



Brief Overview

This demonstrator will leverage EOSC resources to enhance science reproducibility of datasets uploaded to the European Genome-phenome Archive (EGA). By doing this the new dataset will also be made available in a FAIR manner, adding metadata according to the attributes that have been chosen to contribute the strongest to the FAIR principles. Pipelines will be developed as part of this demonstrator to automate the later process. This pilot will have a pragmatic impact by demonstrating how to make analyses portable (tools and workflows), how to increase findability, how to leverage security technologies for sensible data, how to deploy the workflow into a cloud and how to make data FAIR. It will also have a long-term impact by increasing the usability of EGA hosted data by assuring to potential users that up-to-date versions of an assured quality are available to download.

Objectives

The European Genome-phenome Archive (EGA) (<https://ega-archive.org>) is a repository that facilitates access and management for long-term archival of bio-molecular data. Enhancing data analysis reproducibility and exploring new added-value services by leveraging EOSC resources are the main objectives of this SD. Applying the FAIR principles (Findability, accessibility, interoperability and reusability) to our data sets and information associated is a great mission we have accepted from the community.



**EUROPEAN
GENOME-PHENOME
ARCHIVE**

10TH ANNIVERSARY



EOSC_{pilot}

The European Open Science
Cloud for Research Pilot Project

Main achievements

- » A set of results data has been reproduced using a portable version of the pipeline.
- » The same result set has been updated by re-analyzing it with a current version of the pipeline and the reference data.
- » FAIRfied metadata on both result sets is available at a testing EGA server and/or at an appropriate repository

Recommendations for the implementation

Being in possession of huge amounts of data is a first step but not enough to achieve the main goal: foster research. There exists a need for adding usefulness to the bio-molecular data the repositories currently store. The EOSC project is a unique framework to add this necessary layer of standardization and interoperability while unifying and discovering the files and associated pipelines

Partners of the SD

Centre for Genomic Regulation (CRG), Leiden University Medical Center (LUMC) and European Bioinformatics Institute (EBI)



“Bioinformatics tools and pipelines are very sensitive to small details, like versions of the tools, computing environment or parameter values provided; having a comprehensive view of the analysis applied on the raw data to get to the final results will greatly enhance the reproducibility of the data deposited at the EGA and the papers associated to them. Given that managing tools and pipelines are out of the core scope of the EGA, leveraging EOSC resources for that purpose, and linking EGA Studies to them, will be a significant step forward in the added value offered to the community”. Jordi Rambla EGA Head at CRG.

Contacts

✉ petr.knoth@open.ac.uk

