Infrastruttura di analisi dati di livello PaaS compliant con i requisiti tecnici e legali per l'analisi di dati genetici e sanitari

Nadina Foggetti – Marica Antonacci – Giacinto Donvito – Marco Antonio Tangaro*

Istituto Nazionale di Fisica Nucleare – Bari

* CNR-IBIOM - Bari

CONDIVISIONI
Conferenza GARR 2022
Palermo 18-20 maggio

EOSC Pillar

Coordinates national EOSC initiatives in different countries: Austria, Belgium, France, Germany, and Italy, to harmonize the different national strategies.
Use Case 6 - Analyse the regulatory compliance of the ELIXIR-ITALY service Laniakea, on-demand Galaxy platform, for the ELIXIR and Life Science communities, in case of clinical and sensitive data.

LANIAKEA is a cloud based Galaxy instance provider. 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.

Three different levels:

1. Data management: encryption
2. Virtual environment isolation, through VPN
3. Legal framework

Legal and Ethical Issue

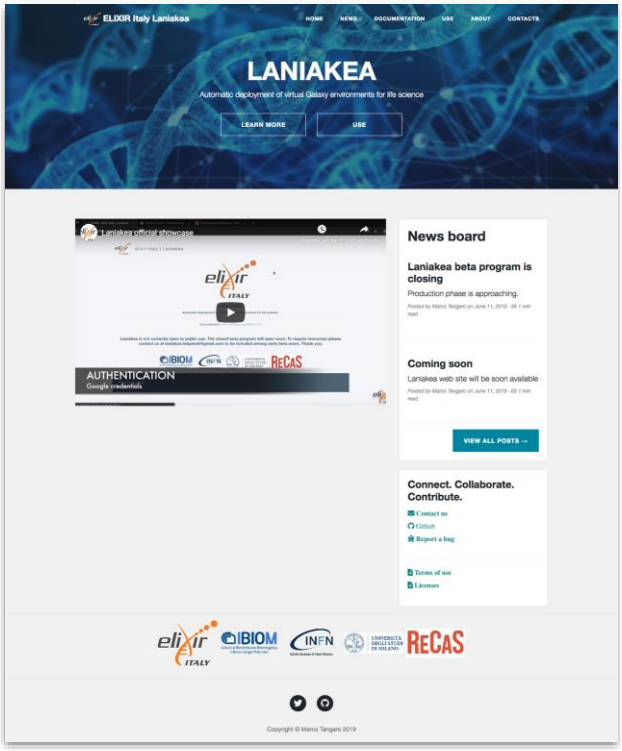
Connessione con il lavoro fatto da T. 4.1

- ❑ Studio dello scenario e dei gap giuridici ed etici (es. Pseudonimizzazione)
- ❑ Definizione della Checklist per lo scenario
- ❑ Applicazione dei principi OS e OA, principi Fair al trattamento dei dati sanitari all'interno dello scenario

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

Version Final: 10.5281/zenodo.6334878

Laniakea



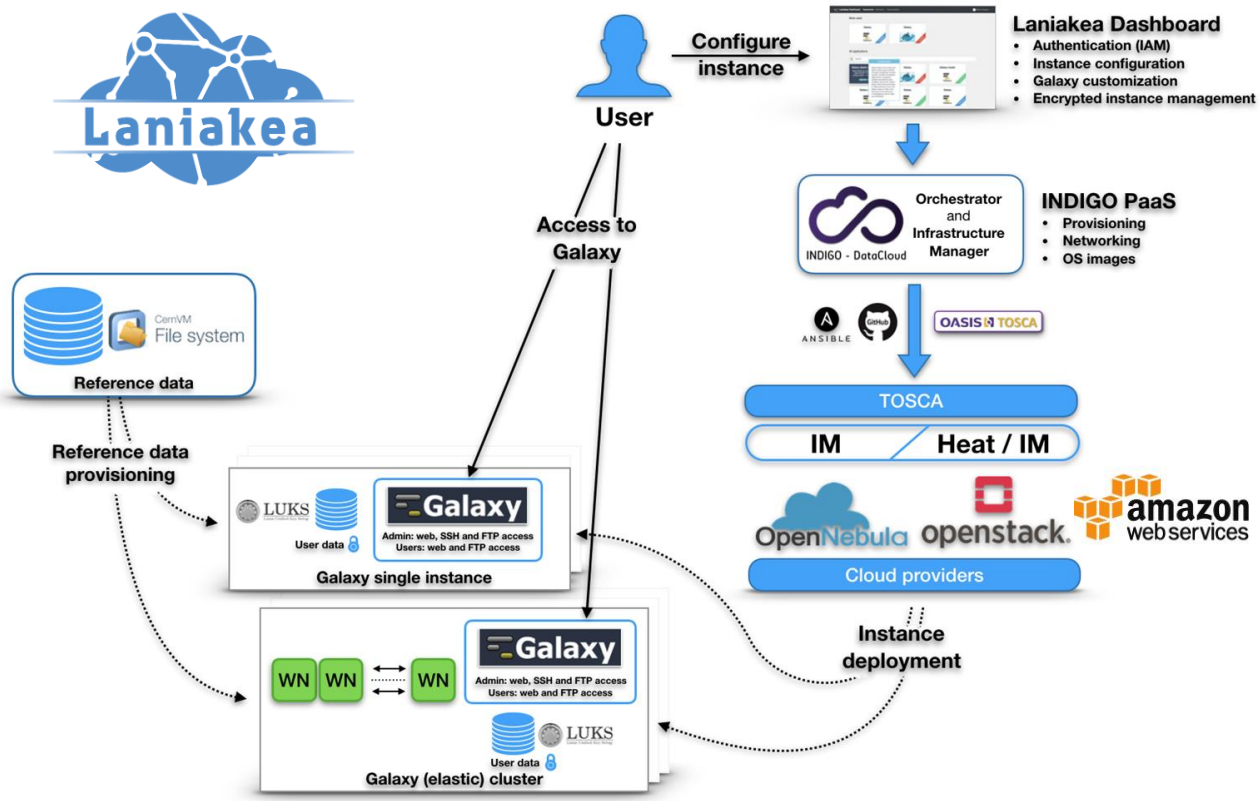
https://laniakea-elixir-it.github.io

LANIAKEA is a cloud Galaxy instance provider, based on INDIGO-DataCloud software catalogue. Its architecture automates the creation of Galaxy-based virtualized environments exploiting the software catalogue provided by the INDIGO-DataCloud project.

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

No need for maintenance of the hardware and software infrastructure.

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.
- **Persistent storage** - With/without encryption.
- **Reference data availability** - With CERN-VM FileSystem.
- **CLUES -** Elasticity manager.

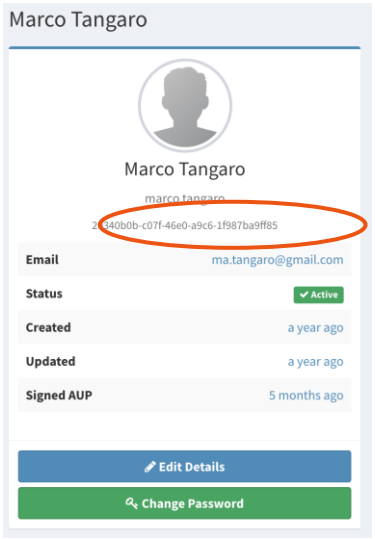
Health and data security aspects

Key management: Vault introduction

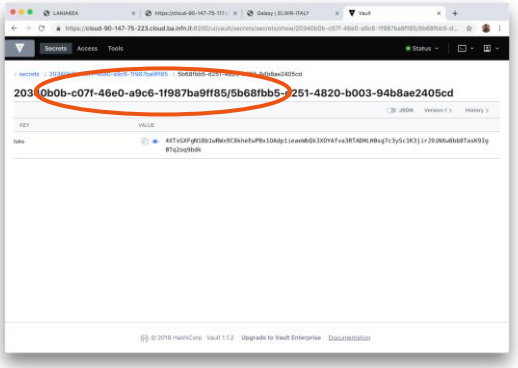
Vault is a tool for securely accessing "secrets". A secret is everything you want to tightly control access to, such as encryption passphrases. Data stored on Vault are encrypted with 256 bit AES (Advanced Encryption Standard) cipher in the Galois Counter Mode (GCM).

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.



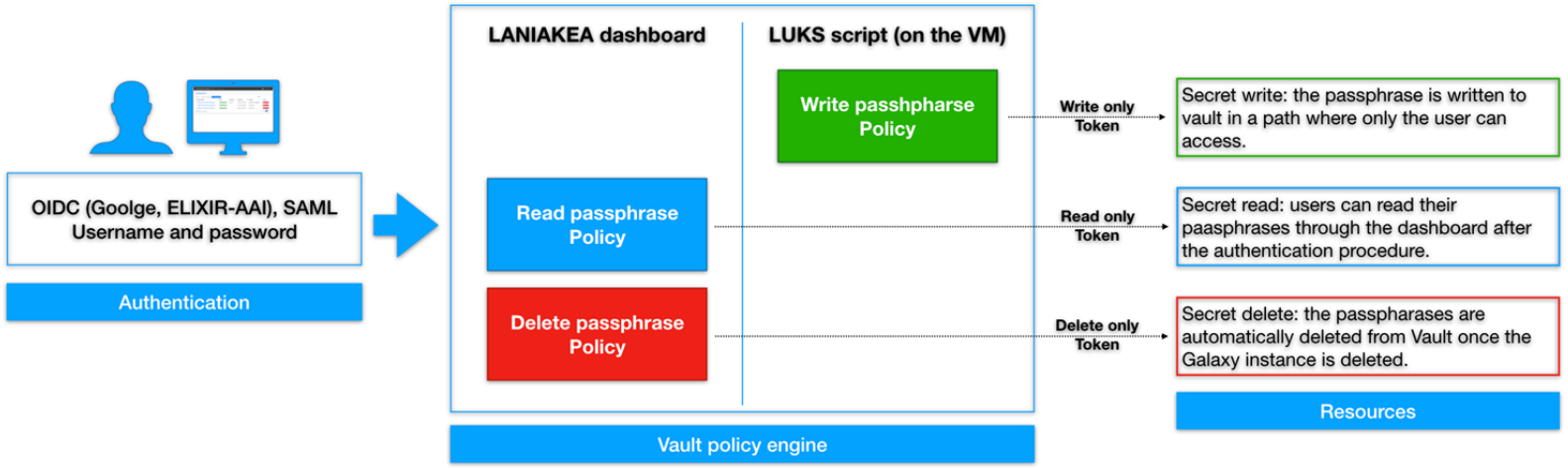
The IAM user subject (user unique ID) is used to store user's secrets in Vault: only the user can access to his path!



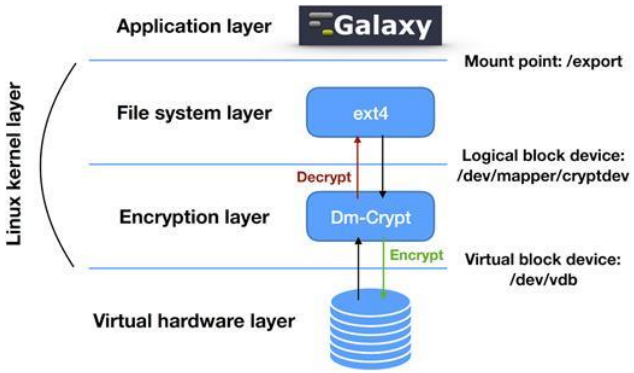
Health and data security aspects

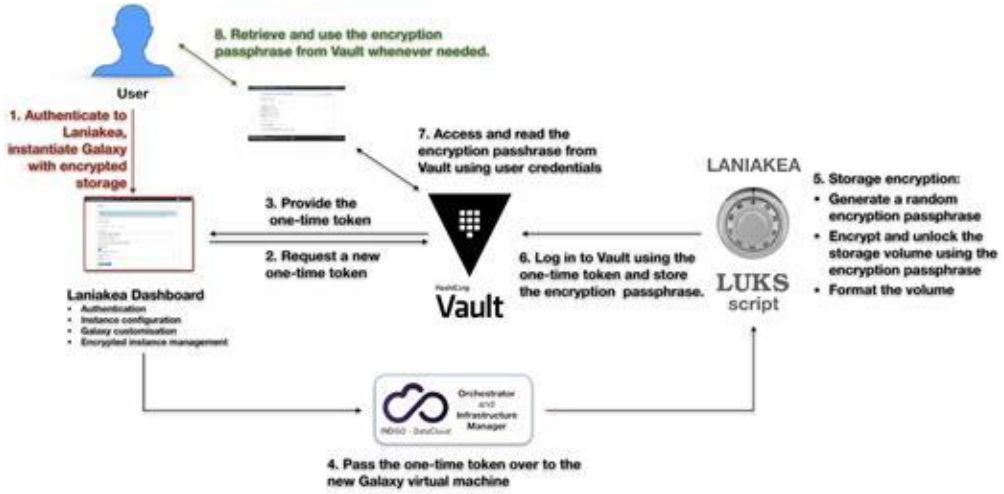
Key management: Vault authentication and authorization workflow

- 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 Delete the passphrase from Vault, once the deployment is deleted.



Data Encryption



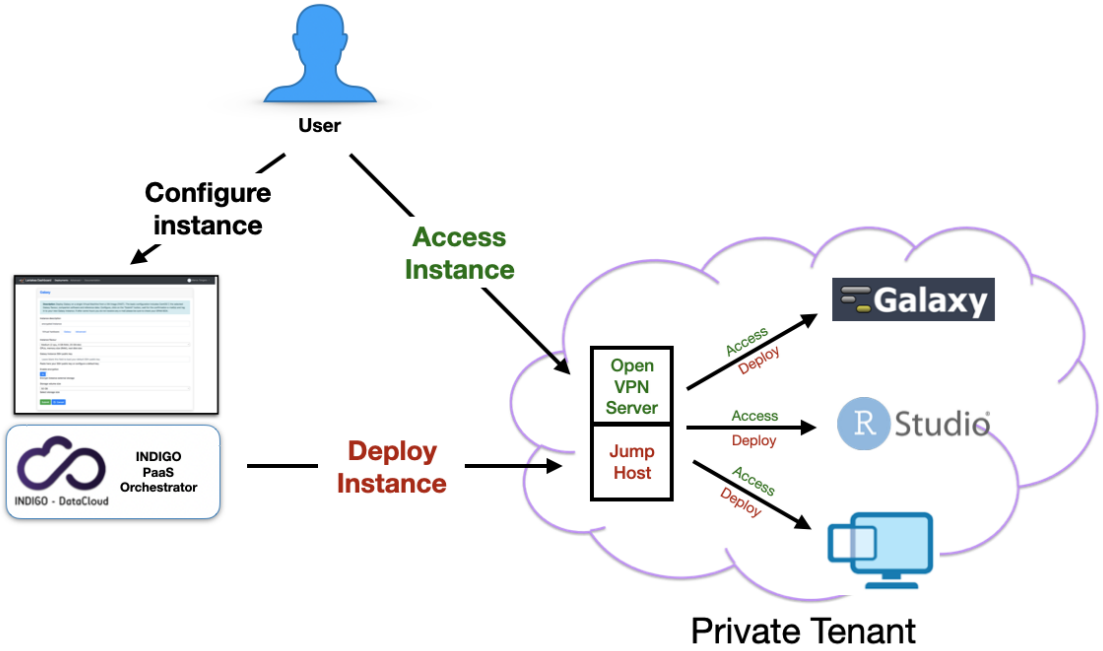


The encryption layer sits between the physical disk and the file system. Galaxy, or any other software is unaware of storage encryption, exploiting a specific mount point in order to store and retrieve files.

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

After the instance has been successfully deployed the user can retrieve his password through the Dashboard.

Service access



- Automatic deployments of virtual environments on private networks, exploiting the isolation features provided by OpenStack tenants and security groups to control the network access.
- User authentication through VPN, integrated with INDIGO IAM.
  - OpenIDConnect compliant

Conclusioni

Informazioni, approfondimenti, contatti:

donvito@infn.it

CONDIVISIONI
Conferenza GARR 2022
Palermo 18-20 maggio