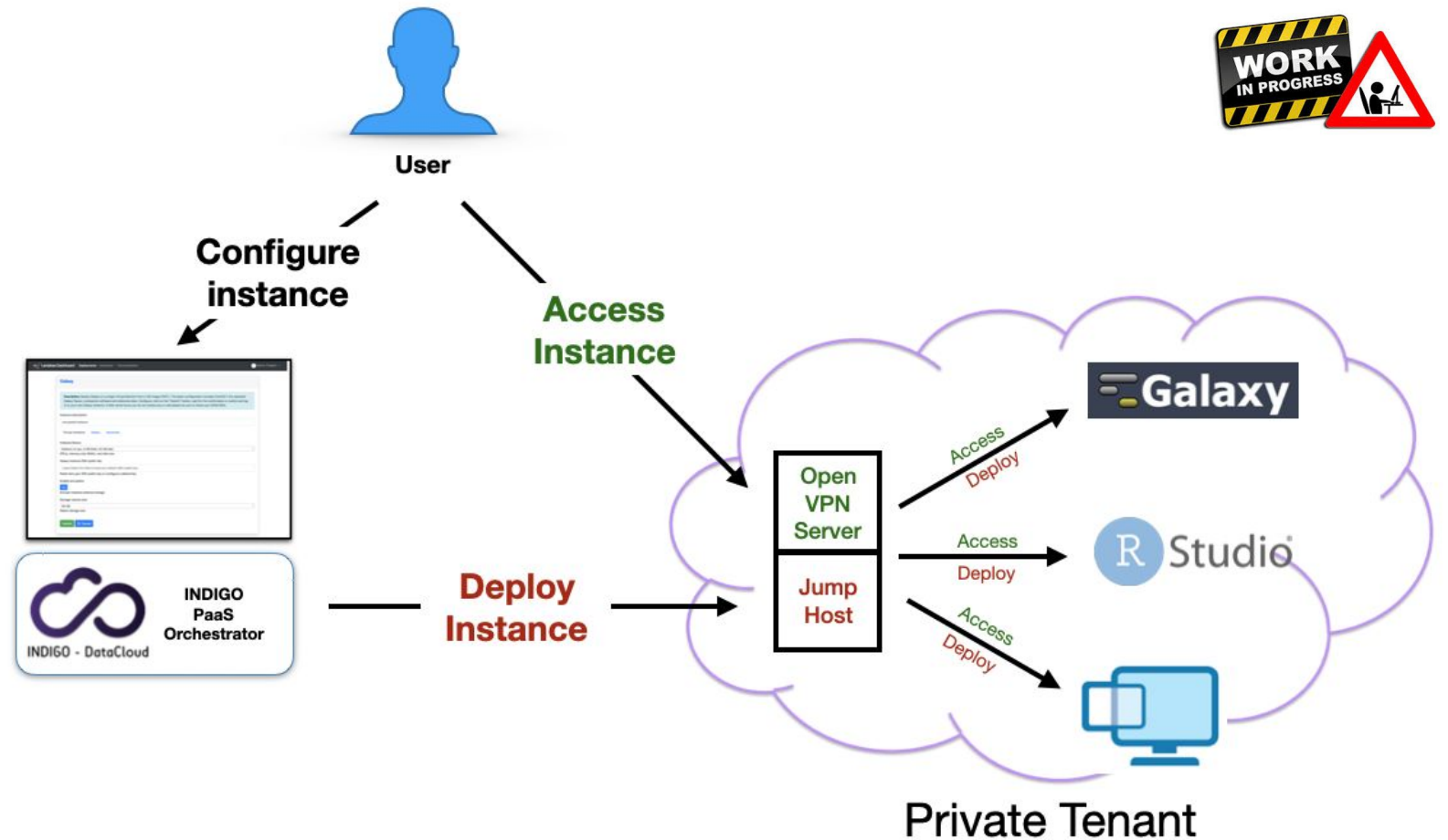
# Deployments under VPN



**VPN isolated environments** - Automatic deployments of virtual environments on private networks.

Isolation is reached using Tenant and security groups properties, granting the access only through VPN authentication.

User authentication to the VPN using the same Laniakea credentials.



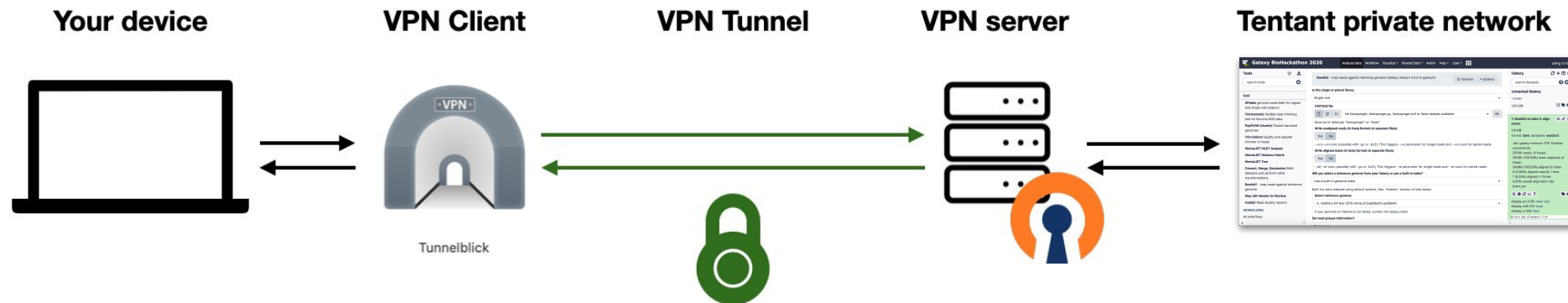
# Deployments under VPN



The VPN is based on OpenVPN, with clients and server are configured to use TPC protocol.

We have developed a PAM plugin to enable authentication through OpenID Connect, exploiting Oauth2 device flow:

1. the user connects to the VPN server using an OpenVPN client
2. PAM is configured to send verification code by mail to the user.
3. the user can authenticate with its own Laniakea credentials.
4. the OIDC provider (INDIGO-IAM) sends the access token to the VPN server, that is now able to verify users identity and authorizations.
5. if the user owns the right tenant permissions, he is granted access to the private network and can finally interact with the deployed application



# Conclusions



Storage encryption solution is already in production and exploited by several Laniakea Galaxy instances.

Data are still potentially exposed to attacks against the VM itself, where Galaxy or other applications need to consume them. We are working to provide Laniakea's users with the possibility to hide deployed applications behind a Virtual Private Network, achieving even more robust isolation of the research environment.

These approaches can help promoting the adoption of the on-demand model for Life Science and biomedical applications, making compute infrastructures more readily available to potential users even in the case of tight requirements for data protection.



# *Thanks for your attention*

## CONTACTS:

**Graziano Pesole (ELIXIR-ITALY Head of Node) [g.pesole@ibiom.cnr.it](mailto:g.pesole@ibiom.cnr.it)**

Federico Zambelli (ELIXIR-ITALY technical coordinator) [federico.zambelli@unimi.it](mailto:federico.zambelli@unimi.it)

Giacinto Donvito (Compute platform ELIXIR-ITALY) [giacinto.donvito@ba.infn.it](mailto:giacinto.donvito@ba.infn.it)

Nadina Foggetti (Legal expert) [nadina.foggetti@ba.infn.it](mailto:nadina.foggetti@ba.infn.it)

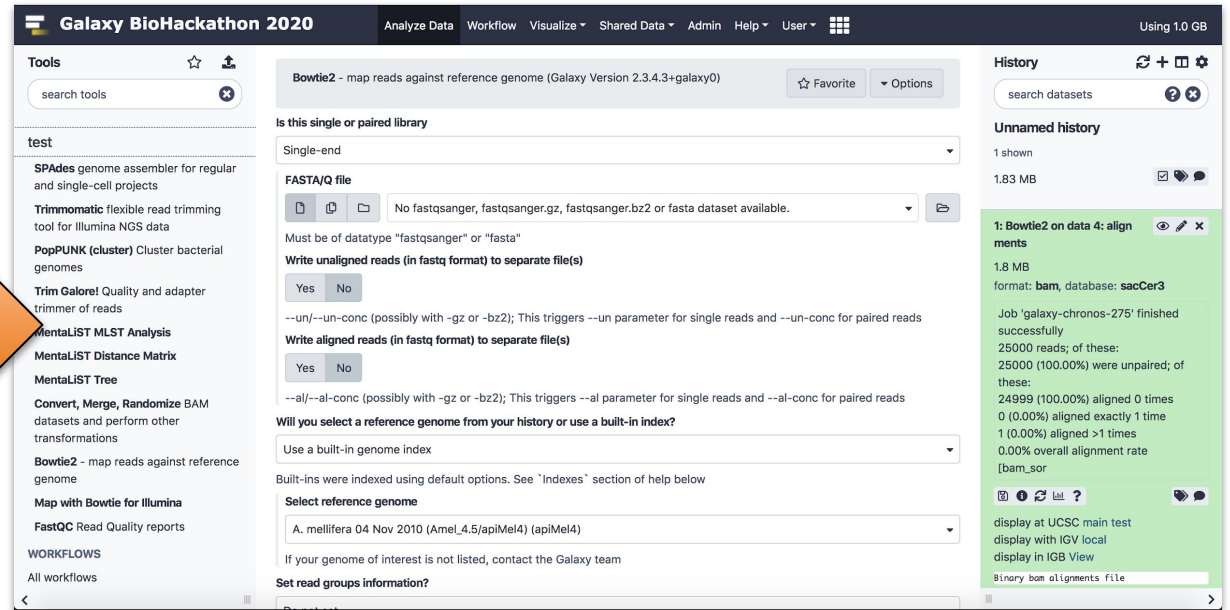
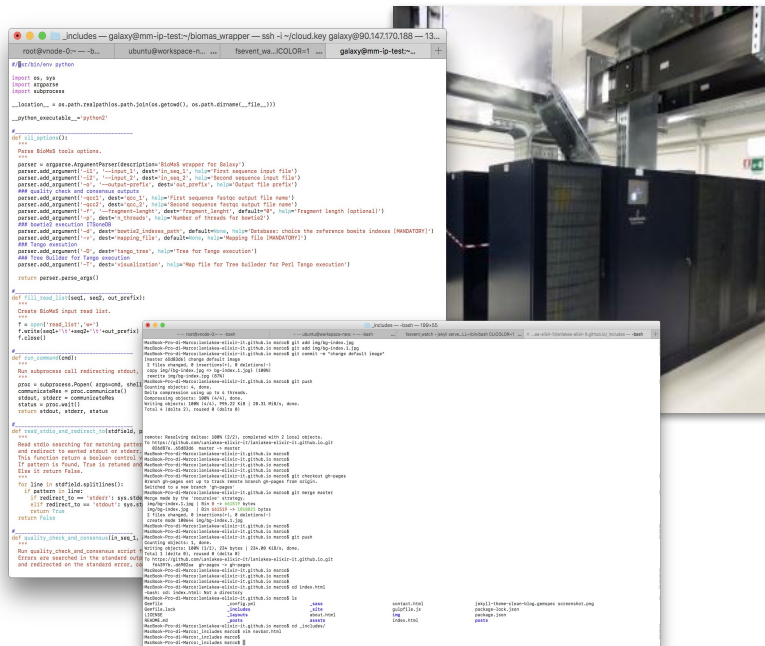
Marica Antonacci (PaaS developer) [marica.antonacci@ba.infn.it](mailto:marica.antonacci@ba.infn.it)

Marco Antonio Tangaro (Laniakea chief developer) [ma.tangaro@ibiom.cnr.it](mailto:ma.tangaro@ibiom.cnr.it)

# *Backup*



Allowing the community to move from command line tools to web user interfaces.



Allowing multiple users to exploit homogenous software environments, enhancing reproducibility.

# Laniakea



ELIXIR-Italy partners are actively involved in the service development and/or also contribute with cloud resources.

**A Laniakea service is in production for ELIXIR-ITALY partner but also for ELIXIR and external users.**

The ELIXIR-ITALY **Laniakea@ReCaS** Call offers access to Cloud resources to be used for the deployment of on-demand Galaxy instances.

[https://laniakea-elixir-it.github.io/laniakea\\_at\\_recas](https://laniakea-elixir-it.github.io/laniakea_at_recas)





# Laniakea



elixir Laniakea Dashboard Deployments Users Documentation Marco Tangaro

Most used

Galaxy Galaxy

Galaxy Galaxy

Galaxy

Galaxy cluster

Galaxy

Galaxy

Full description

Deploy Galaxy from a VM image with cluster support (FAST). The basic configuration includes CentOS 7, SLURM, the selected Galaxy flavour, companion software and reference data. Configure, click on the "Submit" button, wait for the confirmation e-mail(s) 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.

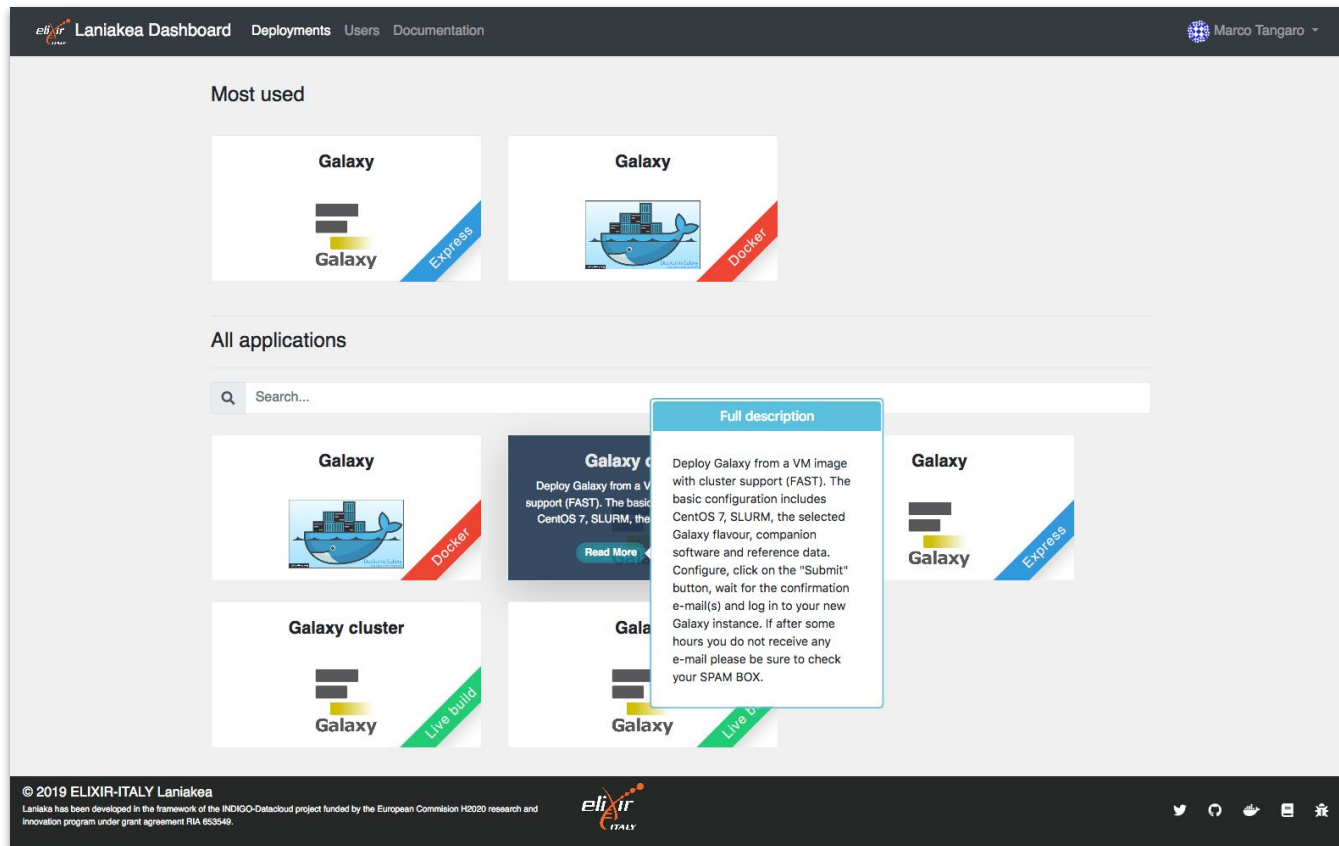
© 2019 ELIXIR-ITALY Laniakea

Laniakea has been developed in the framework of the INDIGO-Datacloud project funded by the European Commission H2020 research and innovation program under grant agreement RIA 663549.

The Laniakea Dashboard home page.

Each tile provides a quick explanation of the application and links to the configuration and launch section.

Also more applications available: Jupyter, RStudio, ...



## Different deployment strategies:

**Live Build**: build Galaxy from scratch -> always up-to-date (deployment time depending by the tools number).

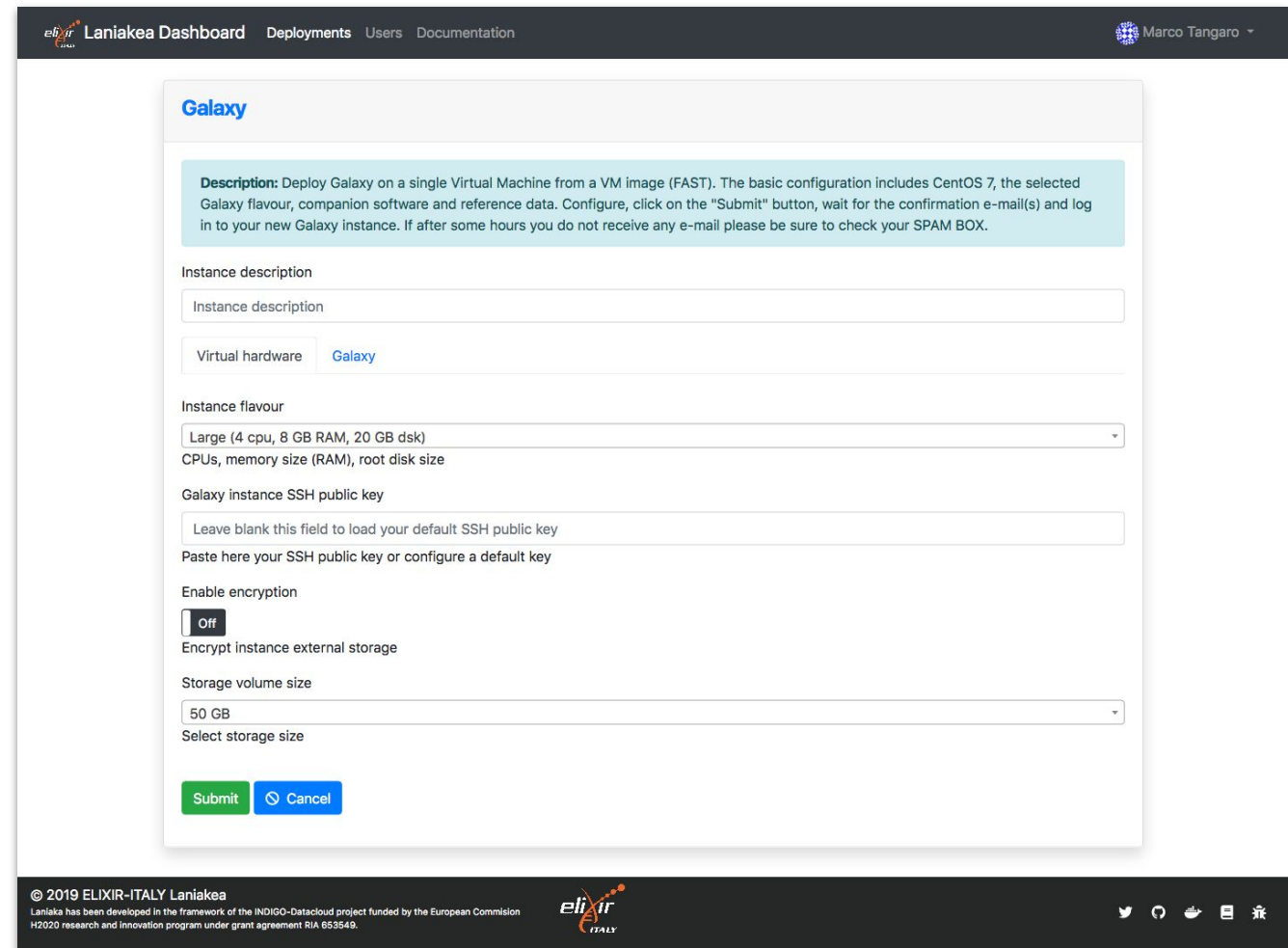
**Express**: pre-built Galaxy images -> fast deployment, but tools not always at the last available version.

**Docker**: fast deployment of new flavours.



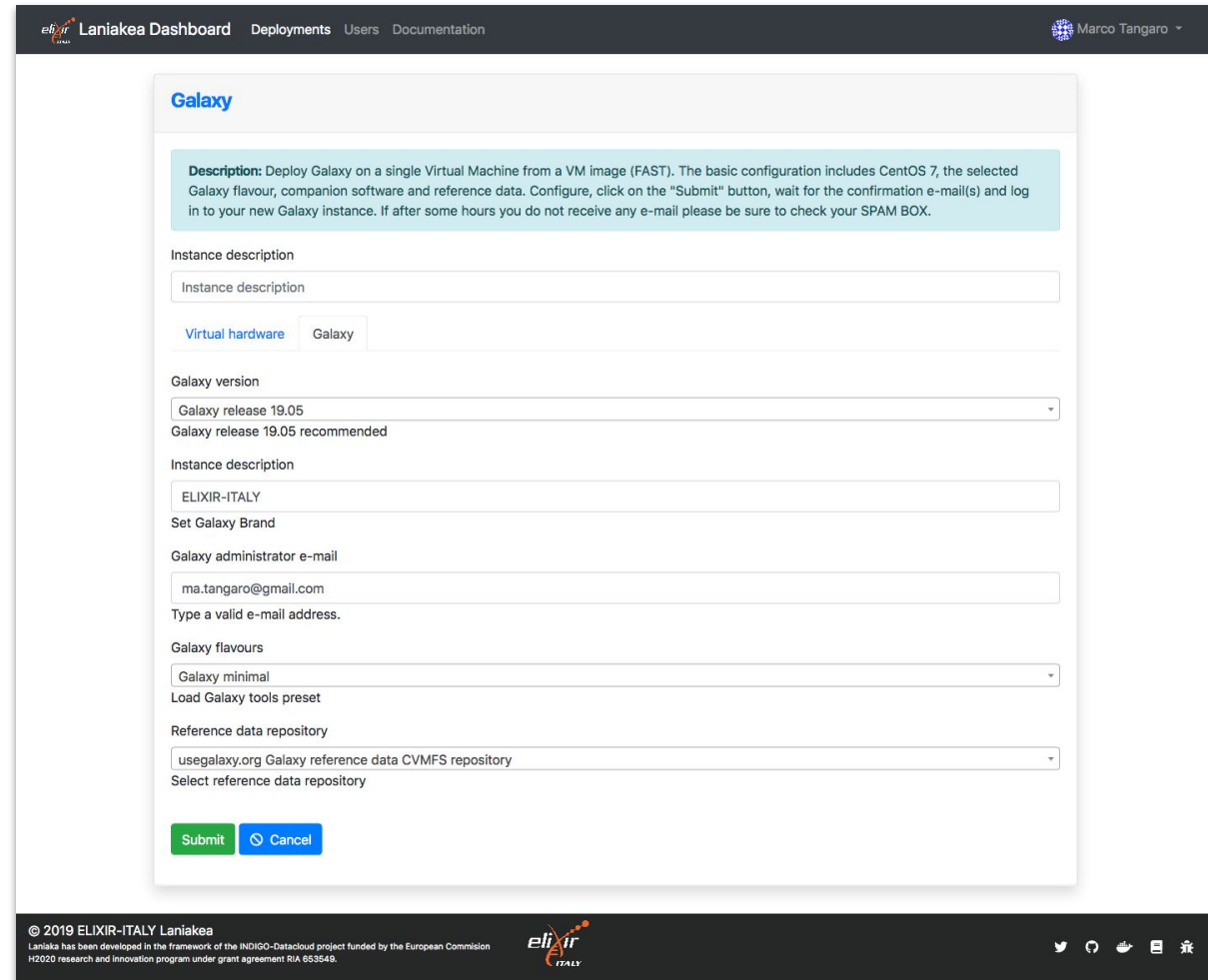
The web front-end provides different tabs to configure your Galaxy.

Virtual hardware: CPU, RAM and Storage

A screenshot of the Laniakea Galaxy configuration dashboard. The page has a dark header with the Laniakea logo, navigation links (Dashboard, Deployments, Users, Documentation), and a user profile (Marco Tangaro). The main content area is titled "Galaxy" and contains a description, an "Instance description" text field, tabs for "Virtual hardware" and "Galaxy", an "Instance flavour" dropdown menu (set to "Large (4 cpu, 8 GB RAM, 20 GB dsk)"), an SSH public key text field, an "Enable encryption" toggle (set to "Off"), and a "Storage volume size" dropdown menu (set to "50 GB"). At the bottom are "Submit" and "Cancel" buttons. The footer contains copyright information, a disclaimer, and social media icons.

The web front-end provides different tabs to configure your Galaxy.

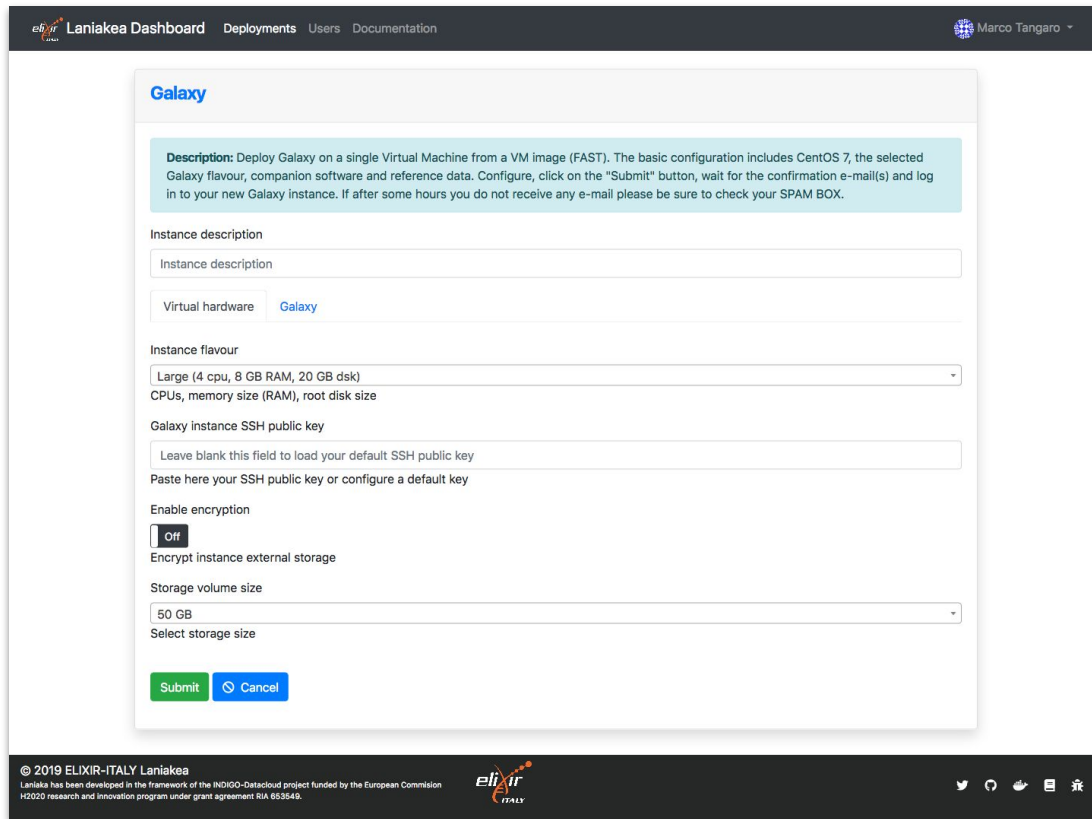
Galaxy software: version, credentials, flavor and reference data.



The screenshot shows the 'Galaxy' configuration page in the Laniakea dashboard. The page has a dark header with navigation links: 'Laniakea Dashboard', 'Deployments', 'Users', and 'Documentation'. A user profile 'Marco Tangaro' is visible in the top right. The main content area is titled 'Galaxy' and contains a description box with instructions on how to deploy Galaxy on a VM. Below this, there are several form fields and tabs:

- Instance description:** A text input field.
- Virtual hardware:** A tab labeled 'Galaxy' is selected.
- Galaxy version:** A dropdown menu showing 'Galaxy release 19.05' (recommended).
- Instance description:** A text input field containing 'ELIXIR-ITALY'.
- Set Galaxy Brand:** A text input field.
- Galaxy administrator e-mail:** A text input field containing 'ma.tangaro@gmail.com'.
- Galaxy flavours:** A dropdown menu showing 'Galaxy minimal'.
- Reference data repository:** A dropdown menu showing 'usegalaxy.org Galaxy reference data CVMFS repository'.

At the bottom of the form are 'Submit' and 'Cancel' buttons. The footer contains copyright information for 2019 ELIXIR-ITALY Laniakea, a small ELIXIR ITALY logo, and social media icons.



**Galaxy**

**Description:** Deploy Galaxy on a single Virtual Machine from a VM image (FAST). The basic configuration includes CentOS 7, the selected Galaxy flavour, companion software and reference data. Configure, click on the "Submit" button, wait for the confirmation e-mail(s) 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.

Instance description

Instance description

Virtual hardware **Galaxy**

Instance flavour

Large (4 cpu, 8 GB RAM, 20 GB dsk)  
CPUs, memory size (RAM), root disk size

Galaxy instance SSH public key

Leave blank this field to load your default SSH public key

Paste here your SSH public key or configure a default key

Enable encryption

Off

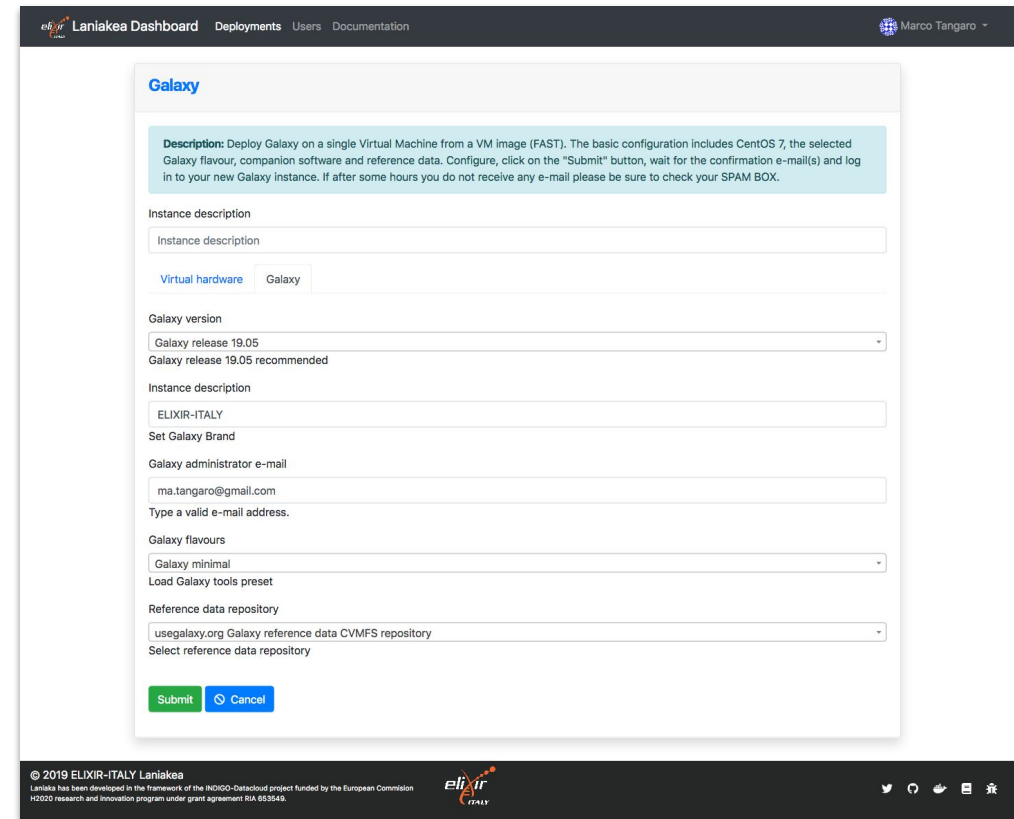
Encrypt instance external storage

Storage volume size

50 GB

Select storage size

© 2019 ELIXIR-ITALY Laniakea  
Laniakea has been developed in the framework of the INDIGO-Datacloud project funded by the European Commission H2020 research and innovation program under grant agreement RIA 853549.



**Galaxy**

**Description:** Deploy Galaxy on a single Virtual Machine from a VM image (FAST). The basic configuration includes CentOS 7, the selected Galaxy flavour, companion software and reference data. Configure, click on the "Submit" button, wait for the confirmation e-mail(s) 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.

Instance description

Instance description

Virtual hardware **Galaxy**

Galaxy version

Galaxy release 19.05  
Galaxy release 19.05 recommended

Instance description

ELIXIR-ITALY

Set Galaxy Brand

Galaxy administrator e-mail

ma.tangaro@gmail.com

Type a valid e-mail address.

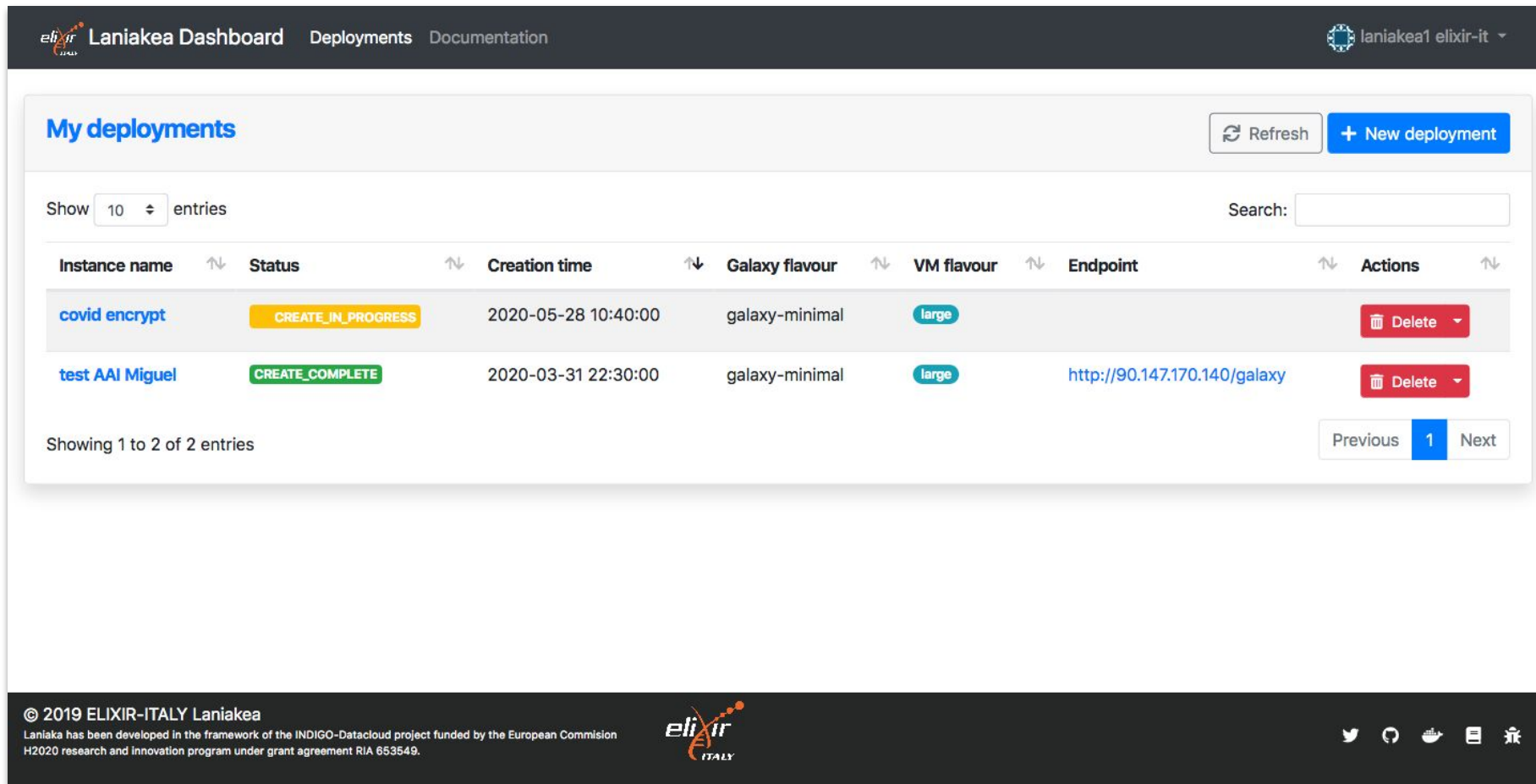
Galaxy flavours

Galaxy minimal  
Load Galaxy tools preset

Reference data repository

usegalaxy.org Galaxy reference data CVMFS repository  
Select reference data repository

© 2019 ELIXIR-ITALY Laniakea  
Laniakea has been developed in the framework of the INDIGO-Datacloud project funded by the European Commission H2020 research and innovation program under grant agreement RIA 853549.



elixir Laniakea Dashboard Deployments Documentation laniakea1 elixir-it

### My deployments



Refresh + New deployment

Show 10 entries Search:

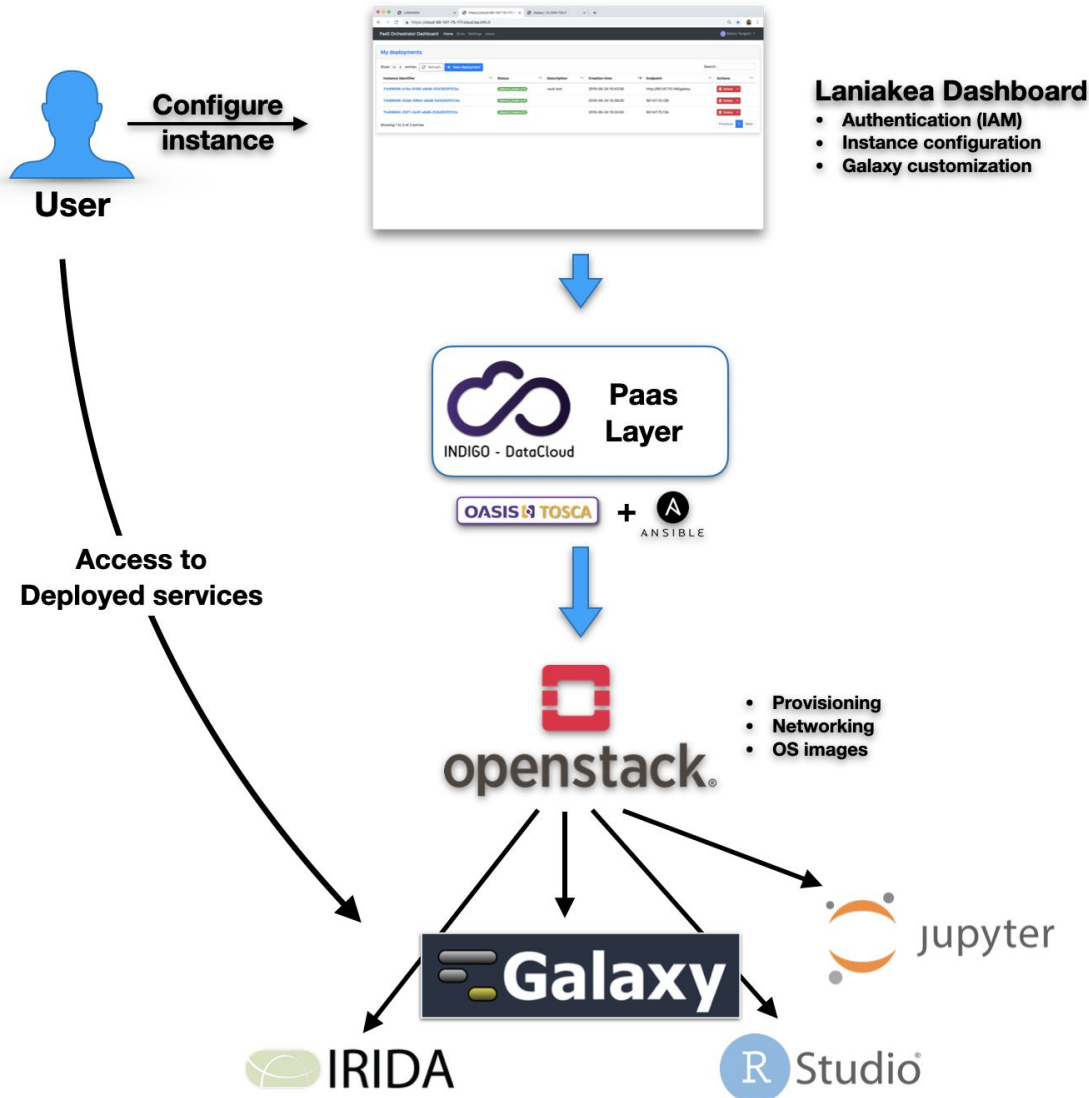
Instance name	Status	Creation time	Galaxy flavour	VM flavour	Endpoint	Actions
covid encrypt	CREATE_IN_PROGRESS	2020-05-28 10:40:00	galaxy-minimal	large		Delete
test AAI Miguel	CREATE_COMPLETE	2020-03-31 22:30:00	galaxy-minimal	large	http://90.147.170.140/galaxy	Delete

Showing 1 to 2 of 2 entries Previous 1 Next

© 2019 ELIXIR-ITALY Laniakea  
Laniakea has been developed in the framework of the INDIGO-Datacloud project funded by the European Commission H2020 research and innovation program under grant agreement RIA 653549.



# Laniakea architecture (simplified view)



- **Dashboard** - User friendly access to configure and launch a Galaxy instance
- **INDIGO PaaS** - Galaxy automatic deployment
- **Cloud Providers** - ReCaS-Bari

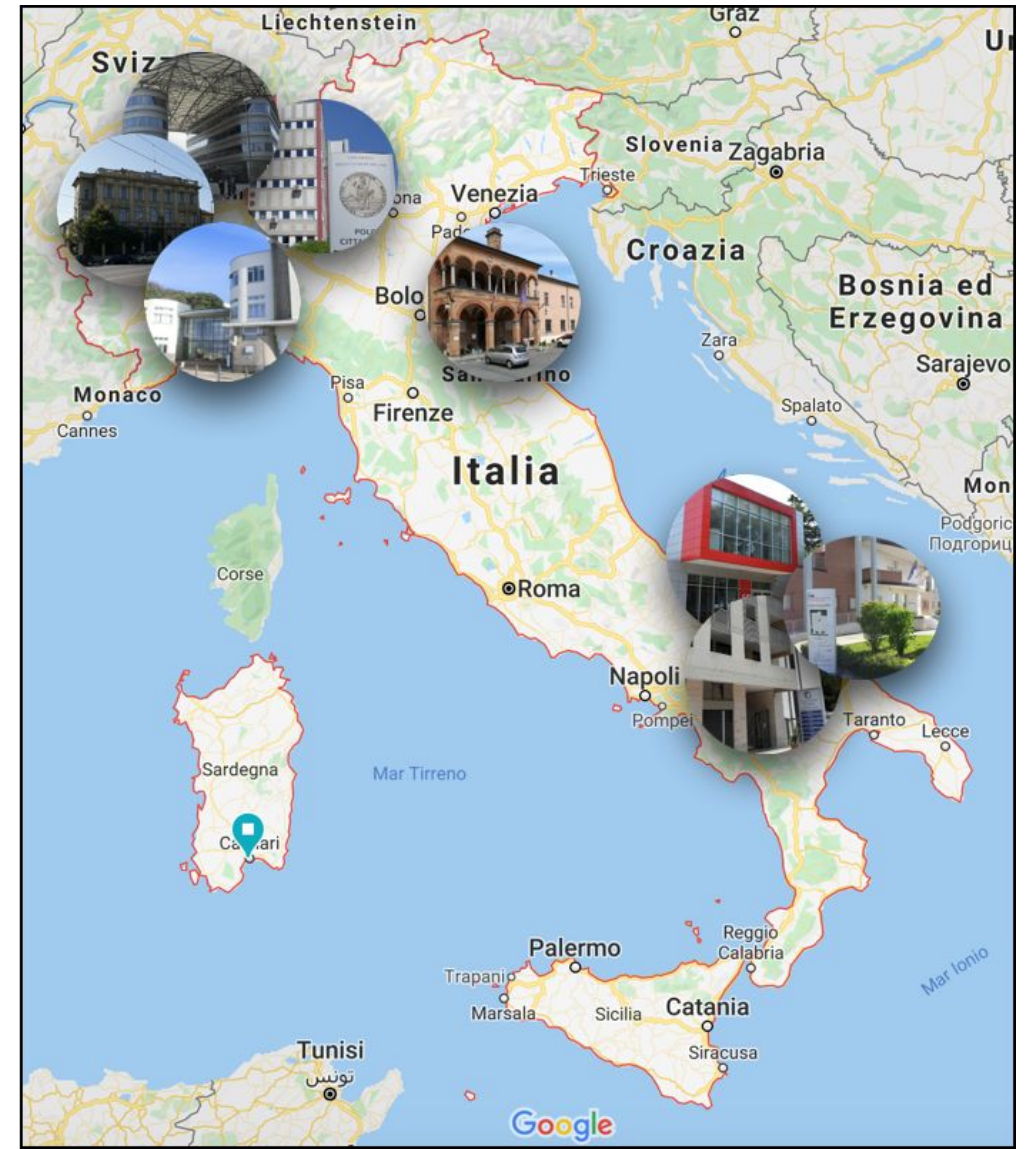
# Laniakea@ReCaS



Currently, some important Italian Institutions are using Laniakea for their daily work:

- 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 (public Galaxy server and tools development).
- IBIOM-CNR (public Galaxy server and tools development).
- University of Turin (training)

... and counting.





# *Laniakea Dashboard*



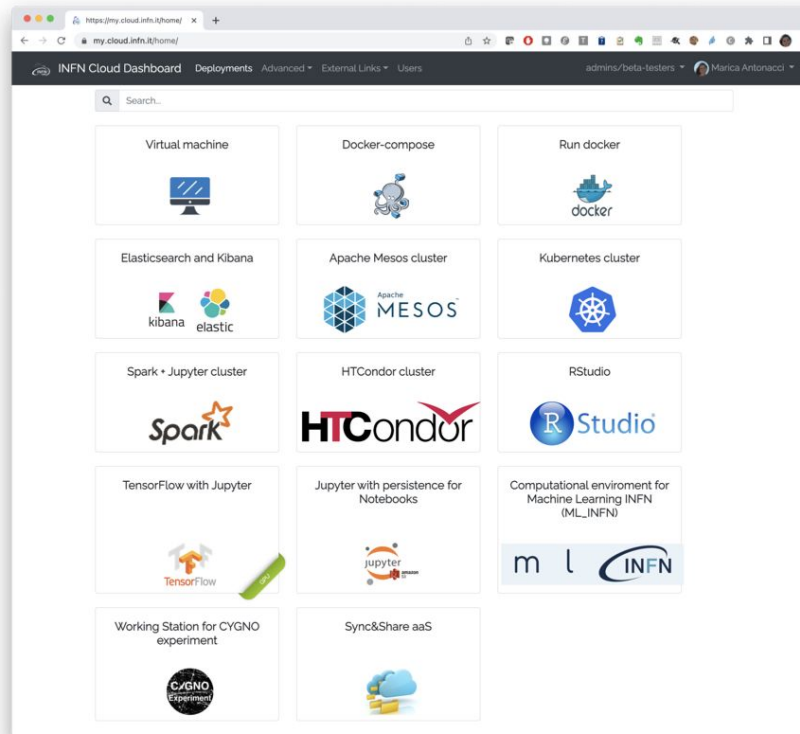
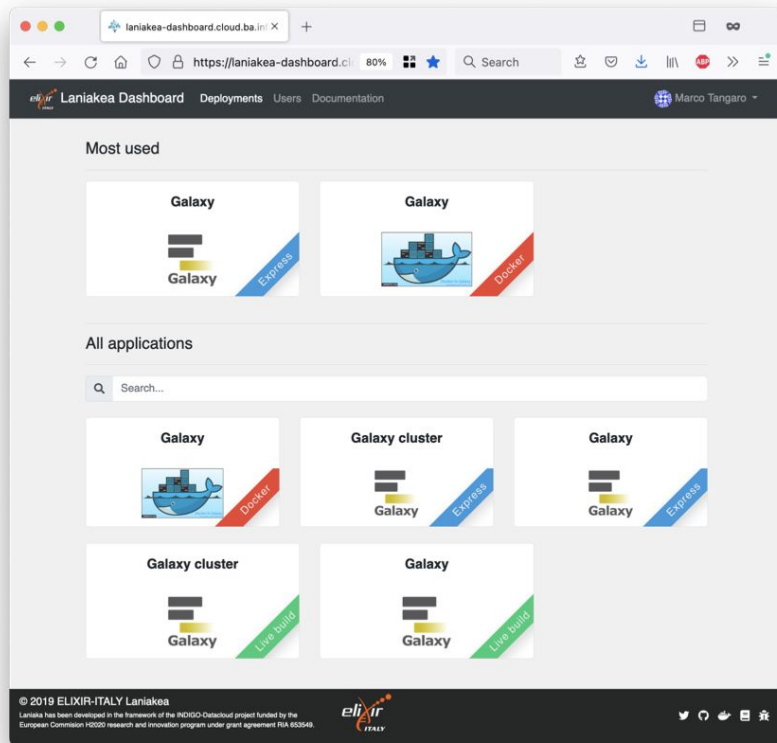
The PaaS Layer accepts deployment requests in the form of TOSCA Templates: a document (YAML syntax) describing the infrastructure to deploy, e.g. the virtual hardware and the software to be installed and configured. The Dashboard parses the TOSCA document automatically and renders the user interface with user friendly forms. This allows to extend Laniakea functionalities just adding new templates without any code modification.

**We extended the TOSCA templates inputs to create configurable forms.** This creates a flexible web interface, allowing straightforward customisation of the user experience through human readable YAML configuration files. The dashboard template is modular and can be easily adapted adding new functionalities to the user interface (e.g. adding a dropdown menu, text fields, toggles...) based on the Laniakea administrator requirements.



# *Laniakea Dashboard*

# Laniakea Dashboard



Flask web micro-framework ([flask.pocoo.org/](https://flask.pocoo.org/)),

Jinja2 template engine ([jinja.pocoo.org/](https://jinja.pocoo.org/))

Bootstrap 4 toolkit ([getbootstrap.com/](https://getbootstrap.com/)).

Integrated with Hashicorp Vault for user secrets management.

# Laniakea Dashboard

```
instance_flavor:  
  type: string  
  description: instance flavor (num_cpu, memory, disk)  
  default: small
```

**TOSCA template**

```
instance_flavor:  
  display_name: "Instance flavour"  
  tag_type: "select"  
  description: "CPUs, memory size (RAM), root disk size"  
  constraints:  
    - { value: "small", label: "Small (1 cpu, 2 GB RAM, 20 GB dsk)" }  
    - { value: "medium", label: "Medium (2 cpu, 4 GB RAM, 20 GB dsk)" }  
    - { value: "large", label: "Large (4 cpu, 8 GB RAM, 20 GB dsk)" }  
    - { value: "xlarge", label: "xLarge (8 cpu, 16 GB RAM, 20 GB dsk)" }  
  tab: "Virtual hardware"
```

**Parameters file**

User input field in the TOSCA template to select instance flavour in terms of vCPUs, RAM and root disk storage.

Virtual hardware **Galaxy** Advanced **Dashboard render**

Instance flavour

Small (1 cpu, 2 GB RAM, 20 GB dsk)

CPUs, memory size (RAM), root disk size

Virtual hardware **Galaxy** Advanced **Dashboard render**

Instance flavour

Small (1 cpu, 2 GB RAM, 20 GB dsk)

Small (1 cpu, 2 GB RAM, 20 GB dsk)

Medium (2 cpu, 4 GB RAM, 20 GB dsk)

Large (4 cpu, 8 GB RAM, 20 GB dsk)

xLarge (8 cpu, 16 GB RAM, 20 GB dsk)

# Laniakea Dashboard

```
users:
  type: list
  description: list of users to create on the VM
  entry_schema:
    type: toasca.datatypes.indigo.User
  default: []
  required: false

instance_flavor:
  type: string
  description: instance flavor (num_cpu, memory, disk)
  default: small
storage_size:
  type: string
  description: storage memory required for the instance
  default: 10 GB

# Storage encryption
storage_encryption:
  type: boolean
  description: Enable storage encryption.
  default: False
  required: true
vault_url:
  type: string
  description: Hashicorp Vault server url
  default: changeit
  required: false
vault_wrapping_token:
  type: string
  description: Vault Wrapping token to write secret
  default: not_a_wrapping_token
  required: false
vault_secret_path:
  type: string
  description: Vault path to store secret
  default: "subject/depuuid"
  required: false
vault_secret_key:
  type: string
  description: Vault secret key name
  default: secret
  required: false
```



### Galaxy

Description: Deploy Galaxy on a single Virtual Machine from a VM image (FAST). The basic configuration includes CentOS 7, the selected Galaxy flavour, companion software and reference data. Configure, click on the "Submit" button, wait for the confirmation e-mail(s) 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.

Deployment description

Virtual hardware **Galaxy** Advanced

Instance flavour

  
CPUs, memory size (RAM), root disk size

Storage volume size

  
Select storage size

Enable encryption

 Off  
Encrypt instance external storage

# Laniakea Dashboard

```
admin_email:  
  type: string  
  description: email of the admin user  
  default: admin@admin.com  
admin_api_key:  
  type: string  
  description: key to access the API with admin role  
  default: not_very_secret_api_key  
version:  
  type: string  
  description: galaxy version to install  
  default: master  
instance_description:  
  type: string  
  description: galaxy instance description  
  default: "INDIGO Galaxy test"  
export_dir:  
  type: string  
  description: path to store galaxy data  
  default: /export  
  
flavor:  
  type: string  
  description: Galaxy flavor for tools installation  
  default: "galaxy-no-tools"
```



### Galaxy

Description: Deploy Galaxy on a single Virtual Machine from a VM image (FAST). The basic configuration includes CentOS 7, the selected Galaxy flavour, companion software and reference data. Configure, click on the "Submit" button, wait for the confirmation e-mail(s) 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.

Deployment description

Virtual hardware  Galaxy  Advanced

Galaxy administrator e-mail  
  
Type a valid e-mail address.

Galaxy version  
  
Galaxy release 20.05 recommended

Instance description  
  
Set Galaxy Brand

Galaxy flavours  
  
Load Galaxy tools preset

# Video demo



<https://youtu.be/bzJbHby8gAY>

Video outline:

- Authentication
- SSH key management
- Galaxy Deployment with encrypted storage
- RStudio/JupyterLab deployment
- How to build and encrypted deployment (from a simple VM to a full encrypted instance with Galaxy).