

ELIXIR Cloud & AAI: FAIR Data & Analytics Services for Human Data Communities

Thursday, 27 August 2020 16:00 (30 minutes)

Currently, patient data is geographically dispersed, difficult to access, and often stored in siloed project-specific databases preventing large-scale data aggregation, standardisation, integration/harmonisation and advanced disease modelling. The ELIXIR Cloud and Authentication & Authorisation Infrastructure (AAI) for Human Data Communities project aims to leverage a coordinated network of ELIXIR Nodes to deliver a Global Alliance for Genomic Health (GA4GH) standards-compliant FAIR federated environment to enable population-scale genomic and phenotypic data analysis across international boundaries and a potential infrastructure to enable 1M Genome analysis. The Global Alliance for Genomics and Health (GA4GH) is a policy-framing and technical standards-setting organization, seeking to enable responsible genomic data sharing within a human rights framework.

The ELIXIR Cloud & AAI project will lay the groundwork to deliver the foundational FAIR capability of “federation” of identities, sensitive data access, trusted hybrid HPC/Cloud providers and sensitive data analysis services across ELIXIR Nodes by underpinning the bi-directional conversation between partners with the GA4GH standards and specifications and ELIXIR trans-national expertise. The project is also developing a framework for secure access and analysis of sensitive human data based on national federations and standardised discovery protocols. The secure authentication and authorisation process alongside guidelines and compliance processes is essential to enable the community to use this data without compromising privacy and informed consent.

Indico rendering error

Could not include image: [404] Error fetching image

Running complex data analysis workflows from the command line often requires IT and programming experience that make such workflows inaccessible to many scientists. Constant changes from new software versions, different operating systems and HPC/Cloud installations add to the complexity. To address this, it is vital to work not only with tool developers but also with users and infrastructure providers, to enable users to access workflows composed by different bioinformatics software containers that can be easily reused and deployed in both academic HPC and commercial clouds. Well documented, containerised workflows are also inherently reproducible, thus addressing one of the key challenges in computational life science. The project provides a FAIR repository to store bioinformatics software containers and workflows, a FAIR data registry to discover and resolve the locations of datasets and a standardised and interoperable workflow and task execution service to leverage the federated life-science infrastructure of ELIXIR.

The project therefore provides a globally available curated FAIR repository to store bioinformatics software containers and workflows (Biocontainers - GA4GH TRS), a FAIR service to discover and resolve the locations of datasets (RDSDS - GA4GH DRS) and a distributed workflow and task execution service (WES-ELIXIR/TESK - GA4GH WES/TES) to leverage the federated life-science infrastructure of ELIXIR.

The ambition of the project is to provide a global ecosystem of joint sensitive data access and analysis services where federated resources for life science data are used by national and international projects across all life science disciplines, with widespread support for standard components securing their long-term sustainability. Connecting distributed datasets via common standards will allow researchers unprecedented opportunities to detect rare signals in complex datasets and lay the ground for the widespread application of advanced data analysis methods in the life sciences.

Primary author: Dr VARMA, Susheel (EMBL-EBI)

Co-authors: Dr KANITZ, Alexander (University of Basel); Mr CRISTOFORI, Andrea (EMBL-EBI); Mrs NIEWIELSKA, Ania (EMBL-EBI); Dr GRUENING, Bjorn (University of Frieberg); Dr HARROW, Jennifer (ELIXIR); Ms CHIEN, Jinny (EMBL-EBI); Dr TEDDS, Jonathan (ELIXIR); Mr TÖRNROOS, Juha (CSC); Mr SAYERS, Kevin (University of Basel); Dr CAPELLA-GUTIERREZ, Salvador (BSC); Dr BUTCHER, Sarah (EMBL-EBI); Mr KAPOOR,

Shubham (CSC); Mr DE, Soumyadip (EMBL-EBI); Dr NEWHOUSE, Steven (EMBL-EBI); Mr NYRÖNEN, Tommi (CSC); Dr PEREZ-RIVEROL, Yasset (EMBL-EBI)

Presenters: Dr NEWHOUSE, Steven (EMBL-EBI); Dr VARMA, Susheel (EMBL-EBI)

Session Classification: Infrastructure Clouds and Virtualisation Session

Track Classification: Infrastructure Clouds and Virtualisation