

## BITS :: Call for Abstracts 2023 - Oral communication

<i>Type</i>	Oral communication
<i>Session</i>	Biological data management
<i>Title</i>	Alleanza Contro Il Cancro: secure data management for multicentric clinical trials in the HBD-DataCloud.
<i>All Authors</i>	Arnaud Ceol (1), Barbara Martelli (2), for the ACC genomics and bioinformatics working groups and INFN.

<i>Affiliation</i>	(1) European Institute of Oncology IRCCS, Milan, Italy. (2) INFN, CNAF Division, Bologna, Italy
--------------------	--

*Motivation*

Alliance Against Cancer (ACC) is an Italian Oncology Network of more than 28 Institutes with the objectives to promote translational research and to quickly bring precision medicine of excellence in national oncology. ACC, together with the networks of Neuroscience and Neurorehabilitation and Cardiology, is now part of the Health Big Data (HBD) project, funded by the MEF and coordinated by the Ministry of Health and managed by the three Networks in collaboration with the Politecnico of Milan and INFN. In the context of the Health Big Data project (HBD), several projects and trials are ongoing, most of them supported by the transversal genomics and radiomics workgroups. To permit the sharing, tracking, and processing of the data generated by the project, the ACC bioinformatics team of the genomics workgroup and INFN CNAF have jointly developed a secure, centralized cloud platform named HDB-DataCloud.

*Methods*

To follow the “need-to-know” principle, the HBD-DataCloud platform is implemented as a dual tenant of an OpenStack cloud. The first tenant (ACC-TEST) is used for testing the installation and updates of the virtual machines and software and should not store patient data. All updates on the TEST tenant are repeated on the ACC-PROD tenant, where the ACC data is stored, and only a restricted group of administrators has access.

The HBD-DataCloud platform is composed of the following stacks:

- An authentication and authorization server, based on Keycloak (<https://www.keycloak.org/>), with multi-factor authentication (MFA) enabled. This server enables fine grained access based on application, project, institute, and role.
- A data storage and sharing server, based on OneData (<https://onedata.org>),
- A web server hosts the ACC LIMS, which allows to register and track on a pseudonymized form all patient and data from the ACC projects,
- A computation server, for automatized processing of the data and bioinformatics research,
- A log server, which receives logs and audits from all serves of the platform,
- A backup and archiving system able to store data in remote locations for disaster recovery.

The platform is managed in the context of the Information Security Management System of EPIC Cloud (Enhanced Privacy and Compliance Cloud, <https://www.cnaf.infn.it/en/epic-cloud-en/>), the region of INFN Cloud certified under information security standards like ISO/IEC 27001:2013, ISO/IEC 27017:2015 and ISO/IEC 27018:2017.

*Results*

For ACC projects, patient, image, sample and sequencing information is registered to the ACC LIMS platform. Images and sequences data is uploaded to the OneData platform deployed for ACC. Data collected on the platform are of the following types: i) genomics: germline and tumor samples, DNA and RNA, in BAM or VCF format; and ii) radiomics: radiomic features belonging to different families: morphological, textural, statistical, in DICOM format.

The granularity of the authorizations allows each researcher or manager to be assigned access to data of one or more types, projects, or institutes in compliance with the data minimization, security by design and security by default principles stated in the GDPR. A unique barcode of the image stack or sample sequencing is automatically generated by the system (ACC LIMS) and is used to name uploaded files, ensuring their traceability throughout the project lifecycle.

The computational resources of the platform allow the validation and processing of the uploaded data. The ACC Bioinformatics Workgroup implements the automation of reporting for ACC studies.

More than 80 researchers have been registered up to now on the platform and have registered more than 2800 patients for 5 projects currently ongoing, with genomics data, radiomics data or both. The amount of “live” data available on the platform is 5,5 TB (plus backups and remote archive copies).

<i>Info</i>	-
-------------	---

<i>filename</i>	bits2023-ceol-acc.png
-----------------	-----------------------

<i>Figure</i>	
---------------	--

**Tracking and reporting**

**ACC Lung Report**

**Authentication (MFA) and Authorization**

**Storage and sharing**

**Automatic, Centralized Processing and reporting**

Availability -

**Corresponding Author**

Name, Surname Arnaud, Ceol

Email arnaud.ceol@ieo.it

Submitted on 02.05.2023

**Società Italiana di Bioinformatica**  
 C.F. / P.IVA 97319460586  
 E-mail bits@bioinformatics.it

Sede Legale Viale G. Mazzini, 114/B - 00195 Roma  
 Website bioinformatics.it