Integrazione
e configurazione
dell'Identity and
Access Manager
INDIGO IAM su GARR
Cloud Platform

Pietro Mandreoli Università degli studi di Milano



OUTLINE

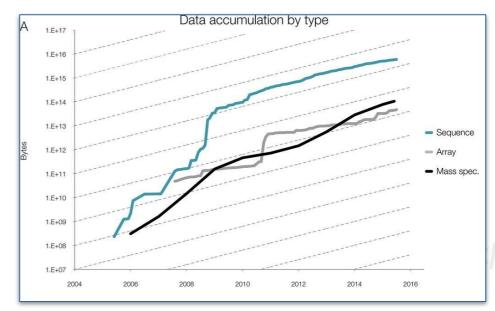
- I dati nelle Scienze della Vita
- Laniakea
- Laniakea@Recas
- Integrazione IAM su GARR Cloud Platform
- Prospettive future



Volume dati raddoppia ogni 6 mesi. Il trend è confermato per i prossimi 10 anni

Non solo quantità ma varietà!

Necessarie forme innovative di analisi e nuove infrastrutture



Crescita del volume dei dati presso EMBL-EBI Source: Charles E. Cook et al. Nucl. Acids Res.2016;44:D20-D26

Sono numerosi siti i che producono dati genomici in Europa:

Disciplina	Data size	N strumenti
HEP	15PB/year	1
Astronomy	15PB/year	1
Genomics	0.4TB/genome	>1000

Distribuzione dei dati in diversi centri di sequenziamento e su diverse infrastrutture IT



GDPR riconosce i dati genomici umani come dati sensibili, ma allo stesso tempo ne permette l'analisi, senza consenso esplicito, ai fini della ricerca scientifica a condizione che ciò sia consentito dal quadro legislativo dell'UE o degli Stati membri e che siano in atto misure di salvaguardia adeguate, per esempio la crittografia dei dati e la (pseudo)anonimizzazione.



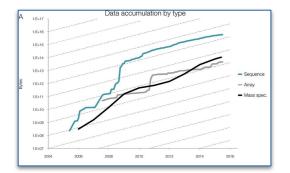


I tool l'analisi dei dati sono utilizzabili solo da riga di comando.

Non sono facilmente utilizzabili da chi non possegga adeguate competenze informatiche:

- familiarità con ambienti unix/linux
- programmazione Python, PERL...
- uso della riga di comando;
- ...





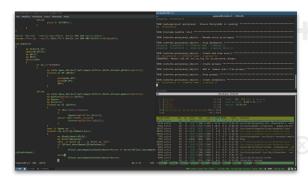
Dati in rapido aumento.



Dati distribuiti in Europa.



Dati genomici spesso protetti da GDPR.



Dati difficili da analizzare.









Servizio per la creazione di istanze Galaxy in Cloud. Basato su tecnologia INDIGO-DataCloud.

https://academic.oup.com/gigascience/article/9/4/giaa033/5816668

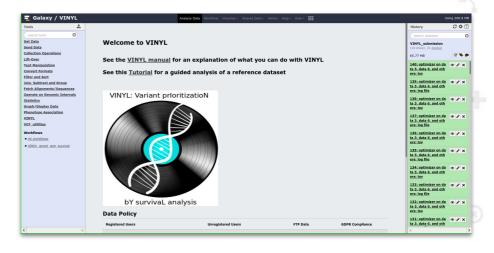


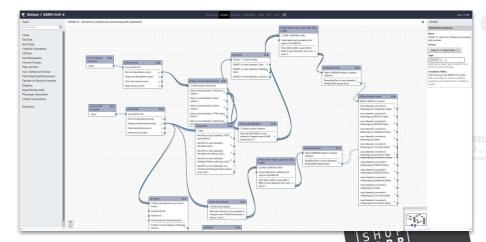




Galaxy è una piattaforma per la gestione di workflow per analisi di bioinformatica, open source e basato su interfaccia web.

- Promuove la riproducibilità delle analisi
- Permette la condivisione di tool e pipeline.
- Migliaia di tool disponibili
- Documentazione
- Più di 20000 utenti
- UI per la creazione di workflow





Perchè Galaxy su Cloud

SERVER PUBBLICI		
PRO	CONTRO	
Non sono richieste conoscenze IT particolari. Non richiede infrastruttura IT	Privacy dei dati	
Molti tools sono già installati e configurati	limiti nella quota utilizzabile	
Uso gratuito	nessuna possibilità di personalizzazione	

SERVER PRIVATI			
PRO	CONTRO		
Completamente personalizzabile.	Installazione richiede avanzate competenze IT		
controllo sugli utenti	Necessita di una infrastruttura IT adatta		
privacy dei dati	Costoso		

Laniakea

Laniakea nasconde la complessità e la potenza delle e-infrastrutture Cloud dietro l'interfaccia di Galaxy.

Consente agli utenti senza competenze IT di usare tool da interfaccia web (su cloud). Consente ai propri utenti di riutilizzare i tool a disposizione, migliorando la riproducibilità

delle analisi dati.





INDIGO-DataCloud



E' un progetto H2020 avviato nell'aprile 2015 e conclusosi nel settembre 2017, coordinato dall'Istituto Nazionale di Fisica Nucleare (INFN).

INDIGO ha come obiettivo lo sviluppo di una piattaforma dati e computazionale adatta a diverse comunità scientifiche, compatibile con diverse e-infrastrutture sia private che pubbliche.

Lo sviluppo di INDIGO continua nei progetti H2020:

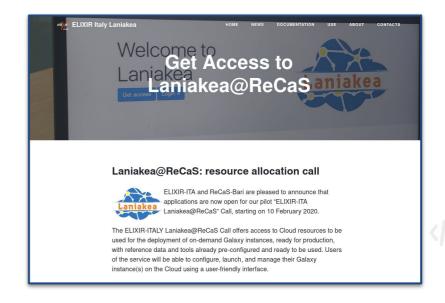
- Deep-Hybrid-DataCloud: https://deep-hybrid-datacloud.eu/
- Extreme-DataCloud: http://www.extreme-datacloud.eu/
- EOSC-Hub: https://www.eosc-hub.eu/

Salomoni, et al. 2018. "INDIGO-DataCloud: A Platform to Facilitate Seamless Access to E-Infrastructures." Journal of Grid Computing



Laniakea@ReCaS

- Operativo da Febbraio 2020
- Servizio offerto da Elixir-Italy
- Open-ended call: first-come first served.
 Board tecnico/scientifico per valutare le domande.
- Risorse su ReCaS Bari.
- Permette ai ricercatori un accesso più facile alle risorse di calcolo attraverso Laniakea.

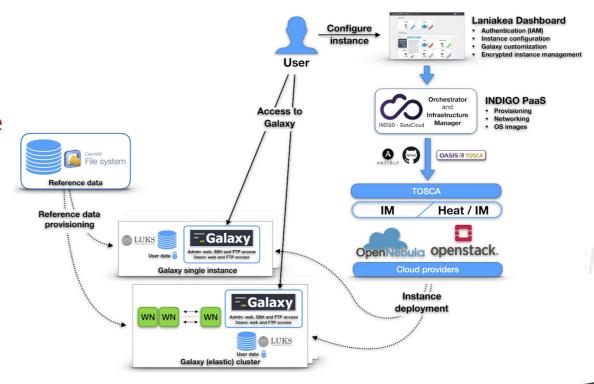




Architettura



- Dashboard Accesso user friendly alla configurazione ed al deployment di istanze Galaxy
- IAM servizio di autenticazione e autorizzazione
- INDIGO PaaS Deployment automatico di Galaxy
- Cloud Providers ReCaS-Bari
- Storage Persistente con o senza crittografia
- Disponibilità dei Reference data attraverso CERN-VM FS





Laniakea

- Creazione on-demand di istanze Galaxy.
- Supporto a VM, Docker e cluster di VM
- Istanze Galaxy con set di tool pre-configurati "Galaxy flavors".
- Supporto Reference data.
- Criptazione on-demand dello Storage associato all'istanza Galaxy.







Cosa succede quando non è possibile o non è conveniente spostare i dati?





Federazione delle infrastrutture

INDIGO Cloud stack consente:

- la federazione di infrastrutture cloud differenti attraverso IAM.
- di sfruttare una singola istanza della PaaS di INDIGO per creare e configurare VM e servizi sulle Cloud federate, in modo trasparente all'utente finale.

Laniakea dispone di tutti i componenti necessari per poter sfruttare questa possibilità, consentendo, ad esempio, la creazione istanze Galaxy presso il sito che dispone dei dati da analizzare, se federato*.

(*) Per i siti non federati la PaaS di INDIGO offre la possibilità di accedere alle risorse cloud attraverso username e password (solo utenti esperti).



WORK IN PROGRESS

Supporto GARR Cloud a Laniakea@ReCaS.

Potenziare il servizio ELIXIR-IT già in produzione da quasi un anno per offrire maggiori risorse e possibilità agli utenti ELIXIR(IT) e non.

IAM

- Supporto: SAML, Open ID Connect, Username/ password
- Servizio di registrazione (automatico o dall'admin)
- Gestione di utenti e gruppi.
- Gestione dei token.
- Servizio di delega: offre ai servizi (INDIGO PaaS) la possibilità "di operare per conto degli utenti".
- ..







Installazione di Laniakea su GARR Cloud

Ognuno dei micro-servizi che compongono la PaaS è stato installato su una VM differente su GARR cloud.

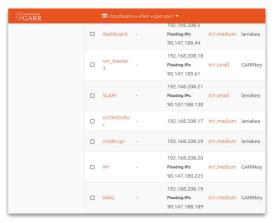
Il servizio IAM tra quelli installati è stato integrato su GARR cloud di pre-produzione.

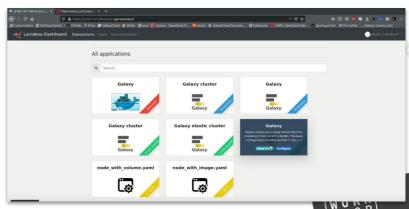
Test eseguiti:

- Deployment di VM (Ubuntu, CentOS)
- Deployment di Galaxy
- Deployment di Galaxy + cluster SLURM (elastico)
- Deployment Cluster Mesos

Documentazione step-by-step:

https://tinyurl.com/y8uoaaeg





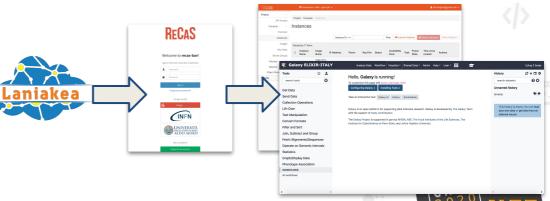


Work in progress:

Test dell'Integrazione dell'istanza IAM presente su ReCaS Bari andando però a garantire:

- Invariata la modalità di accesso a GARR cloud Platform (username/password, IDEM/eduGAIN, Google).
- Consentire agli utenti la possibilità di creare istanze Galaxy presso GARR Cloud, solo attraverso Laniakea.







Work in progress:

Test dell'Integrazione dell'istanza IAM presente su ReCaS Bari andando però a garantire:

- Invariata la modalità di accesso a GARR cloud Platform (username/password, IDEM/eduGAIN, Google).
- Consentire agli utenti la possibilità di creare istanze Galaxy presso GARR Cloud, solo attraverso Laniakea.

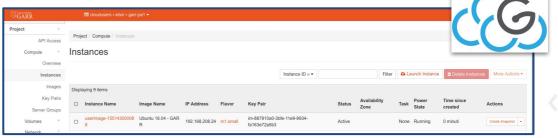






Creazione di istanze Galaxy sia su risorse ReCaS-Bari che su Cloud GARR in maniera trasparente, considerando le richieste dell'utente, del servizio e le risorse fornite dalle singole istanze cloud in modo da rendere semplice la scelta dell'istanza cloud da usare in base alle richieste.









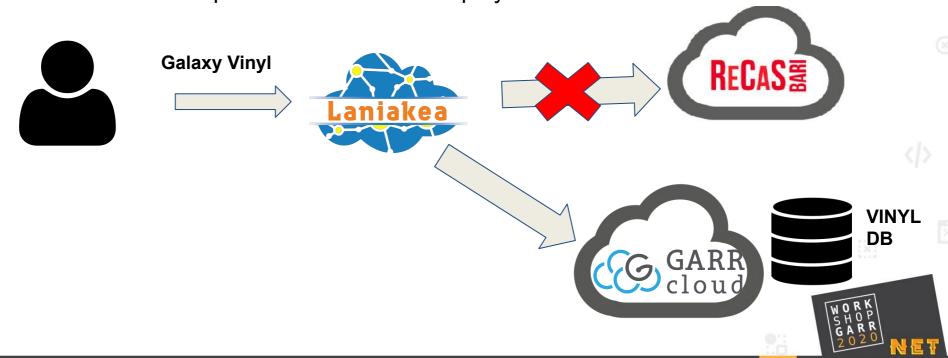
Ad esempio:

- Risorse virtuali disponibili sui siti Cloud
- Flavour (configurazioni di vCPU, RAM e storage) disponibili sul sito Cloud
- Immagini disponibili sul sito Cloud
- Dati disponibili
- Immagini richieste dagli utenti
- Risorse virtuali (vCPU, RAM, storage, ...) richieste dagli utenti
- Galaxy flavour (pre-set di tool) richieste dagli utenti





Se un flavour Galaxy ha dipendenze specifiche presenti in un datacenter (e.g Database varianti) è possibile rendere disponibile l'immagine del flavour sulla Cloud specifica in modo da indirizzare su quella infrastruttura il deployment dell'istanza.



Conclusioni

- È stata testata l'installazione di Laniakea su GARR eseguendo il deployment di istanze su un tenant GARR cloud di pre-produzione.
- Attualmente stiamo testando l'integrazione dell'istanza IAM presente presso il datacenter ReCaS@Bari su GARR cloud. Questo permetterà alla PaaS del servizio laniakea@ReCaS, già in fase di produzione, di poter eseguire il deployment delle istanze Galaxy su GARR cloud.



Link utili



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











M. Antonacci G. Donvito

G.Pesole

M.Chiara

F. Zambelli

M.Alverà

M. A. Tangaro















LANIAKEA@ReCaS

- Istituto Ortopedico Rizzoli
- Istituto Zooprofilattico Sperimentale della Puglia e della Basilicata
- Ospedale Pediatrico Giannina Gaslini
- University of Milan
- IBIOM-CNR
- University of Turin (training).

Galaxy pubblici:

- VYNIL
- CorGAT









prima integrazione di IAM su GARR

Per integrare IAM su garr cloud di preproduzione è stato necessario:

- Modificare su i 3 nodi keystone dei file di configurazione dell'identity service di openstack e di Apache
- mappare in modo specifico gli utenti IAM
- modificare la configurazione della dashboard Horizon

