EGI Conference 2022



Contribution ID: 22

Type: Poster

Exploring reference data through existing computing services for the bioinformatics community : an EOSC-Pillar use-case

Tuesday, 20 September 2022 19:00 (1 hour)

Galaxy is a widely adopted workflow management system for bioinformatics, aiming to make computational biology accessible to research scientists that do not have computer programming or systems administration experience.

How can scientists connect this useful, reproducibility-oriented tool seamlessly with many data sources? How can they do so in a coherent way using different instances of Galaxy? Can they run it locally or on a secured infrastructure that handles patient data? Can they compare the results of those different scenarios?

The proposed poster presents the work done as part of one of the EOSC-Pillar project's scientific use-cases to address those questions, achieving the following objectives:

- Allow access to reference data from different Galaxy deployments to all EOSC users.
- Facilitate the deployment of Galaxy instances in the same infrastructure hosting the data to analyse
- Provide coherency in the deployment of different Galaxy instances
- Ensure sensitive (e.g., health) data security requirements are met throughout the process

The poster describes four scientific scenarios based on concrete needs from the ELIXIR community. It also describes the technical services the use-case is relying on, namely: Laniakea (Galaxy as a service provided by IBIOM-CNR and INFN), Inserm data repository, IFB cloud Galaxy instances and the INDIGO-IAM authentication service provided by INFN. It demonstrates the interest of EOSC Pillar's Federated Data Space (F2DS) for connecting different data sources to the Galaxy in a simple and coherent way.

The poster also highlights the need to conform to data protection regulations concerning health personal data, by deploying Galaxy in a private, secured environment while still ensuring the data analysis workflow remains similar to its public counterpart.

Finally, it shows proposed solutions to provide access to the service to all users within the EOSC community through roles management and by integrating it into a global authentication framework.

Any relevant links

https://eosc-pillar.eu/use-cases/exploring-reference-data-through-existing-computing-services-bioinformatics-community

Topic

Data Spaces

Primary authors: MATHIEU, Gilles (INSERM); Mrs SANAA, Yosra (Inserm); BLANCHET, Christophe (French Institute of Bioinformatics - CNRS UMS3601); Mr MANDREOLI, Pietro (CNR-IBOM, University of Milan); TAN-GARO, Marco Antonio (CNR); DONVITO, Giacinto (INFN); Mrs FOGGETTI, Nadina (INFN); ANTONACCI, Marica (INFN); Mrs BELHAJ SALEM, Marwa (Inserm) **Co-authors:** Mrs BURLOT, Laura (French Institute of Bioinformatics (IFB)); Mr LORENZO, Jonathan (French Institute of Bioinformatics (IFB)); Mr COLOMBO, Daniele (University of Milan); ZAMBELLI, Federico (CNR (ELIXIR-ITA)); Mr SALGADO, David (Inserm, Aix-Marseille University)

Presenter: MATHIEU, Gilles (INSERM)

Session Classification: Posters (presenters at poster)

Track Classification: Data Spaces