

WORK  
SHOP  
GARR  
2024

NET  
MAKERS

# Managing patient data across national networks: Examples from the HEALTH BIG DATA cloud platform.

Arnaud Ceol





Milano, 13 settembre - **L'Istituto Europeo di Oncologia e il Centro Cardiologico Monzino** sono primi in Italia per il quarto anno consecutivo nella classifica "*World Best Specialised Hospital 2024*" elaborata dal settimanale internazionale **Newsweek**, in collaborazione con la società di ricerca "Statista".



EU-LIFE is an alliance of independent European research institutes in the life sciences. EU-LIFE contributes to the improvement of research by **influencing European science policies** and by developing, implementing and disseminating **best practices in the organisation and management** of research institutes.





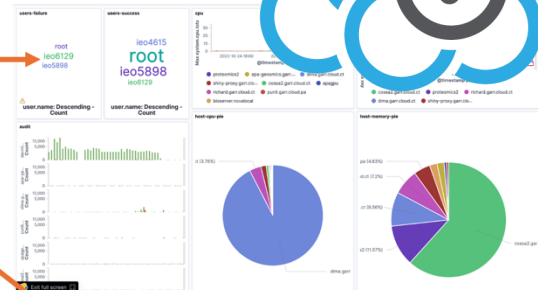
**1994-2019**  
25 anni di ricerca e innovazione  
per la lotta al cancro,  
25 anni di Istituto Europeo di Oncologia



```

aform > benchmarks > terraform.tfvars > image_name
research_instance_floatingip = ""
user_name                    = "ieo3004"
project_name                 = "benchmarks"
volume_size                  = "500"
flavor                       = "d1.xlarge"
gocryptfs_key                = ""
teleport_token               = "f48a343e2d4de784659093eafd19e042"
availability_zone            = "nova"
image_name                   = "RockyLinux-8 - GARR"
network_name                 = "minsalute-IEO-VM-net"
add_data_volume              = true
assign_floating_ip           = false

```



Resources	Address	Labels	Connect
ieo3004	...	...	CONNECTED
ieo3005	...	...	CONNECTED
ieo3006	...	...	CONNECTED
ieo3007	...	...	CONNECTED
ieo3008	...	...	CONNECTED
ieo3009	...	...	CONNECTED
ieo3010	...	...	CONNECTED
ieo3011	...	...	CONNECTED
ieo3012	...	...	CONNECTED
ieo3013	...	...	CONNECTED
ieo3014	...	...	CONNECTED
ieo3015	...	...	CONNECTED
ieo3016	...	...	CONNECTED
ieo3017	...	...	CONNECTED
ieo3018	...	...	CONNECTED
ieo3019	...	...	CONNECTED
ieo3020	...	...	CONNECTED

```

test.py
Time = 14, 1, 2, 3
partition = 10, 200, 300
p1 = p1[0:10]
p2 = p2[0:10]
p3 = p3[0:10]
TESTER: B.A. "PARTITION (100%)
300
250

```



**1994-2019**  
 25 anni di ricerca e innovazione  
 per la lotta al cancro,  
 25 anni di Istituto Europeo di Oncologia



```

terraform.tfvars > image_name
research_instance_floatingip = ""
user_name                    = "ieo3004"
project_name                 = "benchmarks"
volume_size                  = "500"
flavor                       = "d1.xlarge"
gocryptfs_key                = ""
teleport_token               = "f48a343e2d4de784659093eafd19e042"
availability_zone            = "nova"
image_name                   = "RockyLinux-8 - GARR"
network_name                 = "minsa-lute-IEO-VM-net"
add_data_volume              = true
assign_floating_ip           = false
  
```



**1994-2019**  
 25 anni di ricerca e innovazione  
 per la lotta al cancro,  
 25 anni di Istituto Europeo di Oncologia



**KPI CLINICAL DATA PLATFORM**  
 Analisi KPI arricchimento Clinical Data Platform IEO



## IEO charts the future of healthcare with Value Based Medicine

Laife Reply collaborates with IEO to innovate diagnosis-therapeutic paths through personalized medicine, combining AI and Big Data in the Value Based Medicine platform.

#Artificial Intelligence  
 #Big Data  
 #Healthcare

The platform's objective is to use clinical information to predict, through statistical and artificial intelligence models, the outcomes of therapeutic and interventional treatment options, prioritising the patient's quality of life, both in terms of health and psychophysical well-being.



European Journal of  
 Volume 183, April 2023, Page

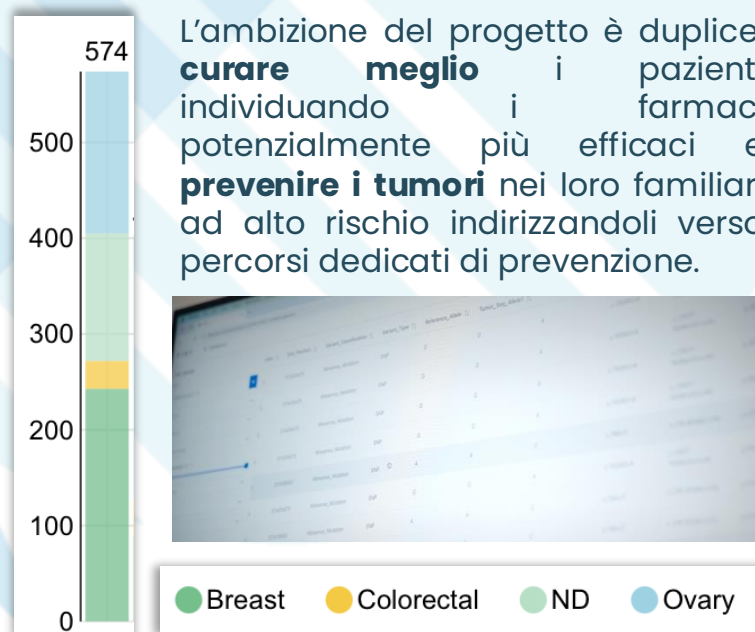
Original Research

### Molecular tumour board at European Institute of Oncology: Report of the first three year activity of an Italian precision oncology experience

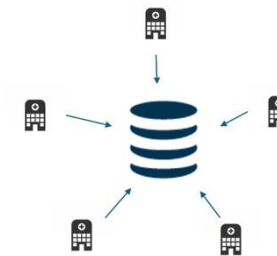
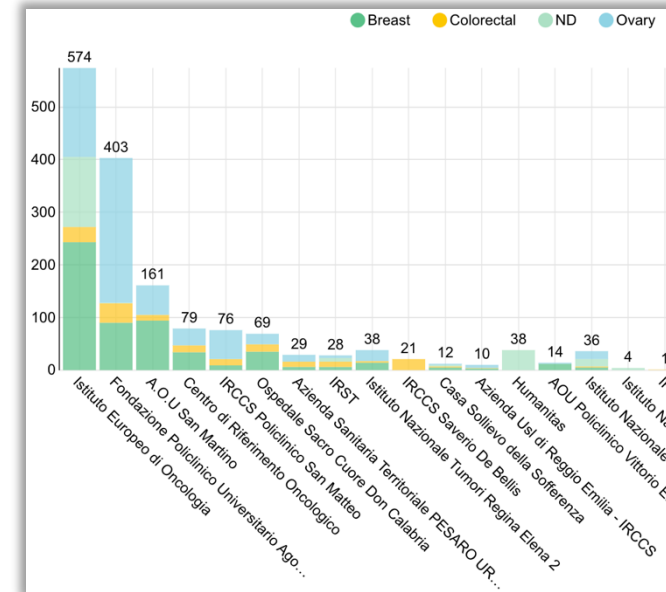
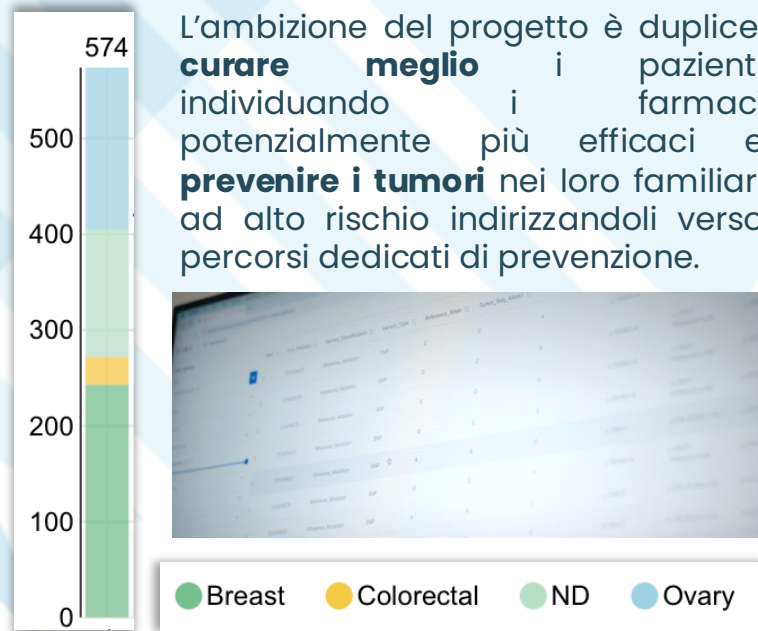
Matteo Repetto <sup>a,b,1</sup>, Edoardo Crimini <sup>a,b,1</sup>, Luca Boscolo Bielo <sup>a,b</sup>, Elena Guerini-Rocco <sup>b,d</sup>, Liliana Ascione <sup>a,b</sup>, Andrea Bonfanti <sup>b</sup>, Cristina Zanzottera <sup>d</sup>, Luca Mazzarella <sup>a</sup>, Alberto Ranghiero <sup>c</sup>, Carmen Belli <sup>a</sup>, Carmen Criscitiello <sup>a,b</sup>, Angela Esposito <sup>a</sup>, Massimo C.P. Barberis <sup>c</sup>, Giuseppe Curigliano <sup>a,b</sup>



**GerSom:** validazione di un pannello genico a **costi contenuti** ed eseguibile in **tempi brevi** (<1 settimana) per l'analisi di 467 geni alterati nei tumori, tra cui 172 CPGs.



**GerSom:** validazione di un pannello genico a **costi contenuti** ed eseguibile in **tempi brevi** (<1 settimana) per l'analisi di 467 geni alterati nei tumori, tra cui 172 CPGs.



Compliant with the standards of security (ISO/IEC 27001-27017-27018) and privacy (GDPR)



Compliant with the standards of security(ISO/IEC 27001-27017-27018) and privacy (GDPR)



## Extraction/Ingestion (WG1, WG2, WG3)

### Estrazione di:

- Informazioni scritte in linguaggio naturale nei documenti clinici tramite NLP
  - Informazioni dai dati omici, dai segnali, dalle immagini e dalla radiomica
  - Dati clinici (CCE e CTMS)
- nel **formato più conveniente** all'interpretazione e all'**integrazione** tra essi

## Management ed Integration (WG4, WG5)

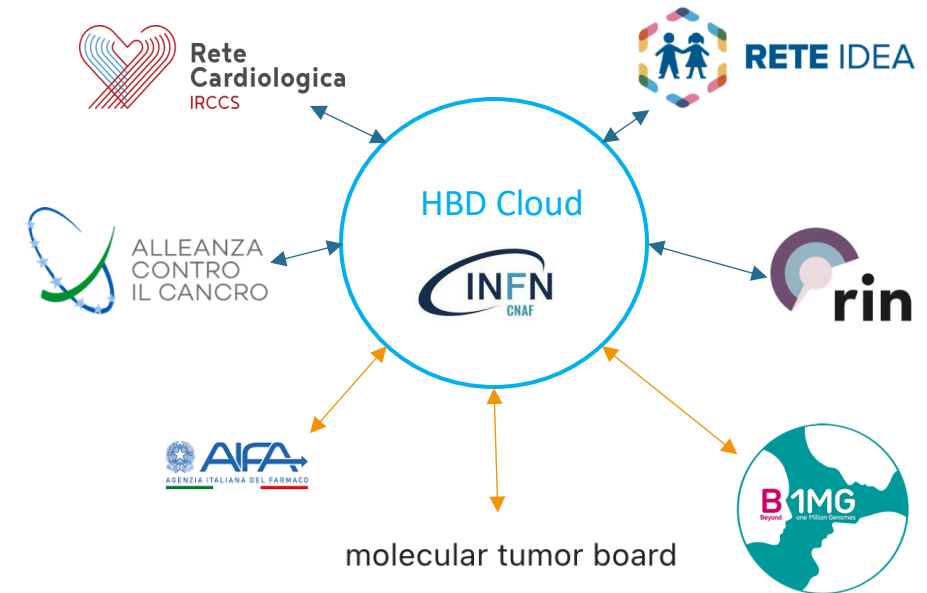
- Identificazione ed implementazione di **strumenti metodologici e tecnologici** necessari per l'estrazione e l'integrazione dei dati
- Identificazione ed implementazione della **soluzione logica architetturale e dell'infrastruttura fisica** della piattaforma cloud integrata e federata HBD

## Analysis (WG6)

Elaborazione di **sistemi di health analytics** modulari e a complessità crescente per i dati raccolti e strutturati nelle architetture di progetto

## Privacy, Sicurezza ed Etica (WG7)

Identificazione ed implementazione di soluzioni per **privacy e di sicurezza**, e ad altre criticità di **carattere etico** che potrebbero presentarsi



Compliant with the standards of security (ISO/IEC 27001-27017-27018) and privacy (GDPR)

# The CNAF – ACC platform


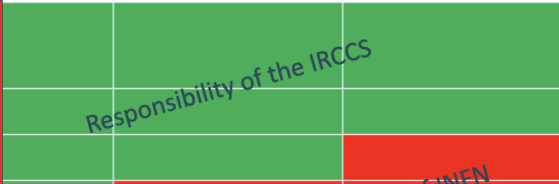
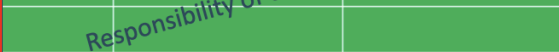








Information Security Management System Certificate  
**ISO/IEC 27001:2013**

We certify that the the Information Security Management System of the Organization:

**ISTITUTO NAZIONALE DI FISICA NUCLEARE**

is in compliance with the Standard UNI CEI EN ISO/IEC 27001:2017 for the following products/services:

Hosting of physical and virtual systems for biomedical data access and conservation, and management of data analytics applications targeted to biomedical and genomic research.  
 Delivering cloud services in IaaS mode by applying ISO/IEC 27017:2015 and ISO/IEC 27018:2019 guidelines.

Responsibility model	IaaS (Infrastructure as a Service)	PaaS (Platform as a Service)	SaaS (Software as a Service)
GRC (Security Governance Risk & Compliance)			
Data Security			
Application Security			
Platform Security			
Infrastructure Security	 <b>ISO 27001</b>	 <b>ISO 27001</b>	 <b>ISO 27001</b>
Physical Security	 <b>ISO 27001</b>	 <b>ISO 27001</b>	 <b>ISO 27001</b>

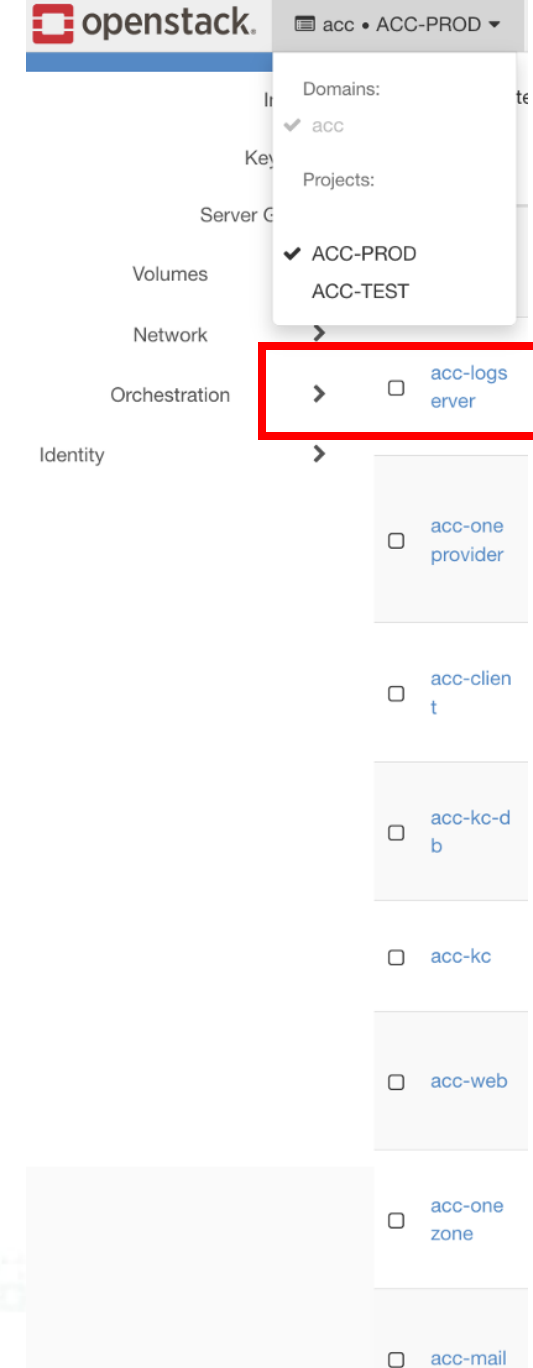
2 twin tenants:

- **Test tenant**

- Install and test software, configuration, updates
- No sensible data

- **Production tenant**

- Reproduce the test installation
- Stores ACC data
- Only accessible by authorized users (ACC)



# Authentication and authorization



Istituto Nazionale di Fisica Nucleare  
CNAF  
R25bis MODULO ACCESSO RISORSE EPIC Cloud  
Sistema di Gestione della Sicurezza delle Informazioni certificato ISO/IEC 27001  
Rev. 2.3 del 17/03/2022

## Modulo di autorizzazione all'accesso alle risorse EPIC Cloud

Il/la sottoscritto/a \_\_\_\_\_  
email \_\_\_\_\_ Telefono: \_\_\_\_\_  
documento Carta di identità \_\_\_\_\_ Num. Documento: \_\_\_\_\_  
ricercatore/collaboratore/dipendente di Fondazione Policlinico Universitario A. Gemelli IRCSS

CHIEDE

l'accesso alle risorse EPIC Cloud (Enhanced Security and Privacy Cloud) - Sistema di Gestione di Sicurezza delle Informazioni certificato UNI CEI ISO/IEC 27001:2014 ISO/IEC 27017:2015 ISO/IEC 27018:2019 definito all'interno della sede operativa CN dell'INFN in Viale Bert Pichat 6/2 a Bologna, per la seguente motivazione (*indicare il nome del progetto ed il ruolo ricoperto all'interno di esso*):

ACC e-platform, ricercatore

Risorse e relativi progetti ACC a quale si richiede l'accesso:

Risorse: LIMS, OneData. Progetti: Radiomics (tutti) di tutti i centri coinvolti

Nome del Referente presso ACC:

Dr. Luca Boldrini

Nome del Referente presso l'INFN-CNAF o RTP (Responsabile Tecnico del Progetto):

Barbara Martelli

DICHIARA

di aver ricevuto e di accettare le condizioni specificate nei seguenti documenti:

- Regolamento per l'accesso e l'utilizzo delle risorse INFN  
[https://www.cnaf.infn.it/wp-content/uploads/2020/03/Disciplinare\\_2020\\_IT.pdf](https://www.cnaf.infn.it/wp-content/uploads/2020/03/Disciplinare_2020_IT.pdf)
- Informativa sul trattamento dei dati personali (art. D.Lgs. 30 giugno 2003, n.196)  
[https://dpo.infn.it/wp-content/uploads/2019/01/Informativa\\_generale\\_INFNO\\_181204.pdf](https://dpo.infn.it/wp-content/uploads/2019/01/Informativa_generale_INFNO_181204.pdf)
- Norme per il trattamento dei dati personali  
[https://dpo.infn.it/wp-content/uploads/2018/10/Norme\\_uso\\_sistemi\\_informatici.pdf](https://dpo.infn.it/wp-content/uploads/2018/10/Norme_uso_sistemi_informatici.pdf)
- Linee Guida sulla Sicurezza Informatica INFN consegnate contestualmente a questo modulo

Platform access module + video call



Institutional public IP



HBD Cloud

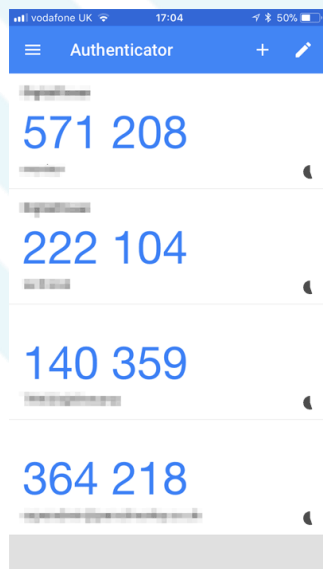
IP whitelist

# Authentication and authorization

Two factor authentication  
(password + cell phone)



<https://www.keycloak.org/>



Log In

Username  
arnaud.ceol

Password  
.....

Log In

New user? [Register](#)

KEYCLOAK

Acc

Configure

- Realm Settings
- Clients
- Client Scopes
- Roles
- Identity Providers
- User Federation
- Authentication

Manage

- Groups**
- Users
- Sessions
- Events
- Import
- Export

User Groups

Groups Default Groups

Search... View all groups

- acclims-admin
- acclims-genomics-gers-readwrite-01\_jeo
- acclims-genomics-lung-read-01\_IEO
- acclims-genomics-lung-readwrite-01\_jeo
- acclims-genomics-lung-readwrite-03\_asmn
- acclims-genomics-lung-readwrite-04\_cro
- acclims-genomics-lung-readwrite-05\_fpo
- acclims-genomics-lung-readwrite-07\_h
- acclims-genomics-lung-readwrite-08\_irst
- acclims-genomics-lung-readwrite-10\_iov
- acclims-genomics-lung-readwrite-13\_intre
- acclims-genomics-lung-readwrite-14\_osr
- acclims-genomics-lung-readwrite-15\_aousm
- acclims-genomics-lung-readwrite-17\_itgpil
- acclims-genomics-lung-readwrite-18\_css
- acclims-genomics-lung-readwrite-19\_aoupve
- acclims-genomics-lung-readwrite-21\_ircscrocrb
- acclims-genomics-lung-readwrite-22\_orvs
- acclims-genomics-lung-readwrite-24\_aoup
- acclims-genomics-lung-readwrite-26\_intre2

# Manage and monitor

ACC Lims Thursday, 22nd June 2023 01:10 Welcome, Arnaud. D

Home » Accbarcode » Patients

Keyword  Project: ACC Gersom Institution  Search 942 results 3000 total

<input type="checkbox"/>	Barcode	Project	Enrolment id	Institution	Internal id	Gender	Samples	Studies	Annotations
<input type="checkbox"/>	GERS0106DH	ACC Gersom	13-083	Istituto Europeo di Oncologia	13-083	Female	3 +	0 +	0 +
<input type="checkbox"/>	GERS0108CN	ACC Gersom	13-228	Istituto Europeo di Oncologia	13-228	Unknown	1 +	0 +	0 +
<input type="checkbox"/>	GERS010BRS	ACC Gersom	13-071	Istituto Europeo di Oncologia	13-071	Female	1 +	0 +	0 +
<input type="checkbox"/>	GERS010CE6	ACC Gersom	791-021	Istituto Europeo di Oncologia	791-021	Unknown	1 +	0 +	0 +
<input type="checkbox"/>	GERS010D8Y	ACC Gersom	13-004	Istituto Europeo di Oncologia	13-004	Female	1 +	0 +	0 +
<input type="checkbox"/>	GERS010ITK	ACC Gersom	13-212	Istituto Europeo di Oncologia	13-212	Unknown	2 +	0 +	0 +
<input type="checkbox"/>	GERS010MOD	ACC Gersom	13-062	Istituto Europeo di Oncologia	13-062	Female	1 +	0 +	0 +
<input type="checkbox"/>	GERS010NQO	ACC Gersom	13-254	Istituto Europeo di Oncologia	13-254	Unknown	1 +	0 +	0 +
<input type="checkbox"/>	GERS010X3G	ACC Gersom							
<input type="checkbox"/>	GERS0114W5	ACC Gersom							
<input type="checkbox"/>	GERS011BK3	ACC Gersom							
<input type="checkbox"/>	GERS011GYI	ACC Gersom							
<input type="checkbox"/>	GERS011IRI	ACC Gersom							
<input type="checkbox"/>	GERS011KZQ	ACC Gersom	13-067	Istituto Europeo di Oncologia	13-067	Unknown	2 +	0 +	0 +
<input type="checkbox"/>	GERS011TH1	ACC Gersom	13-092	Istituto Europeo di Oncologia	13-092	Unknown	2 +	0 +	0 +
<input type="checkbox"/>	GERS011TWA	ACC Gersom	13-195	Istituto Europeo di Oncologia	13-195	Unknown	1 +	0 +	0 +
<input type="checkbox"/>	GERS0125SE	ACC Gersom	791-043	Istituto Europeo di Oncologia	791-043	Unknown	7 +	0 +	0 +
<input type="checkbox"/>	GERS0127ZU	ACC Gersom	SR D	Istituto Europeo di Oncologia	SR D	Unknown	1 +	0 +	0 +

<input type="checkbox"/>	Barcode	Sequencing id	Vial id	Analyte	Sequencing Batch
<input type="checkbox"/>	lung01AL31-2TAR201	2	A	RNA	38 2019-10-25 Istituto Europeo di Oncologia
<input type="checkbox"/>	lung01AL31-2TAD101	1	A	DNA	38 2019-10-25 Istituto Europeo di Oncologia
<input type="checkbox"/>	lung01AL31-1NAD101	1	A	DNA	38 2019-10-25 Istituto Europeo di Oncologia

KEYCLOAK

Acc

Configure

- Realm Settings
- Clients
- Client Scopes
- Roles
- Identity Providers
- User Federation
- Authentication

Manage

- Groups
- Users
- Sessions
- Events
- Import
- Export

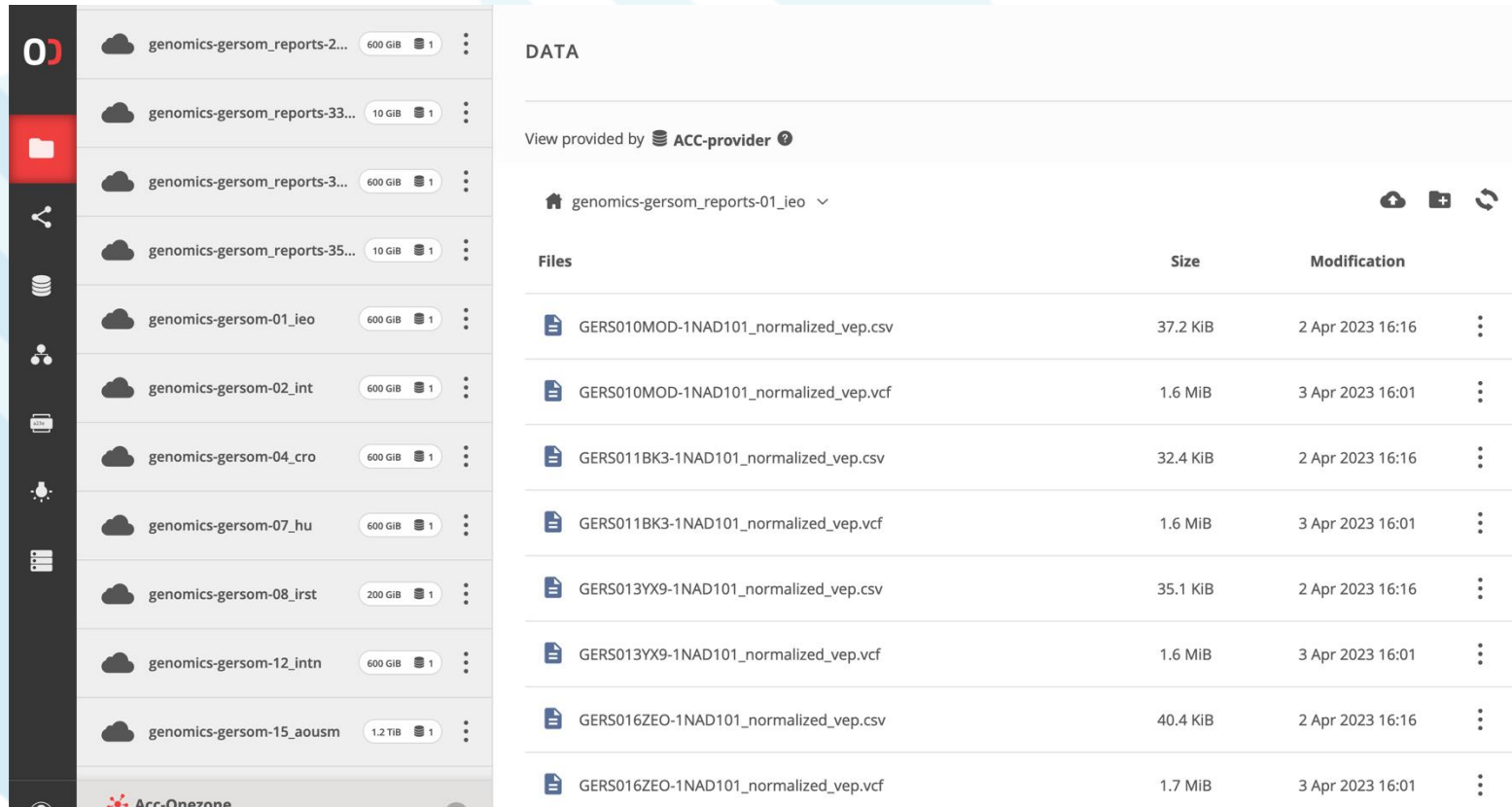
User Groups

Groups Default Groups @

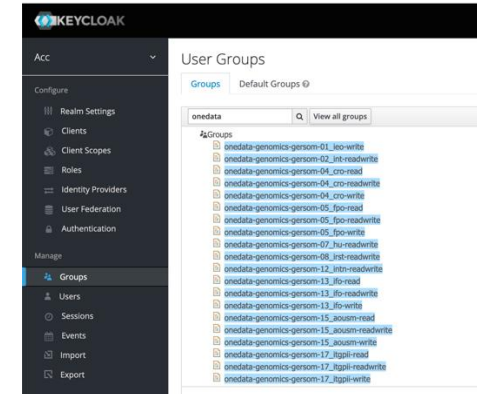
Search... View all groups

- Groups
- acclms-admin
- acclms-genomics-gers-readwrite-01\_jeo
- acclms-genomics-lung-read-01\_jeo
- acclms-genomics-lung-readwrite-01\_jeo
- acclms-genomics-lung-readwrite-03\_asnm
- acclms-genomics-lung-readwrite-04\_cro
- acclms-genomics-lung-readwrite-05\_fpo
- acclms-genomics-lung-readwrite-07\_in
- acclms-genomics-lung-readwrite-08\_inst
- acclms-genomics-lung-readwrite-10\_jov
- acclms-genomics-lung-readwrite-13\_intr
- acclms-genomics-lung-readwrite-14\_osr
- acclms-genomics-lung-readwrite-15\_aousm
- acclms-genomics-lung-readwrite-17\_aggpl
- acclms-genomics-lung-readwrite-18\_cis
- acclms-genomics-lung-readwrite-19\_oupev
- acclms-genomics-lung-readwrite-21\_jccscrobrb
- acclms-genomics-lung-readwrite-22\_ovns
- acclms-genomics-lung-readwrite-24\_soup
- acclms-genomics-lung-readwrite-26\_intr2

# Data exchange & storage



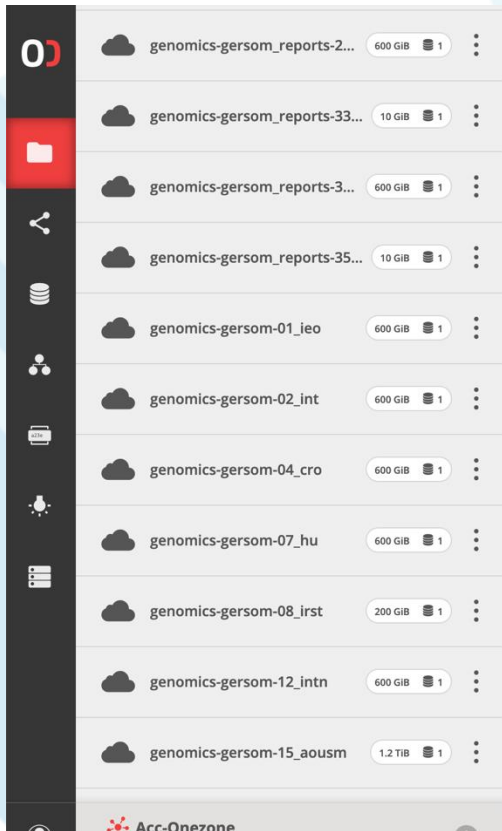
Files	Size	Modification
gersom-gersom_reports-01_ieo	600 GiB	1
gersom-gersom_reports-02_int	600 GiB	1
gersom-gersom_reports-03...	10 GiB	1
gersom-gersom_reports-04_cro	600 GiB	1
gersom-gersom_reports-05...	600 GiB	1
gersom-gersom_reports-06...	10 GiB	1
gersom-gersom_reports-07...	600 GiB	1
gersom-gersom_reports-08...	600 GiB	1
gersom-gersom_reports-09...	600 GiB	1
gersom-gersom_reports-10...	600 GiB	1
gersom-gersom_reports-11...	600 GiB	1
gersom-gersom_reports-12...	600 GiB	1
gersom-gersom_reports-13...	600 GiB	1
gersom-gersom_reports-14...	600 GiB	1
gersom-gersom_reports-15...	600 GiB	1
gersom-gersom_reports-16...	600 GiB	1
gersom-gersom_reports-17...	600 GiB	1
gersom-gersom_reports-18...	600 GiB	1
gersom-gersom_reports-19...	600 GiB	1
gersom-gersom_reports-20...	600 GiB	1
gersom-gersom_reports-21...	600 GiB	1
gersom-gersom_reports-22...	600 GiB	1
gersom-gersom_reports-23...	600 GiB	1
gersom-gersom_reports-24...	600 GiB	1
gersom-gersom_reports-25...	600 GiB	1
gersom-gersom_reports-26...	600 GiB	1
gersom-gersom_reports-27...	600 GiB	1
gersom-gersom_reports-28...	600 GiB	1
gersom-gersom_reports-29...	600 GiB	1
gersom-gersom_reports-30...	600 GiB	1
gersom-gersom_reports-31...	600 GiB	1
gersom-gersom_reports-32...	600 GiB	1
gersom-gersom_reports-33...	600 GiB	1
gersom-gersom_reports-34...	600 GiB	1
gersom-gersom_reports-35...	600 GiB	1
gersom-gersom_reports-36...	600 GiB	1
gersom-gersom_reports-37...	600 GiB	1
gersom-gersom_reports-38...	600 GiB	1
gersom-gersom_reports-39...	600 GiB	1
gersom-gersom_reports-40...	600 GiB	1
gersom-gersom_reports-41...	600 GiB	1
gersom-gersom_reports-42...	600 GiB	1
gersom-gersom_reports-43...	600 GiB	1
gersom-gersom_reports-44...	600 GiB	1
gersom-gersom_reports-45...	600 GiB	1
gersom-gersom_reports-46...	600 GiB	1
gersom-gersom_reports-47...	600 GiB	1
gersom-gersom_reports-48...	600 GiB	1
gersom-gersom_reports-49...	600 GiB	1
gersom-gersom_reports-50...	600 GiB	1
gersom-gersom_reports-51...	600 GiB	1
gersom-gersom_reports-52...	600 GiB	1
gersom-gersom_reports-53...	600 GiB	1
gersom-gersom_reports-54...	600 GiB	1
gersom-gersom_reports-55...	600 GiB	1
gersom-gersom_reports-56...	600 GiB	1
gersom-gersom_reports-57...	600 GiB	1
gersom-gersom_reports-58...	600 GiB	1
gersom-gersom_reports-59...	600 GiB	1
gersom-gersom_reports-60...	600 GiB	1
gersom-gersom_reports-61...	600 GiB	1
gersom-gersom_reports-62...	600 GiB	1
gersom-gersom_reports-63...	600 GiB	1
gersom-gersom_reports-64...	600 GiB	1
gersom-gersom_reports-65...	600 GiB	1
gersom-gersom_reports-66...	600 GiB	1
gersom-gersom_reports-67...	600 GiB	1
gersom-gersom_reports-68...	600 GiB	1
gersom-gersom_reports-69...	600 GiB	1
gersom-gersom_reports-70...	600 GiB	1
gersom-gersom_reports-71...	600 GiB	1
gersom-gersom_reports-72...	600 GiB	1
gersom-gersom_reports-73...	600 GiB	1
gersom-gersom_reports-74...	600 GiB	1
gersom-gersom_reports-75...	600 GiB	1
gersom-gersom_reports-76...	600 GiB	1
gersom-gersom_reports-77...	600 GiB	1
gersom-gersom_reports-78...	600 GiB	1
gersom-gersom_reports-79...	600 GiB	1
gersom-gersom_reports-80...	600 GiB	1
gersom-gersom_reports-81...	600 GiB	1
gersom-gersom_reports-82...	600 GiB	1
gersom-gersom_reports-83...	600 GiB	1
gersom-gersom_reports-84...	600 GiB	1
gersom-gersom_reports-85...	600 GiB	1
gersom-gersom_reports-86...	600 GiB	1
gersom-gersom_reports-87...	600 GiB	1
gersom-gersom_reports-88...	600 GiB	1
gersom-gersom_reports-89...	600 GiB	1
gersom-gersom_reports-90...	600 GiB	1
gersom-gersom_reports-91...	600 GiB	1
gersom-gersom_reports-92...	600 GiB	1
gersom-gersom_reports-93...	600 GiB	1
gersom-gersom_reports-94...	600 GiB	1
gersom-gersom_reports-95...	600 GiB	1
gersom-gersom_reports-96...	600 GiB	1
gersom-gersom_reports-97...	600 GiB	1
gersom-gersom_reports-98...	600 GiB	1
gersom-gersom_reports-99...	600 GiB	1
gersom-gersom_reports-100...	600 GiB	1



```
[arnaud.ceol@acc-client wp0_bot]$ export ONEPROVIDER_HOST=acc-oneprovider.cnaf.infn.it
[arnaud.ceol@acc-client wp0_bot]$ export ONEPROVIDER_TOKEN=MDAyNmMxY2F0aW9uIGFjYy1vbmV6b251LmNmLmUyYTVhMzZkNWU4ODk1OGMzM2U3ZTAwZWVhOGZlNmJyY2gxYjE1CjAwMmVjajQgaW50ZmZlY2YyYWNlID00b251Y2xwZm50OCjAw
[arnaud.ceol@acc-client wp0_bot]$ oneclient -t $ONEPROVIDER_TOKEN -H $ONEPROVIDER_HOST onemount
Connecting to provider 'acc-oneprovider.cnaf.infn.it:443' using session ID: '4862659865158986788'
Getting configuration...
Oneclient has been successfully mounted in 'onemount'.
[arnaud.ceol@acc-client wp0_bot]$ ls onemount/
acclatpadmin      genomics-gersom-33_scdc      genomics-gersom-reports-26_ifo2
genomics-archives genomics-gersom-35_A0SCF    genomics-gersom-reports-27_psm
genomics-gersom-01_ieo genomics-gersom-35_OUSCF    genomics-gersom-reports-33_scdc
genomics-gersom-02_int genomics-gersom-refdata     genomics-gersom-reports-35_A0SCF
genomics-gersom-04_cro genomics-gersom-reports-01_ieo genomics-gersom-reports-35_OUSCF
genomics-gersom-07_hu genomics-gersom-reports-02_int genomics-lung-05_fpo
genomics-gersom-08_inst genomics-gersom-reports-04_cro genomics-lung-07_hu
genomics-gersom-12_intn genomics-gersom-reports-07_hu genomics-lung-08_inst
genomics-gersom-15_aouism genomics-gersom-reports-08_inst genomics-lung-10_iov
genomics-gersom-17_itgpii genomics-gersom-reports-12_intn genomics-lung-13_ifo1
genomics-gersom-19_aoupve genomics-gersom-reports-15_aouism genomics-lung-14_osr
genomics-gersom-22_orvs genomics-gersom-reports-17_itgpii genomics-lung-15_aouism
genomics-gersom-25_fpgu genomics-gersom-reports-19_aoupve genomics-lung-18_css
genomics-gersom-26_ifo2 genomics-gersom-reports-22_orvs genomics-lung-19_aoupve
genomics-gersom-27_psm genomics-gersom-reports-25_fpgu genomics-lung-21_irccscrocbrcb
[arnaud.ceol@acc-client wp0_bot]$
```

<https://www.onedata.org>

# Data exchange & storage

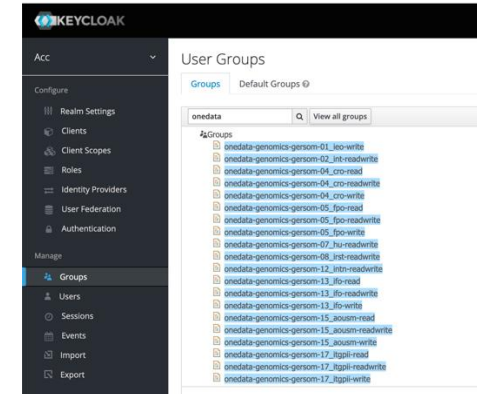


DATA

View provided by ACC-provider

genomics-gersom\_reports-01\_jeo


Files	Size	Modification
gers010MOD-1NAD101_normalized_vep.csv	37.2 KiB	2 Apr 2023 16:16
gers010MOD-1NAD101_normalized_vep.vcf	1.6 MiB	3 Apr 2023 16:01
gers011BK3-1NAD101_normalized_vep.csv	32.4 KiB	2 Apr 2023 16:16
gers011BK3-1NAD101_normalized_vep.vcf	1.6 MiB	3 Apr 2023 16:01




<https://rucio.cern.ch/>

```
[arnaud.ceol@acc-client wp0_bot]$ export ONEPROVIDER_HOST=acc-oneprovider.cnaf.infn.it
[arnaud.ceol@acc-client wp0_bot]$ export ONEPROVIDER_TOKEN=MDA9NmxvY2F00aW9uIGFjYy1vbmV6b251LmNm...
```

- Distributed data centres (federated platform)
- Large data management



<https://www.onedata.org>

# Tracking, Monitoring, QC

Django REST framework

Schema / Sample List

Sample List

GET /api/samples/

```
HTTP 200 OK
Allow: GET, HEAD, OPTIONS
Content-Type: application/json
Vary: Accept

{
  "count": 3249,
  "next": "https://acc-web.cnaf.infn.it/api/samples/?page=2",
  "previous": null,
  "results": [
    {
      "uuid": "95067c54-bfa2-4564-8cdf-8ce04adb4004",
      "id": 940,
      "internal_id": null,
      "barcode": "LLUNG220PRZ-IN",
      "sample_type": "germline",
      "project": "ACC Lung / RATIONALE",
      "ircs": "Ospedali Riuniti Villa Sofia",
      "sample_cellularity": null
    }
  ]
}
```

Client (python)

```
# Get sample sequencing information given a barcode
def getSampleSequencing(self, barcode):

    schema = self.client.get(self.acclimsApiUrl)

    params = {"barcode": barcode, }

    result = self.client.action(schema, ["samplesequencings", "list"], params)

    if len(result['results']) == 0:
        logger.error("Barcode not found : %s" % barcode)
        return None

    return result
```

Bot (Audit)

```
def doAudit(self):

    patientBarcodes = []
    errors = []

    try :
        logger.info("Start auditing")

        # Get files to process
        sourceFiles = self.getSourceFileList()

        # Get list of barcodes:
        patientBarcodes = self.acclimsClient.getPatientBarcodes()

        # to upper
        for i in range(len(patientBarcodes)):
            patientBarcodes[i] = patientBarcodes[i].upper()
```

For each file uploaded  
 ↓  
 Has it a valid extension (bam, vcf) ?  
 Is the naming correct (barcode.extension)?

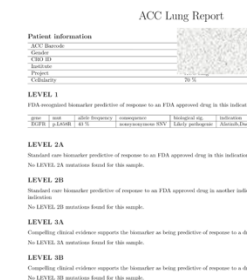
↓  
 Is the barcode in the LIMS?

↓  
 Is there a md5? Is the md5 correct?

↓  
 Is the file valid (samtools, picard)

↙  
 Update the LIMS

↓  
 Generate Report





# Tracking, Monitoring, QC

Django REST framework

Schema / Sample List

Sample List

GET /api/samples/

```
HTTP 200 OK
Allow: GET, HEAD, OPTIONS
Content-Type: application/json
Vary: Accept

{
  "count": 3249,
  "next": "https://acc-web.cnaf.infn.it/api/samples/?page=2",
  "previous": null,
  "results": [
    {
      "uuid": "95067c54-bfa2-4564-8cdf-8ce04adb4004",
      "id": 940,
      "internal_id": null,
      "barcode": "LLUNG220PRZ-IN",
      "sample_type": "germline",
      "project": "ACC Lung / RATIONALE",
      "ircs": "Ospedali Riuniti Villa Sofia",
      "sample_cellularity": null
    }
  ]
}
```

Client (python)

```
# Get sample sequencing information given a barcode
def getSampleSequencing(self, barcode):

    schema = self.client.get(self.acclimsApiUrl)

    params = {"barcode": barcode, }

    result = self.client.action(schema, ["samplesequencings", "list"], params)

    if len(result['results']) == 0:
        logger.error("Barcode not found : %s" % barcode)
        return None

    return result
```

Bot (Audit)

```
def doAudit(self):

    patientBarcodes = []
    errors = []

    try :
        logger.info("Start auditing")

        # Get files to process
        sourceFiles = self.getSourceFileList()

        # Get list of barcodes:
        patientBarcodes = self.acclimsClient.getPatientBarcodes()

        # to upper
        for i in range(len(patientBarcodes)):
            patientBarcodes[i] = patientBarcodes[i].upper()
```

For each file uploaded  
 ↓  
 Has it a valid extension (bam, vcf) ?  
 Is the naming correct (barcode.extension) ?

↓  
 Is the barcode in the LIMS?

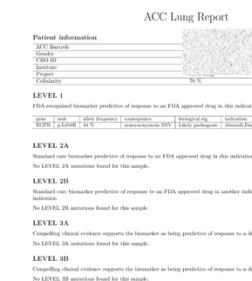
↓  
 Is there a md5? Is the md5 correct?

↓  
 Is the file valid (samtools, picard)

Update the LIMS

Generate Report

Fill the data catalogue



# Validazione dataset GerSom: FHIR & OMOP



## Rete ACC

- Applicazione degli **standard di interoperabilità OMOP e FHIR**
- Definizione **linee guida per la strutturazione dei dataset** dei progetti di ricerca per garantire l'interoperabilità
- Implementazione del **Data Catalog** della Rete ACC presso il nodo INFN
- Implementazione di una soluzione di **Molecular Tumor Board** sfruttando le esperienze in atto presso IRE e HSR
- Identificazione di ambiti di **gestione di immagini (anatomia patologica) e dati omici (Gersom)**.

VARIABILI GENERALI DATASET			FOLLOW UP DATASET	
FHIR	OMOP		FHIR	OMOP
84%	46%	Variabili allineate agli standard FHIR e OMOP in termini di nomenclatura	80%	42%
11%	9%	Variabili parzialmente disallineate rispetto agli standard FHIR e OMOP	17%	10%
5%	45%	Variabili disallineate rispetto agli standard FHIR e OMOP	3%	48%

### VARIABILI NON ALLINEATE

Variabili che includono **più concetti** al loro interno o che richiedono informazioni aggiuntive/iterative rispetto ad altre variabili

Variabili che **includono aggettivi di tempo** come successivo/post, precedente/antecedente, contemporaneamente

Utilizzo di **nomenclatura specifica del Dataset GerSom** (esempio: GerSom guidelines, GerSom test)



- Riformulare singoli elementi del **Dataset GerSom**
- Avanzare la proposta di **estensione dei linguaggi FHIR e OMOP**
- Definizione di **linee guida di strutturazione dei dataset** di progetti di ricerca affinché sia garantita l'interoperabilità applicabile a tutte le Reti

<https://www.hl7.org/fhir/>

# Data catalog: data profiling



- **Data Profiling:** Analyzes and evaluates datasets to understand their **structure, quality, and characteristics**.
- **Benefits:** Provides insights for **informed decision-making** on data cleaning, transformation, and usage.
- **Application:** Essential for insights in Health Big Data, supporting Business Intelligence and Advanced Analytics.



Apache Atlas



Key	New Value
Allele Frequency Spectrum (3)	0: 952, 1: 2005, 2: 5
Average Variant Quality	6089.4644
Base Changes (13)	CG: 82, CT: 258, TC: 262, GA: 217, AG: 230, None: 244, CA: 39, GT: 62, AC: 50, GC: 78, TG: 53, AT: 29, TA: 41
Command Line	N/A
File Date	20240113
File Format	VCFv4.1
Memory Size (MB)	0.98588085
Numero Campioni	1

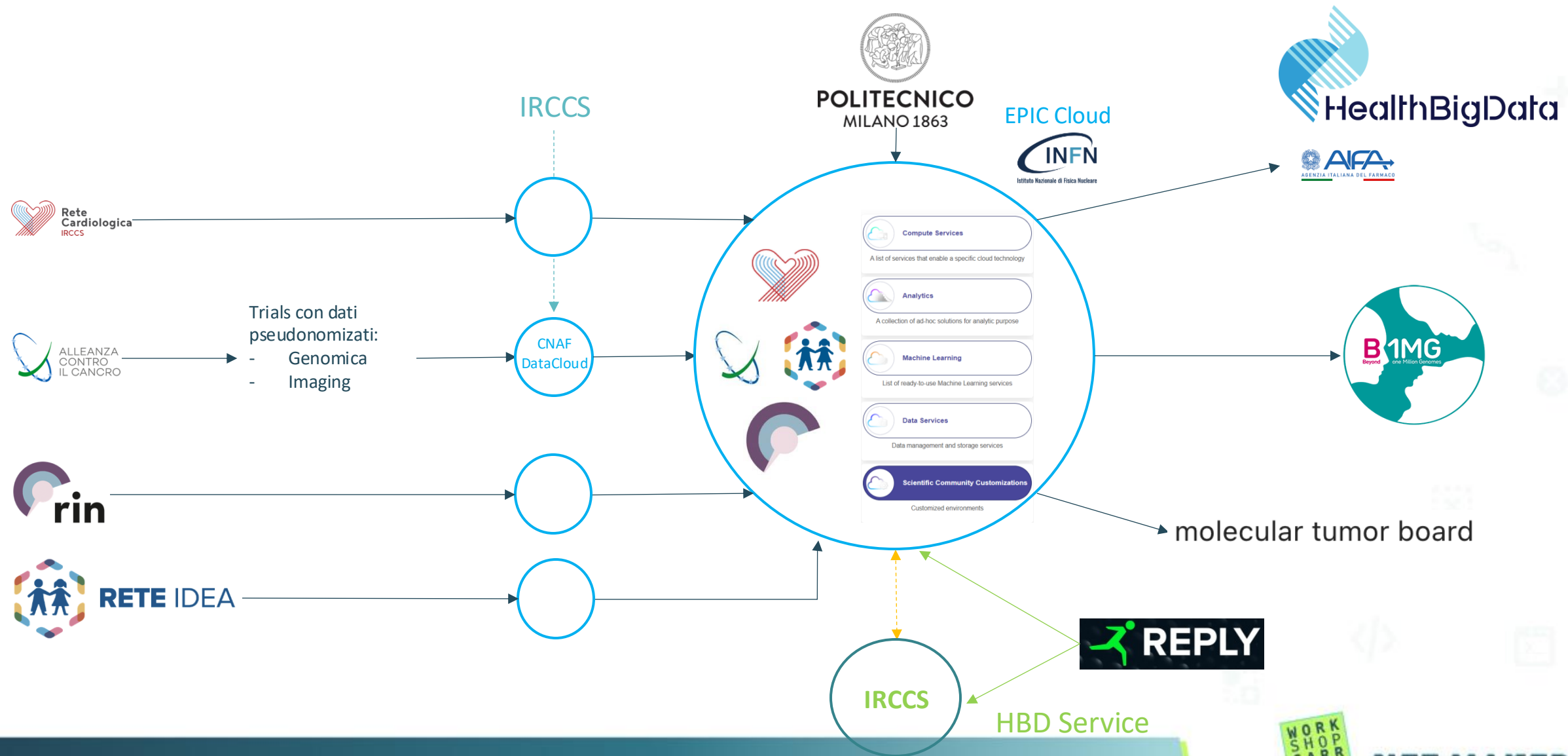
Classification	Attributes	Action				
Metadata_Quality_V02	<table border="1"><thead><tr><th>Name</th><th>Value</th></tr></thead><tbody><tr><td>Quality Level</td><td>BUONA</td></tr></tbody></table>	Name	Value	Quality Level	BUONA	
Name	Value					
Quality Level	BUONA					

Properties Lineage Relationships Classifications Audits

Showing 1 - 1  Show Propagated Classifications

○ Current Entity → Lineage → Impact

No lineage data found





Ministero della Salute



ALLEANZA  
CONTRO  
IL CANCRO



HealthBigData



POLITECNICO  
MILANO 1863



IEO  
Istituto Europeo di Oncologia

- Il numero degli IRCCS in Italia è in **continua crescita**: dal 2000 al 2020 il numero di Istituti è passato da 35 a 52.
- Degli attuali 52 IRCCS:
  - 22 sono di **diritto pubblico**
  - 30 sono di **diritto privato**
- **Reti strutturate tra gli IRCCS** per mettere a fattor comune le risorse (competenze scientifiche cliniche, sperimentali e biotecnologiche, tecnologie avanzate) e ottimizzare il loro impiego per rafforzare la posizione italiana in Europa

Nome della Rete	N° IRCCS coinvolti
Alleanza Contro il Cancro - ACC	29
Rete delle Neuroscienze e della Riabilitazione - RIN	30
Rete Cardiologica	20
Rete Italiana salute Dell'Età evolutiva – IDEA	10
Associazione Rete Aging	13
Rete Apparato Muscolo Scheletrico – RAMS	7

Numero di IRCCS per Regione, 2021



Aree di expertise:

- Cardiologia
- Dermatologia
- Diagnostica Immagini
- Farmacologia
- Gastroenterologia
- Genetica
- Geriatria
- Malattie Infettive
- Medicina della complessità
- Neurologia
- Neuroriabilitazione
- Oculistica
- Oncologia
- Ortopedia
- Pediatria
- Psichiatria
- Riabilitazione

WORK  
SHOP  
GARR  
2024

NET  
MAKERS

[arnaud.ceol@ieo.it](mailto:arnaud.ceol@ieo.it)





**KEYCLOAK**

Acc

Configure

- Realm Settings
- Clients
- Client Scopes
- Roles
- Identity Providers
- User Federation
- Authentication

Manage

- Groups**
- Users
- Sessions
- Events
- Import
- Export

User Groups

Groups Default Groups @

571 208

222 104

140 359

364 218

Search...

View all groups

- acclims-radiomics-unic-14\_osr-readwrite
- acclims-radiomics-unic-14\_osr-write
- acclplatformadmin
- admins
- onedata-genomics-admin
- onedata-genomics-gers-01\_ico-read
- onedata-genomics-gers-01\_ico-readwrite
- onedata-genomics-gers-04\_cro-readwrite
- onedata-genomics-gers-05\_fpo-readwrite
- onedata-genomics-gers-07\_h-readwrite
- onedata-genomics-gers-13\_ifo-readwrite
- onedata-genomics-gers-15\_aouim-readwrite
- onedata-genomics-gers-17\_igpi-read
- onedata-genomics-gers-17\_igpi-readwrite
- onedata-genomics-gers-17\_igpi-write
- onedata-genomics-gers-26\_ifo2-readwrite
- onedata-genomics-gers-33\_scdc-readwrite
- onedata-genomics-gers-34\_sdc-readwrite
- onedata-genomics-gers-all-read

## Autentication(MFA) & authorization



## Storing & Sharing

genomics-lung-10\_jov

Overview

Data

Shares

Transfers

Providers

Members

Harvesters

genomics-lung-13\_ifo1

genomics-lung-14\_osr

DATA

View provided by ACC-provider

- genomics-lung-10\_jov
  - lung1000CX-1TAR110\_v1\_Non-Filtered\_2019-12-05\_14\_27\_38.vcf
  - lung1000CX-1TB0210.bam
  - lung1000CX-1B0210\_v1\_Non-Filtered\_2019-12-05\_14\_50\_11.vcf
  - lung1000CX-2NB0110.bam
  - lung1002H-2TAR110.bam
  - lung1002H-2TAR110\_v2\_Non-Filtered\_2019-12-05\_15\_45\_45.vcf
  - lung1002H-2TB0210.bam

```

#####PIPE 2 (patients with SNV files)
process vcf_preprocess {
  errorStrategy = 'Ignore'
  pubInDir "${params.outDir}/01_normalization/${ircs}",
  input:
  set ircs, barcode, project, bam_tumor, bam_normal, hel_tu
  output:
  set ircs, barcode, file("${snv_somatic_baseName}.normaliz
  45.7 KB
  
```

## Tracking & monitoring

Table with columns: ID, Name, Description, Status, etc.

Line graphs showing trends over time.

ACC Lung Report

Patient information

ACC Barcode: [Barcode]

Gender: [Gender]

CBO ID: [CBO ID]

Institute: [Institute]

Project: [Project]

Cellularity: 70 %

**LEVEL 1**

FDA-recognized biomarker predictive of response to an FDA approved drug in this indication

gene	mut	allele frequency	consequence	biological sig.	indication
EGFR	p.L858R	43 %	missense mutation SNV	Likely pathogenic	Adinab,Docosone

**LEVEL 2A**

Standard case biomarker predictive of response to an FDA approved drug in this indication.\*

No LEVEL 2A mutations found for this sample.

**LEVEL 2B**

Standard case biomarker predictive of response to an FDA approved drug in another indication.

No LEVEL 2B mutations found for this sample.

**LEVEL 3A**

Compelling clinical evidence supports the biomarker as being predictive of response to a drug in this indication.

No LEVEL 3A mutations found for this sample.

**LEVEL 3B**

Compelling clinical evidence supports the biomarker as being predictive of response to a drug in this indication.

No LEVEL 3B mutations found for this sample.

**LEVEL 4**

Compelling biological evidence supports the biomarker as being predictive of response to a drug in this indication.

geneticallt pmpipmtp scapmms adhdhca epm pmpmpmtp m pmpit hmpmpmpm ep adhdhca m pmpit m

FEZET 4

50 FEZET 3B mmpmpmpm pmpm p m epm mmpmp-

geneticallt pmpipmtp scapmms adhdhca epm pmpmpmtp m pmpit hmpmpmpm ep adhdhca m pmpit m

FEZET 3B

## Automatized and centralized processing & reporting

# The widest italian oncology research network

Founded in 2002 by the Ministry of Health

L'Associazione è attualmente formata da **27 Istituti di Ricovero e Cura a Carattere Scientifico (IRCCS)** AIMaC, Italian Sarcoma Group, Fondazione CNAO, INFN, Fondazione Policlinico di Milano e Istituto Superiore di Sanità.

ACC promuove il trasferimento dell'innovazione scientifico-tecnologica nella pratica clinica per permettere di curare sempre meglio e in maniera sostenibile i pazienti oncologici.

<https://www.alleanzacontroilcancro.it>



<b>Pathological Anatomy and Biobanks</b> Cross-cutting discipline for cancer research activities	<b>Genomics</b> The country's computer network created to target cancer therapies	<b>Immunotherapy</b> Conversion of non-responsive patients into responsive with development of new combined therapies
<b>Breast</b> Seeking for predictive changes in response to biological drugs	<b>Oncohematology</b> Forever a frontier in genomic and/or personalized medicine	<b>Lung</b> The national cancer network fights the "big killer"
<b>Radiomics</b> Innovative research field and valuable opportunity for patient data collection	<b>Survivorship Care and Nutritional Support</b> Alongside recovered patients, chronic and long-term cancer survivors	<b>Musculoskeletal Tumors</b> Focus on patients not responding to first-line therapies
<b>Colon</b> IRCCS in the front line to shorten the time of translational research	<b>Glioblastoma</b> Liquid biopsy, a priority in the research agenda/programme	<b>Melanoma</b> Identify biomarkers to predict the least collateral rate therapy
<b>Sarcoma</b> Improve clinical management of patients with different sarcoma histotypes		



