Contribution ID: **27** Type: **Oral Presentation**

Open-source and cloud-native solutions for managing and analyzing heterogeneous and sensitive clinical Data

Friday, 25 March 2022 13:50 (20 minutes)

The requirement for an effective handling and management of heterogeneous and possibly confidential data continuously increases within multiple scientific domains.
PLANET (Pollution Lake ANalysis for Effective Therapy) is a INFN-funded research initiative aiming to implement an observational study to assess a possible statistical association between environmental pollution and Covid-19 infection, symptoms and course. PLANET builds on a "data-centric" based approach that takes into account clinical components, environmental and pollution conditions, complementing primary data and many eventual confounding factors such as population density, commuter density, socio-economic metrics and more . Besides the scientific one, the main technical challenge of the project is about collecting, indexing, storing and managing many types of datasets guaranteeing FAIRness as well as adherence to the prescribed regulatory frameworks, such as the GDPR.
In this contribution we describe the developed open-source DataLake platform, detailing its key features: the event-based storage system centered on MinIO, which automates metadata processing; the data pipeline, implemented via Argo Workflows; the GraphQL-based mechanisms to query object metadata; finally, the seamless integration of the platform with a compute multi-user environment, showing how all these frameworks are integrated in the Enhanced PrIvacy and Compliance (EPIC) Cloud partition of the INFN-Cloud federation.

Primary authors: COSTANTINI, Alessandro (INFN-CNAF); MARTELLI, Barbara (INFN - CNAF); DUMA, Cristina (Istituto Nazionale di Fisica Nucleare, CNAF); SALOMONI, Davide (INFN); Dr SPIGA, daniele (INFN-PG); CIANGOTTINI, Diego (Istituto Nazionale di Fisica Nucleare, Sezione di Perugia); RONCHIERI, Elisabetta (INFN CNAF); Mrs SERGI, Giusy (INFN-CNAF); Mr GASPARETTO, Jacopo (INFN-CNAF); Prof. STORCHI, Loriano (Dipartimento di Farmacia Universita' G. d'Annunzio); TRACOLLI, Mirco (Istituto Nazionale di Fisica Nucleare, Sezione di Perugia); Dr LUBRANO, Pasquale (INFN -Sezione Perugia); CUTINI, Sara (INFN Perugia)

Presenter: Dr SPIGA, daniele (INFN-PG)

Session Classification: Data Management & Big Data

Track Classification: Track 6: Data Management & Big Data