

Laniakea: a Galaxy-on-demand Provider Platform Through Cloud Technologies

Tangaro M.A.¹, Donvito G.², Antonacci M.², Chiara M.³, Mandreoli P.³, Pesole G.^{1,4}, Zambelli F.^{1,3}

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



2019 Galaxy Community Conference, 1-6 July, Freiburg, Germany

Galaxy is rapidly becoming the de facto standard workflow manager for bioinformatics. Although several Galaxy public services are currently available, the usage of a private Galaxy instance is still mandatory or preferable for several use cases, including heavy workloads, data privacy concerns or particular customization needs.

In this context, cloud computing technologies and infrastructures can provide a powerful and scalable solution to avoid the onerous deployment and maintenance of a local hardware and software infrastructure.

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.

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.

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.

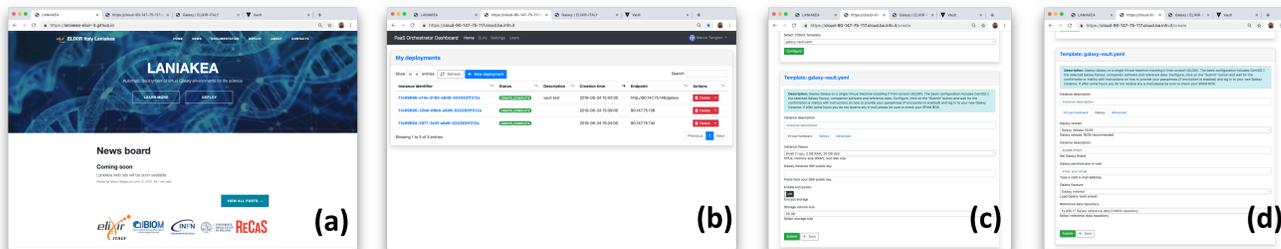
End-users interact with Laniakea through a web front-end allowing for a general setup of a Galaxy instance. The deployment of the virtual hardware and of the Galaxy software ecosystem is subsequently performed by the INDIGO Platform as a Service (PaaS) layer. At the end of the process, the user gains 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 user data volumes and rapid development of novel Galaxy flavours for specific tasks.

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. 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 a duration compatible with the project requirements. A way to request an extension of the project will also be available.

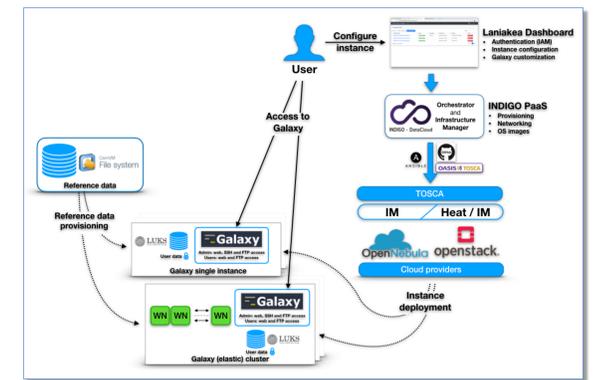
LANIAKEA DASHBOARD

Laniakea dashboard is the new, redesigned and reimplemented, user interface of Laniakea, developed using Flask web micro-framework (flask.pocoo.org/), Jinja2 template engine (jinja.pocoo.org/) and Bootstrap 4 toolkit (getbootstrap.com/). It provides different panels for configuring the virtual hardware and the Galaxy instance (e.g administrator credentials).

Moreover it has been integrated with Hashicorp Vault for user secret management.



Laniakea Dashboard – (a) The new Laniakea website (laniakea-elixir-it.github.io). (b) Dashboard home page. (c) Virtual hardware configuration tab. (d) Galaxy configuration tab.



Laniakea architecture – the service is based on the coordination of separated components part of the INDIGO-DataCloud catalogue.

YAML syntax example

The Laniakea dashboard provides 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) based on the Laniakea administrator requirements.

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

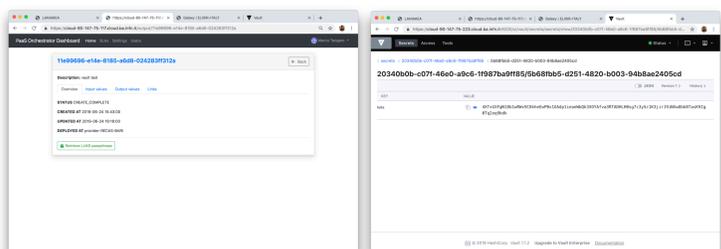
Example of flavour dropdown menu implementation – Each parameter of the menu is specified in the code, interpreted by the Jinja2 template engine and rendered using Flask

USER SECRET MANAGEMENT WITH HASHICORP VAULT

The user 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.

The encryption procedure has been completely re-worked and automated in order to simplify the user experience: now the user can encrypt storage on-demand, using a strong random alphanumeric passphrase, without having to interact with the Galaxy instance through SSH.

This has been achieved integrating the key management system Hashicorp Vault (vaultproject.io) to store encryption keys, which are shown only if explicitly requested by the user.

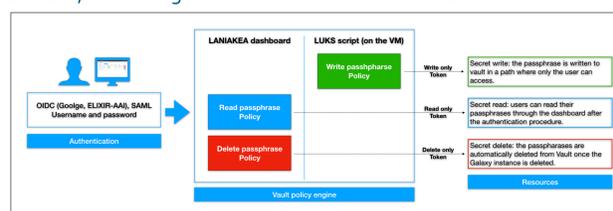


Secret management on Laniakea Dashboard and Vault.

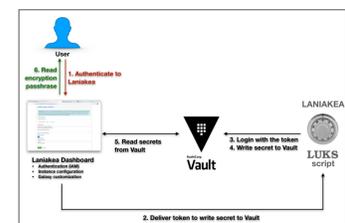
Vault is a tool for securely accessing "secrets". A secret is everything you want to tightly control access to, such as encryption passphrases. Data stored on Vault are encrypted with 256 bit AES (Advanced Encryption Standard) cipher in the Galois Counter Mode (GCM) with a randomly generated nonce.

Vault main concepts:

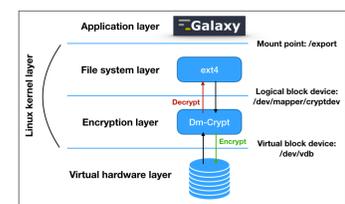
- Everything in Vault is path based: users are able to write their secrets on a specific path, depending on their Identity.
- Tokens are the core method for authentication within Vault. After the authentication on the Laniakea Dashboard, tokens are dynamically generated based on user identity.
- Policies 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.



Vault authentication and Authorization flow



Vault workflow – A short lived write only token, usable only once, is delivered to the Laniakea encryption script on the VM. The Storage volume is encrypted and the password sent to Vault.



File system configuration – 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.

CONCLUSIONS

The development of a brand new user interface and the integration with Vault aims to grant a better user experience, improved functionality for the encrypted storage and easier frontend customisability for the administrator. We expect that the first ELIXIR-IT Galaxy-on-demand service based on this platform will provide 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. We will continue to build on this service and the underlying platform in the framework of the EOSC-Life and EOSC-Pillar H2020 projects.

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 preprint 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

