

Laniakea@ReCaS: first year of activity of a Laniakea-based Galaxy "on-demand" service.

Pietro Mandreoli^{1,3}, Marco Antonio Tangaro^{1,2}, Matteo Chiara^{1,3}, Giacinto Donvito², Marica Antonacci², Graziano Pesole^{1,4}, Federico Zambelli^{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



GCC2021, 6-8 July, Virtual edition



The service aims to offer cloud resources for the creation of production-grade Galaxy instances to accelerate the development of novel bioinformatics tools and services, to facilitate large scale analysis of data and training activities in the context of different scientific projects.

Laniakea is primarily based on the PaaS layer developed by the INDIGO-DataCloud H2020 project and beyond, now largely adopted by the Italian National Institute for Nuclear Physics for its INFN-Cloud.

ELIXIR-IT Open Call

The Laniakea@ReCaS service was launched as a production service on February 2020 together with the associated open-ended call.



Here we present a selection of Laniakea@ReCaS use-cases.

Galaxy / VINYL

VINYL

Institute: University of Milan | IBIOM-CNR
Use: Analysis, development **Users:** 23 **Availability:** Public

VINYL (Variant prioritization BY survival analysis) is a tool suite that integrates an innovative method for variant prioritization. It was developed to mitigate the operator bias in this type of analysis. VINYL, relying on different software modules and an extensive collection of variants annotation databases that needs to be updated regularly (> 200GB), limiting the possibility of distributing it as a standalone tool. Instead, VINYL has been developed as a Galaxy workflow to allow straightforward access to the pipeline and distributed as a public instance through Laniakea@ReCaS.

Galaxy / CorGAT

CorGAT

Institute: University of Milan | IBIOM-CNR
Use: Analysis, Development **Users:** 17 **Availability:** Public

CorGAT (Coronavirus Genome Analysis Tool) is a novel, highly effective and user friendly approach for the functional annotation of SARS-CoV-2 genomes. It provides additional layers of annotation missed by other tools for the functional annotation of SARS-CoV-2 genomic mutations. These include variants associated with regulatory elements, TRS, or to consensus cleavage sites in the ORF1a and ORF1ab polyproteins and finally variants in conserved secondary structure elements.

The database CorGAT relies on requires frequent updates. To ease its distribution and make the usage more user-friendly and general, Laniakea@ReCaS was used to make CorGAT available as a dedicated, public and free to use Galaxy instance.

Galaxy 'S.I.R.I.O'

S.I.R.I.O

Institute: Istituto Zooprofilattico Sperimentale del Piemonte, Liguria e Valle d'Aosta (IZSTO)
Use: Analysis **N users:** 5 **Availability:** Private

IZSPs are a network of public labs in Italy for food safety and animal welfare. S.I.R.I.O. (Staphylococcus aureus Intensive server for Research In Omics data) is a Galaxy instance dedicated to genomic characterization of bacterial strains developed thanks to Laniakea@ReCaS by IZSTO.

With S.I.R.I.O., the IZSP laboratories network now share the same capability of genomic characterization of bacterial strains and the possibility to perform genomic analysis routinely avoiding the need for a local physical infrastructure and highly specialized bioinformatics personnel.

Galaxy UNITO

Learning Platform

Institute: Department of Life Sciences and Systems Biology (University of Turin)
Use: Training **Users:** 30 **Availability:** Private

UNITO teachers used Laniakea@ReCaS to organize hands-on bioinformatics courses. As well as being used during exercises the Galaxy instance was also employed to allow concurrent exams for more than 30 students effortlessly. The approach contributed to the realisation of the objectives of the Biological Sciences Course, providing the students with basic knowledge in the field of bioinformatics.

Galaxy L-PIPE-T

L-PIPE-T

Institute: Istituto Giannina Gaslini
Use: Analysis, Development **Users:** 20 **Availability:** Private

L-PIPE-T has been developed as a Galaxy instance for parsing, filtering, normalising, imputing, and analysing RT-qPCR data. The L-PIPE-T currently hosts 20 active users who ran over 500 analyses and stored over 100 histories and was successfully used to identify new potential biomarkers of hepatic injury and inflammation in a murine model of glycogen storage disease type 1a.

Galaxy rCASC

RCASC

Institute: University of Turin(UNITO)
Use: Development **Users:** 3 **Availability:** under development

rCASC provides a complete and Dockerized set of analysis tools and pipelines for Single-cell RNA-seq. rCASC tools allow performing all the typical steps of scRNAseq analysis, from creating the count matrix to cluster-specific markers detection.

The Galaxy implementation of rCASC, currently under development using Laniakea@ReCaS, will provide a user-friendly training environment to use and learn rCASC.

Galaxy EBD-CSIC

Conservation and Evolutionary Genetics

Institute: Estación Biológica de Doñana (Seville)
Use: Analysis **Users:** 20 **Availability:** Private

The conservation and Evolutionary Genetics Group, located in Seville employs a Laniakea@ReCaS Galaxy instance to analyse data from zoological remains from archaeological sites in Andalucía, Spain including Valencina, Dehesilla and Cueva Chica. Data generated from bones and teeth from these sites are analysed in order to identify the species of origin of bone fragments of archaeological relevance (e.g., associated with a ritual context) or relationships with living animals.

Galaxy IZSPB

Genotyping Bacterial Species

Institute: Istituto Zooprofilattico Sperimentale della Puglia e Basilicata (IZSPB)
Use: Analysis **Users:** 5 **Availability:** Private

IZSPB deployed a dedicated Galaxy instance for the identification of food pathogens. The instance is routinely employed to run workflows available on the Galaxy instance to perform de novo assembly and annotation of bacterial genomes and for the identification of antimicrobial resistance genes and virulence factors. The instance was used by different laboratories of the IZPB allowing the rapid classification, characterization, assessment of the virulence potential of any isolate.

Galaxy IOR

Rare Disease Mutation Detection

Institute: Istituto Ortopedico Rizzoli
Use: Analysis, Development **N Users:** 4 **Availability:** Private, Encrypted

The researchers of Istituto Ortopedico Rizzoli customised the CoVaCS (1) Galaxy instance for variant calling provided by the Laniakea@ReCaS service to facilitate the functional annotation and prioritization of genetic variants in patients affected by rare skeletal diseases. Since CoVaCS(reference) was optimized to work primarily with data produced by Illumina platforms, the workflow was modified to take into account the error profile of the Ion Torrent sequencing platform currently used by the institute.

The IOR Galaxy instance considerably sped up the data analysis process, becoming a crucial support to conduct in-depth analyses of genetic variants in diagnostic and research contexts.

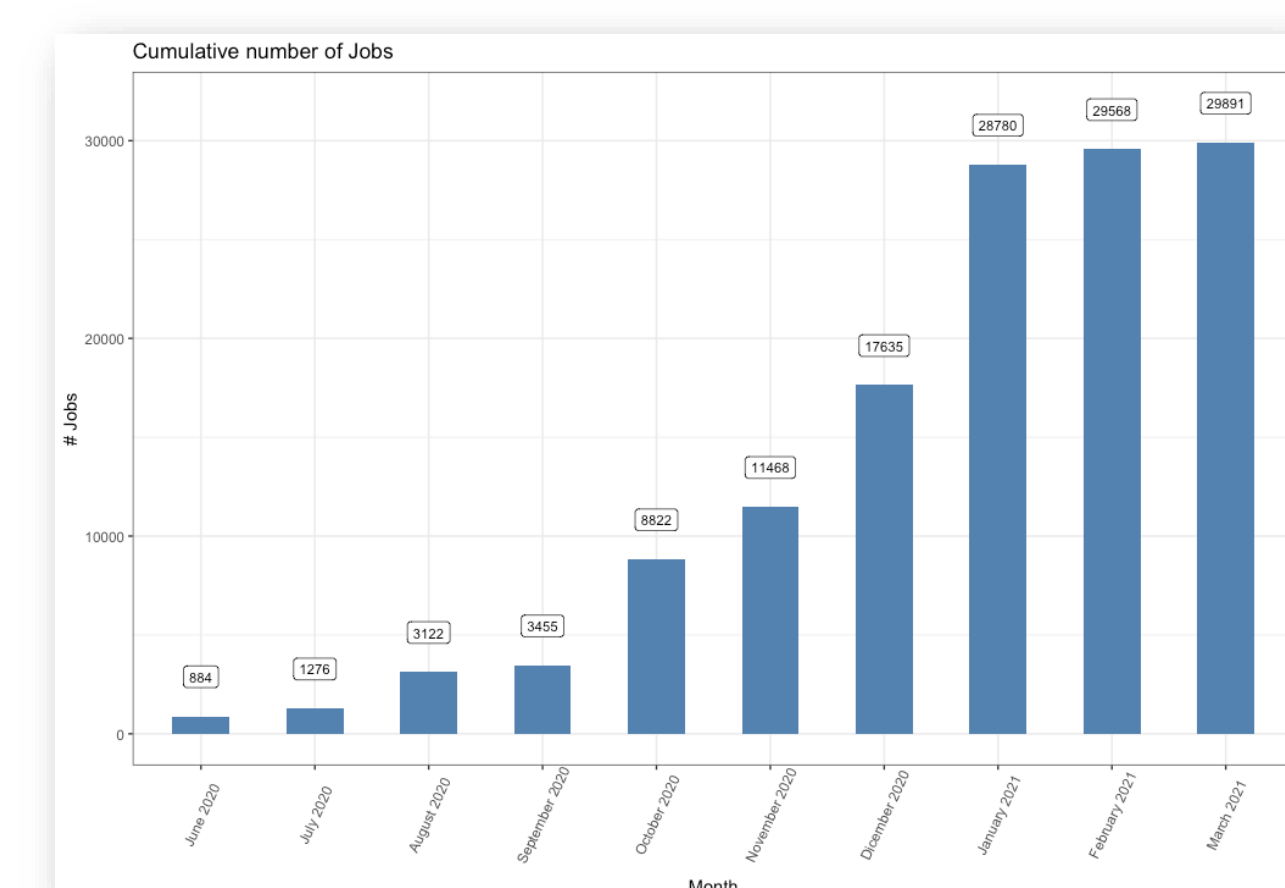
1. <https://bmcgenomics.biomedcentral.com/articles/10.118/s12864-018-4508-1>

Conclusions and future perspectives

Since its opening, Laniakea@ReCaS has accepted proposals for a total of 20 Galaxy instances operating on the ReCaS infrastructure for a total resource budget of 130 vCPUs, 250 GB of RAM and 5 TB of storage. The service proved to be stable and reliable, ensuring the seamless operation of all the Galaxy instances.

Aside from handling the use-cases data analysis needs in terms of resources and tools, Laniakea@ReCaS has proved to be an interesting platform to quickly develop and make available novel Galaxy based public services like VINYL, CorGAT, and L-PIPE-T.

In the meantime, the development and update of the Laniakea software platform, the one on which Laniakea@ReCaS leans, has never stopped: new applications, update of the backend and frontend software and systems are currently in testing phase for the next release. These updates and new features, developed also following the feedback of Laniakea@ReCaS users, will soon be implemented in the production service, providing users with an expanded set of tools beyond Galaxy and an even more reliable environment. Finally, Laniakea@ReCaS is expected to be soon extended to exploit also cloud resources provided by the GARR Cloud, the Italian cloud for university and research.



Cumulative number of jobs launched by all the Galaxy instances deployed on Laniakea@ReCaS up to March 2021.

Laniakea@ReCaS links and info



Contact

- Pietro Mandreoli pietro.mandreoli@unimi.it
- 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

