INFN Cloud per progetti life science

Barbara Martelli

EPIC
Enhanced **PrI**vacy and **C**ompliance Cloud

Enhanced PrIvacy and Compliance Cloud is an ISO certified cloud platform	A region of INFN Cloud with a certified Information Security Management System
EPIC Cloud offers an IaaS Community Cloud for the communities of	Biomedical and genomic researchers Industrial researchers

Site locations: Bologna (active now), Bari and Catania sites will be added in June 2024 enabling for high availability and disaster recovery

Resource available today: about 700 TB of storage, 1440 cores, 10 TB RAM, 6 GPU A100

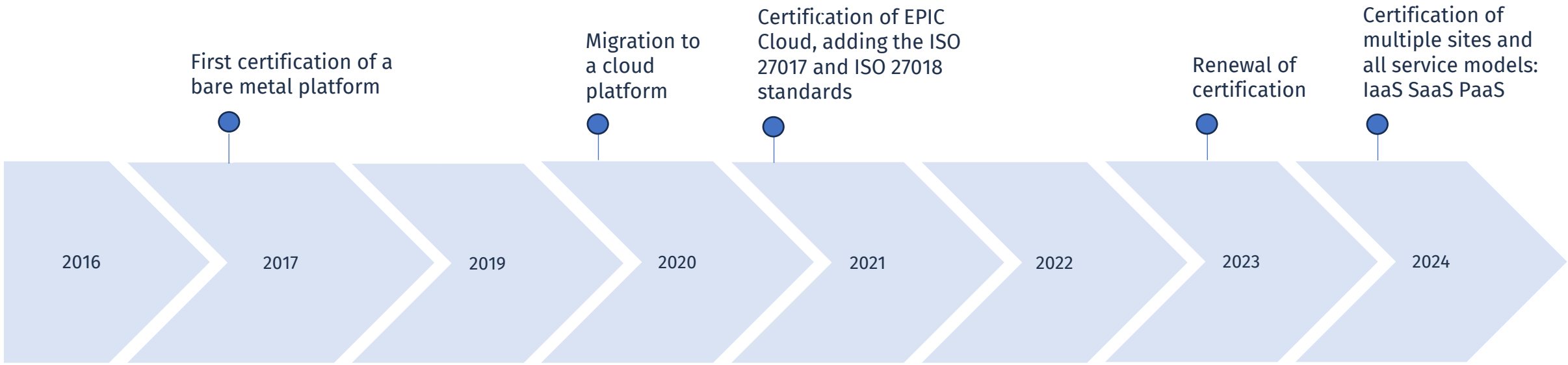
On going expansion with 3M euro of NRRP resources and 4M euro of funds from other projects

Progetti ospitati su EPIC

Progetto	HW (M€)	FTE
Harmony/Harmony+	Progetto in fase di conclusione, si discute della partecipazione alla fondazione	
DARE	3.2 (inseriti in gara TeRABIT)	4 TD CNAF, 1 TD Bari fino a dicembre ’26 (assunti)
ICSC Spoke8	0 (da richiedere a RAC)	1 TD CNAF, 2 TD Catania, 1 borsa Spoke0 dedicata ad attività di S8 (assunti)
HBD	6.5 (procedure d’acquisto da avviare, spalmati su 7 anni)	7/8 AR senior (1 assunto, 1 bandito e deserto, gli altri non ancora finanziati)
Sant’Orsola	0.34 (procedure d’acquisto da avviare, già disponibili)	2 AR senior (non assunti)
Sant’Orsola+	?	
IOR e AlmaHealthDB	?	3-4 ingegneri IOR-UniBO che lavorerebbero part-time al CNAF
Human Technopole	?	
Lafov PET	?	
UNCAN	?	
Numerosi IRCCS richiedono di utilizzare EPIC	?	
Intesa San Paolo (ICSC Spoke0) Smart City (ICSC Spoke9)	?	

Ulteriori dettagli sui progetti (aggiornato a Maggio ’23):
https://agenda.infn.it/event/34683/contributions/197354/attachments/105518/148360/20230523-datacloud-lifescience-v3.pdf

ISO/IEC 27001 27017 27018 certification



- In 2017 we started with a bare-metal that was ISO-27001 certified infrastructure
- In 2019 we moved to a Cloud-based infrastructure, and EPIC Cloud was born
- In 2021 we added the cloud certifications ISO-27017 and ISO-27018

EPIC Cloud Multisito

- il 26 ottobre scorso è stata firmata la disposizione del Presidente 26047, che istituisce lo steering committe di EPIC Cloud multisito.
- Lo steering è composto dai direttori delle quattro sedi coinvolte (Amministrazione Centrale, Bari, Catania e CNAF) ed ha il compito di allocare le risorse e definire gli obiettivi del Sistema Integrato di Gestione nazionale – quello che comunemente chiamiamo “certificazione ISO”.
  - Sempre invitati e informati il coordinatore di DataCloud, il presidente di C3SN ed il membro di Giunta Esecutiva con delega al calcolo
- Kick-off in Presidenza il 19 gennaio ’24. Presente tutto il management del calcolo INFN e i responsabili dei processi del Sistema di Gestione (process owner).

Lo scope del nuovo certificato

Coprogettazione, sviluppo e manutenzione di soluzioni software di DataCloud per il settore della ricerca.

Erogazione di servizi di DataCloud IaaS, SaaS e PaaS in community deployment model.

Obiettivo da conseguire entro giugno ‘24

Stato delle attività

- In fase di avvio la richiesta di offerte per l'audit di terza parte
  - Attualmente c'e' una bozza per 37 persone (~6 FTE), 1 sito e media complessità dell'SGSI spendiamo 6.5 kEuro per il primo anno e 2.6 kEuro per i due anni successivi
- Impostate le attività dei process owner
  - Definizione delle policy da scrivere e approvare (link preliminare qui)
  - Definizione delle attivita' di ciascun processo e del gruppo di lavoro (link alle bozze delle schede processo qui)
- Bozza del piano di attivita' dettagliato preparata dal consulente (a breve sara' discussa nello steering)

Processi e owner di processo (link)

Tipologia	Codice	Processo	Owner	WP
Management processes	MP 01	Information security governance/management interface process	Martelli	7
Core Processes	CP01	Security Policy management process	Bovo	
	CP02	Requirements management process	Donvito	3
	CP03	Information security risk assessment process	Carbone	4
	CP04	Information security risk treatment process	Ciaschini	4
	CP05	Security Implementation management process	V. Spinoso	4
	CP06	Process to control outsourced services	Monforte	
	CP07	Process to assure necessary awareness and competence	Costantini	2
	CP08	Information security incident management process	Peco	4
	CP09	Information security change management process	Stalio	1
	CP10	Internal audit	Belluomo	4
	CP11	Performance evaluation	Vistoli	7
	CP12	Information security improvement process	Lanzi	4
	CP13	Erogazione servizi DataCloud	Diego Michelotto	1
	CP14	Coprogettazione e sviluppo soluzioni software nell'ambito dei servizi DataCloud	Antonacci	5
	CP15	Manutenzione del deployment delle soluzioni sviluppate	Sergi	1
Support Processes	SP01	Records control process	Magenta	7
	SP02	Resource management process	Cesini	3
	SP03	Communication process	Rotondo	2
	SP04	Information-security customer relationship management process	Pellegrino	2

Project plan

Phase	Scope	Owner	Jul	Aug	Sep	23 Oct	Nov	Dec
			27 | 28 | 29 | 30	31 | 32 | 33 | 34| 35	36 | 37 | 38 | 39	40 | 41 | 42 | 42 | 43	44 | 45 | 46 | 47 | 48	49 | 50 | 51 | 52
Compliance	Assicurare il controllo e il soddisfacimento dei requisiti dello standard							
ISMS Organization	Definire l'assetto organizzativo del progetto: teams, ruoli, mansioni e deleghe. Per ciascun ruolo determinare le competenze minime richieste (conoscenze, abilità ed esperienza).							
Project tools	Identificare i tools che saranno utilizzati per pianificare, programmare e gestire tutte le attività di progetto comprese le informazioni documentate.							
Doc. Inf. Mngt	Tenere sotto controllo le informazioni documentate in termini di corretta identificazione, approvazione, distribuzione e conservazione.							
Process Approach	Definire l'approccio per processi utilizzato nello sviluppo del ISMS compresa la loro mappatura, descrizione e monitoraggio.							
Risk Mngt	Definire il nuovo Framework di gestione dei rischi.							
SOA Migration	Ri-mappare gli attuali controlli ISO27002:13 su ISO27002:22 aggiungendo e/o eliminandone altri secondo necessità e attribuendo a ciascuno di essi il livello di maturità conosciuto.							
Risk Assessment	Individuare, quantificare e valutare i rischi in linea con il Risk Framework							
Improvement	Pianificare e attuare le 3 linee di difesa del ISMS con l'emissione di almeno 1 piano di trattamento							

Da spostare in avanti di 6 mesi

Stato delle attivita' dei progetti di calcolo per life science

DARE Digital Lifelong Prevention

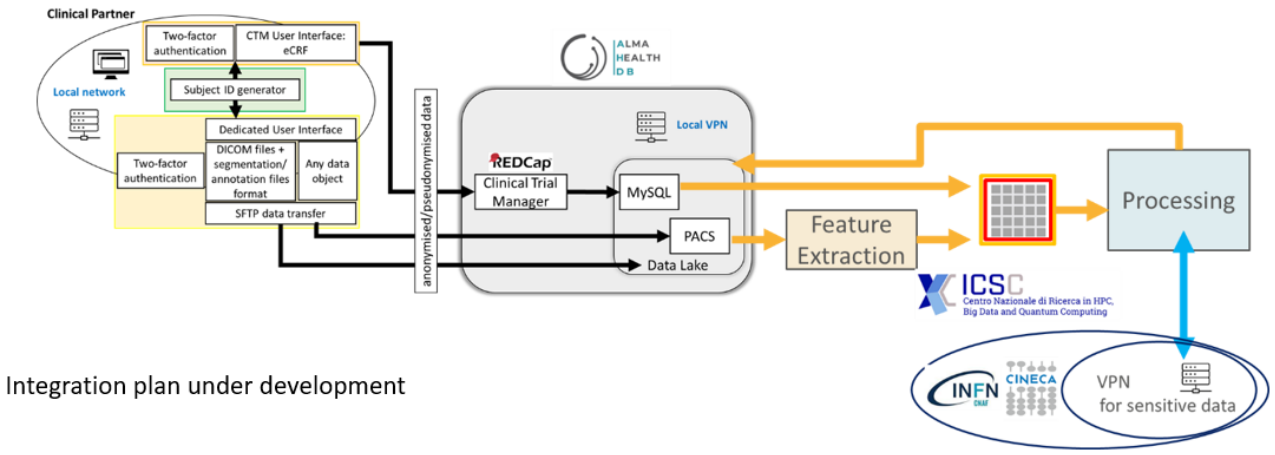
www.fondazionedare.it

- Nell'ambito dello Spoke1 "enabling factors and technologies for digital prevention" INFN partecipa attivamente alle attivita' di sviluppo della piattaforma tecnologica (basata su INFN Cloud EPIC -> ICSC) e coordina il WP3 "Interoperability governance".
  - Pubblicati 2 deliverable:
    - D3.1 Analysis of the Technological Framework and Interoperability - definition of the middleware stack
    - D4.1 Computing solutions deployment strategies
- HW in fase di acquisizione: ~2.3 Meuro (23% a Bari, il resto al CNAF)
  - Aggiunto alla gara quadro TeRABIT
  - Tempi previsti per la messa in produzione: fine 2024
  - ~2k CPU sia a Bari che al CNAF, ~10 GPU A100 al CNAF, ~4PB netti di storage al CNAFe ~1.5PB netti a Bari *(numeri preliminari, da confermare)*
- Richiesto a INFN di sopperire ai ritardi nella costruzione del datacenter di Palermo

Collaborazione con IOR per l'integrazione della loro piattaforma in ICSC

Ministero dell'Università e della Ricerca | Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA | PNC Piano nazionale per gli investimenti complementari al PNRR | DARE Digital Lifelong Prevention

Integration plan between AlmaHealthDB and ICSC architecture

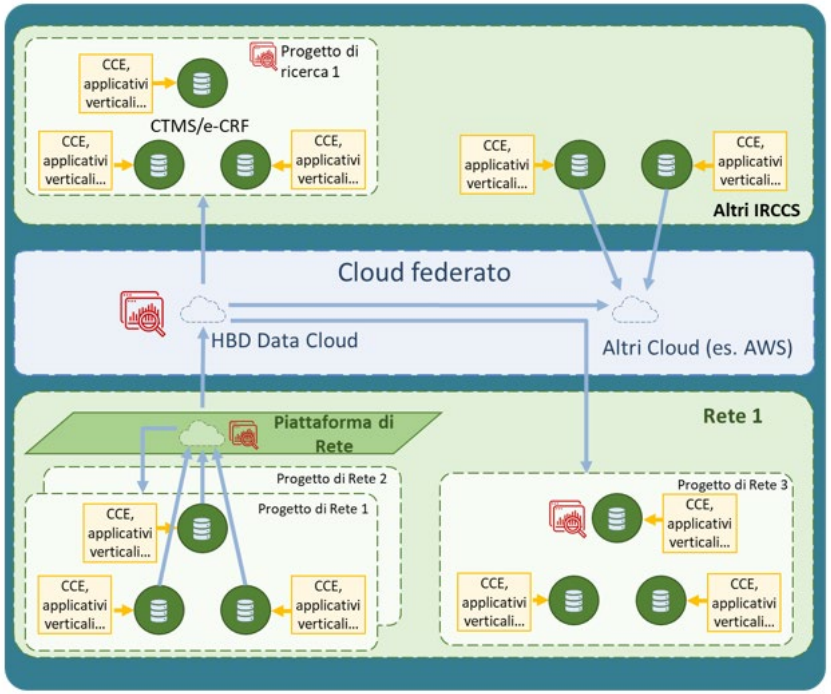


Integration plan under development

30/01/24 © 2024 Fondazione DARE 13

Rimodulazione Health Big Data

- Dopo interlocuzioni con il Ministero, decisa rimodulazione del Progetto
  - Trasformazione dei WG in WP
  - Creato il WP3 su piattaforma federate sotto responsabilita' INFN e IRCCS
  - Budget INFN aumentato (in fase di finalizzazione) sia per HW che per personale
  - Personale IRCCS dedicato alla piattaforma federata: si occupera' di
    - estrazione e armonizzazione dati,
    - integrazione piattaforme IRCCS con HBD DataCloud (-> INFN Cloud EPIC -> ICSC)



Collaborazione con UNCAN

www.uncan.eu

Federazione europea di datalake per la ricerca oncologica.

HBD DataCloud (-> INFN Cloud EPIC -> ICSC) sarà il nodo Italiano.

Le applicazioni ospitate sul nodo saranno quelle dell'ecosistema Elixir.

ICSC
Centro Nazionale di Ricerca in HPC
Big Data and Quantum Computing

Spoke 8
In-silico medicine and omics data

- Stato attivita':
  - Diversi progetti pilota ospitati su piattaforma EPIC.
  - In Corso il porting di Universal Immune System Simulator su piattaforma INFN Catania (che a giugno entrera' a far parte di EPIC).
- News dalla plenaria di Spoke 8 a dicembre '23
  - Cavalli ha proposto di costruire una PoC insieme a Humanitas per replicare il "Sistema Aosta" su piattaforma ICSC.
  - Importante coltivare la collaborazione con il CINECA per l'integrazione delle piattaforme certificate.
  - Pesole ha fatto un intervento sottolineando l'importanza di fare sinergia tra Elixir, ICSC, UNCAN lavorando anche sul FEGA.

Altri contatti in corso

- Human Technopole ha espresso interesse verso la piattaforma INFN Cloud EPIC. Svolte alcune call, si va verso la definizione di un accord di collaborazione di ricercar
- Progetto di ricerca con il gruppo LAFOV PET di Milano (C. Messa). EPIC coinvolta per la parte di gestione sicura dei dati.

Backup Slides

Technology

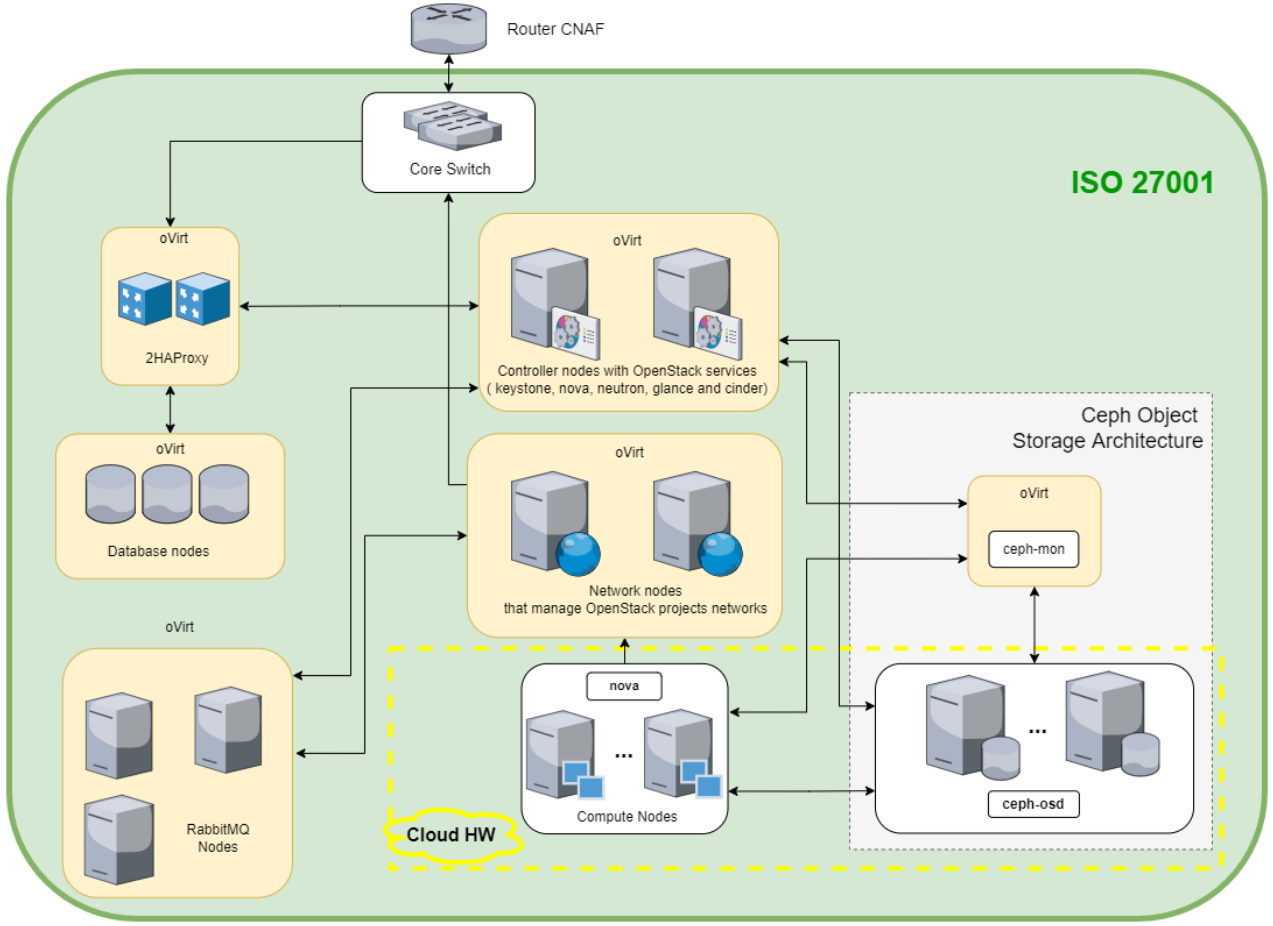
It is **based on the same technologies of INFN Cloud** (OpenStack, CEPH, IAM), with various enhancements introduced to meet higher security and privacy standards.

For example:

- The AuthN/Z system provides **2FA**, integration with web services, SSH and VPN (OpenVPN)
- **Network segregation** between OpenStack tenants is guarantee by **ACLs**
- At-rest and in-transit encryption
- Standard shared responsibility model:
  - User manages data, applications, runtime, middleware and OS
  - EPIC manages networking, storage, servers, virtualization
- Advanced logging and auditing services
  - **centralized syslog** server managed applying the ***segregation of duties*** principle

The EPIC Cloud infrastructure

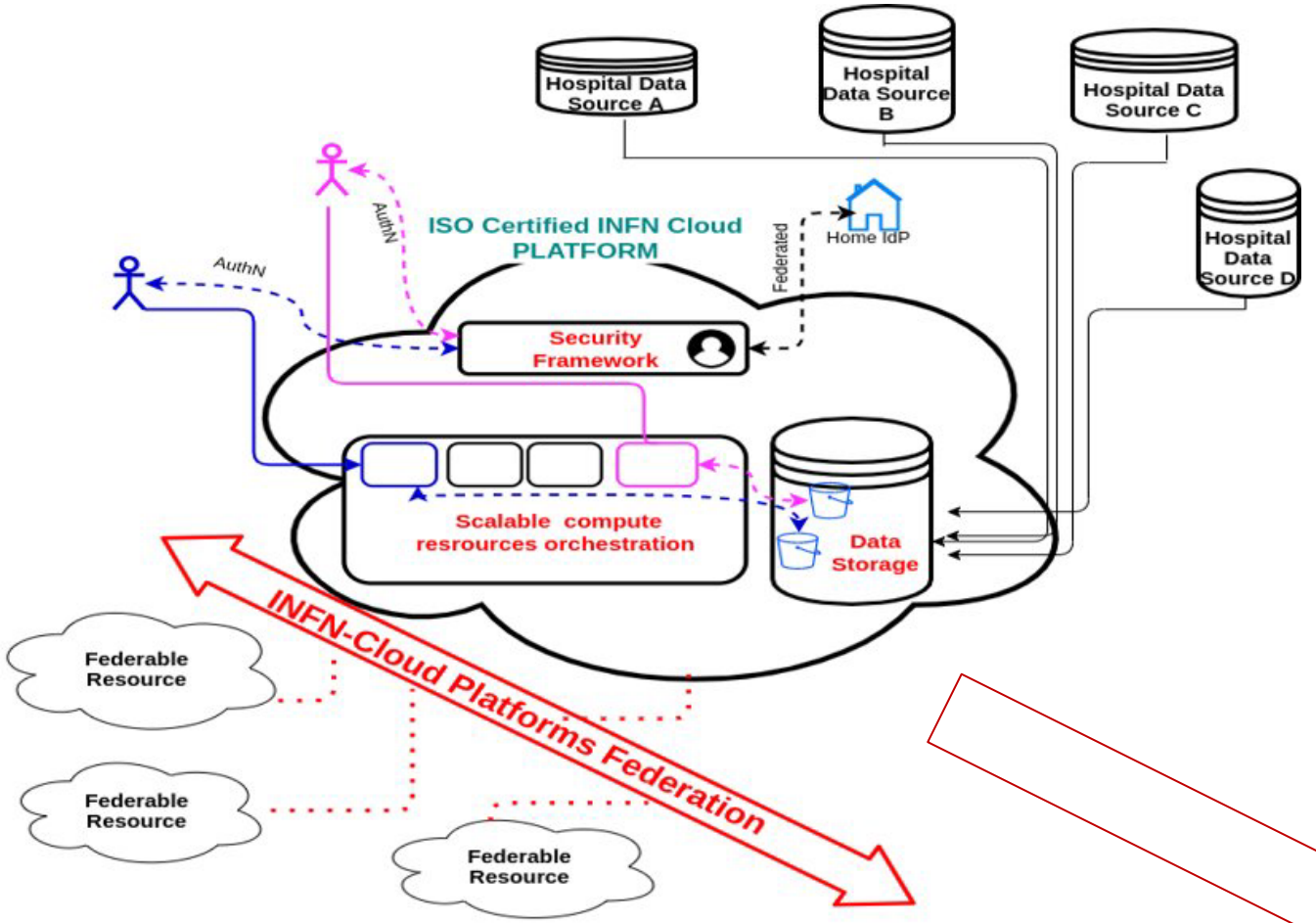
- All services in HA (Availability is one aspect of security)
- It currently hosts 5 Projects
- ~ 1440 ~10TB RAM, ~700TB Disk
- Tenant/domain segregation
- Physical security (defined perimeters, controlled access to racks, TVCC)
- Network isolation from Tier-1 resources, Next-Generation Firewall (NGFW) in place



OpenStack Infrastructure

EPIC Cloud enables the federated data lake for health-related projects

- In general, we receive heterogeneous requirements form life-science communities: some need central repository, others need to store data locally
- Possible scenarios:
  1. Central harvesting of data collected remotely
  2. Edge-level anonymization and central ingestion and analysis of data
  3. Edge-level Feature Extraction and central ingestion and analysis of features
  4. Federated learning (training at local sites and algorithm publishing)



https://www.physicamedica.com/article/S1120-1797(21)00320-3/fulltext

Operational and support requirement to offer sensitive data management services within INFN EPIC Cloud

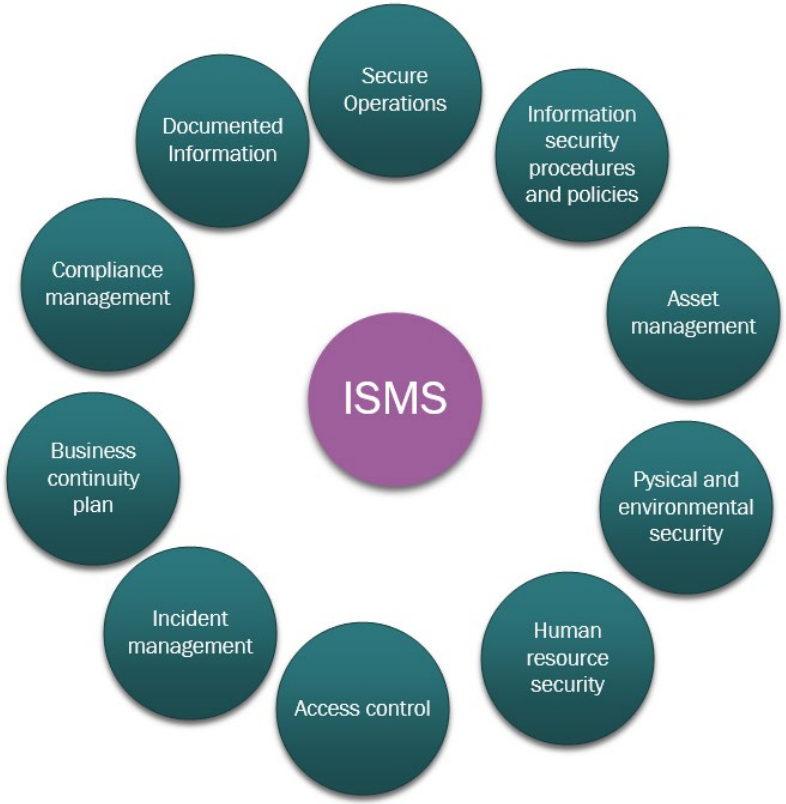
ISO/IEC 27001 27017 27018 certification

- This is a certification of the correct implementation of an **Information Security Management System** (**ISMS**)
- Information Security Management is about preserving the **Confidentiality, Integrity and Availability** (**CIA**) of information and associated facilities (systems, services, infrastructure or physical locations)
- It ensures **business continuity** by preventing and reducing the impact of security incidents
- Other properties can also be involved, such as **authenticity, accountability, non-repudiation, reliability** and **FAIRness**
- The **objectives** of the ISMS **are NOT fixed**, they depend on the context and are defined by the organization

ISMS: what's all about

Information Security Management System

- It is an **organizational framework** linking all the elements relevant to the information security, to assure that **policies**, **processes** and **security objectives** are implemented, communicated and assessed.
- It needs to **continually improve -> Deming Cycle**
- It is centered to the **risk assessment process** -> all decisions are based on the output of this process
- Goal: achieving the optimal **CIA balance**, i.e., ensuring Confidentiality of information, while still ensuring the information remains accessible to authorized persons and is not altered



Deming Cycle and Risk Assessment

Plan
(context, leadership, planning & support)

Do
(risk assessment & operations)

Check
(performance evaluation)

Establish the organization

- Define the scope
- Analyze the context and define the security requirements and objectives
- Define the Security Policy
- Determine roles and responsabilities
- Define the risk assessment approach

Risk Assessment

Control Selection

Control Implementation (Operations)

Monitoring KPI

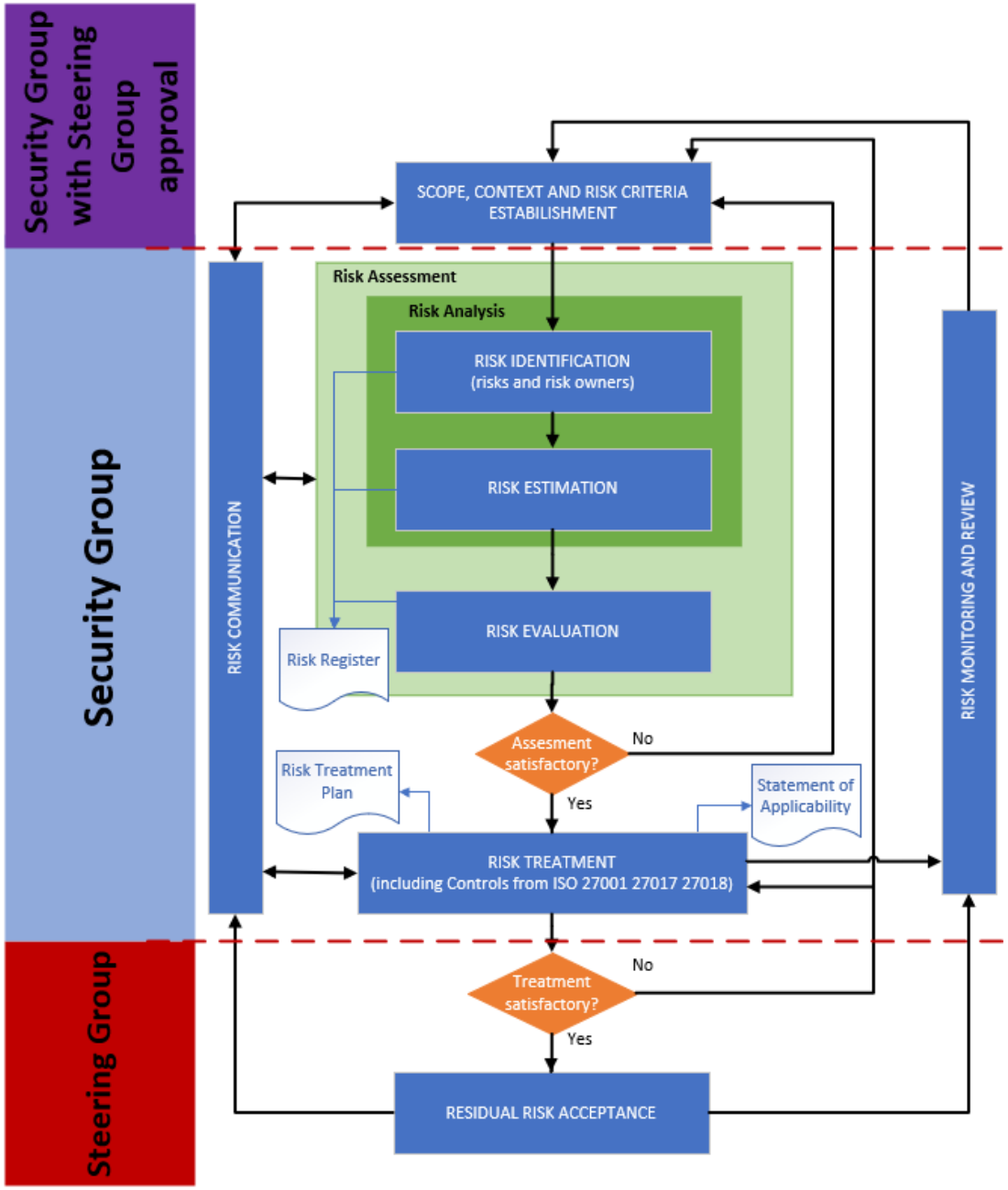
Internal Audit
External Audit

Effectiveness

Act
(continual improvement)

Risk Management Process in EPIC

- Iterative process aimed at supporting the decision-making process
- In EPIC is performed **once a month** and whenever a **relevant change** in the system **occurs**
  - ISO 27001 clause 8.2 requires to perform it *"at planned intervals or when significant changes are proposed or occur, taking account of the criteria established in 6.1.2 a). "*
- Established criteria to define Risk Owners



Security Incident Management

- Security incidents impacting CIA are managed following a documented security management process
- In case of high risk, the incident is escalated to INFN **CSIRT** (**Computer Security Incident Response Team**) https://csirt.infn.it
- The CSIRT team performs:
  - Incident identification
  - Incident categorization
  - Incident prioritization
  - Incident response
  - Incident closure
- A root cause analysis is also performed to prevent that the same incident occurs again

Required Document Information

8.1 Evidence Operational Planning and Control

A.5.1.1 Information Security Policy

9.2 Evidence of Audit Programme and Results

4.3 Scope

A.13.2.1 Information Transfer Policy and Procedures

A.10.1.2 Key Management Policy

9.3 Evidence of Results of Management Reviews

5.3 Organization Chart

6.1.3 Risk Treatment Process

A.11 Physical and Environmental Security Policy

A.12.2.1 Controls Against Malware Policy

A.6.2.1 Mobile Device Policy

6.2 Security Objectives

A.8.1.3 Acceptable Use of Asset Policy

8.3 Information Security Risk Treatment

A.12.3.1 Backup Policy and Procedures

ISO/IEC 27001

A.10.1.1 Cryptographic Control Policy

A.14.2.1 Secure Development Policy

A.11.2.9 Clean Desk Clean Screen Policy

A.15.1.1 Information Security Policy for Supplier Relationships

A.6.2.2 Teleworking Policy

A.12.6.2 Restrictions on Software Installation

A.14.1 Information Security Requirements Analysis and Specification

A.8.2.1 Information Classification Policy

7.2 Evidence of competence

A.16.1.4 Assessment of & Decision on Information Security Events

A.9.1.1 Access Control Policy

A.8 Asset Management Policy

A.17.2 Redundancies Policy

10.1 Actions Resulting from Nonconformities and Corrective Actions

6.1.2 Risk Assessment Process

8.2 Risk Assessment Results

A.9.3 Password Policy

9.1 Evidence of Monitoring and Measurement Results

Use case examples

HARMONY / HARMONY PLUS

Healthcare Alliance for Resourceful Medicines Offensive against Neoplasms in Hematology

Two IMI-2 European Projects. HARMONY PLUS, build up on the success of HARMONY, involves 39 partners and 8 Associated Partners from 10 countries.

- Use of big data analytics to accelerate blood cancer research
- Budget of 42.3M€ for HARMONY and 11.8M€ for HARMONY PLUS
- Over 100.000 patient datasets
- Harmonization of datasets
- Open and Standard interfaces

https://www.harmony-alliance.eu/

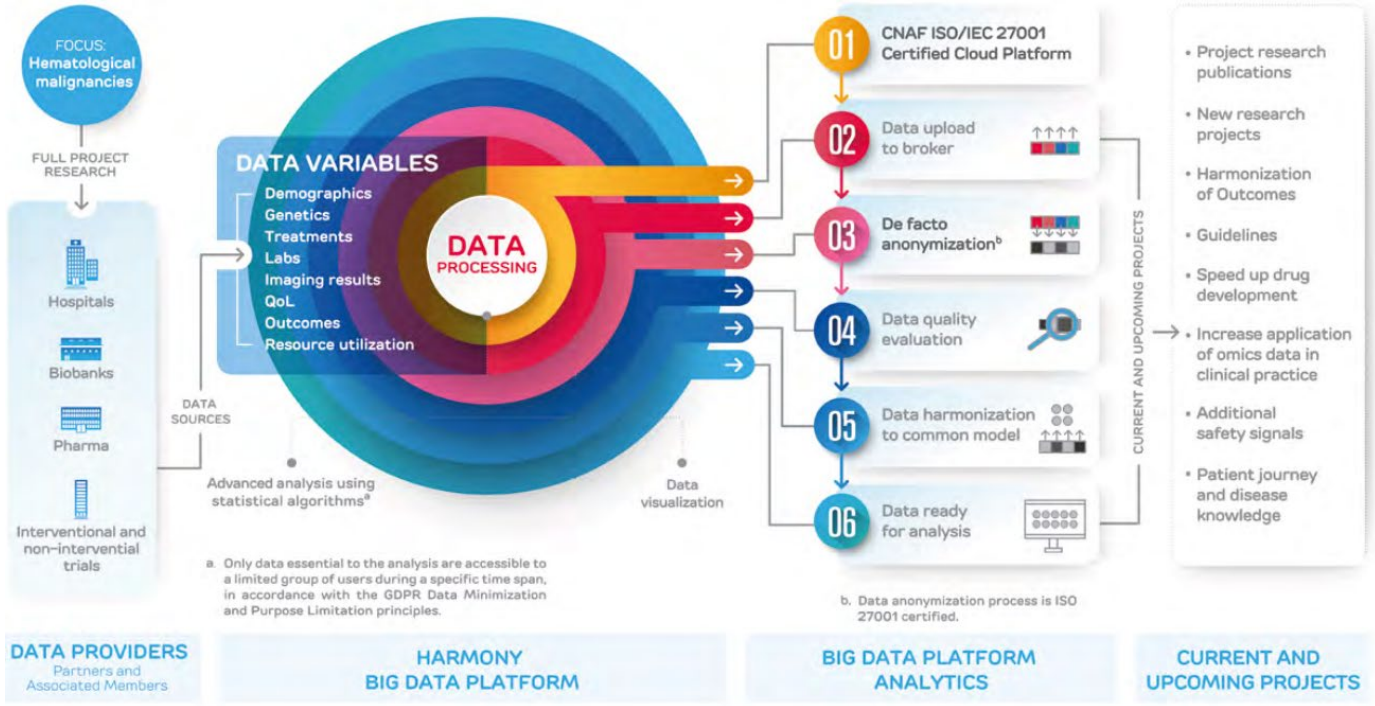


Figure 1 – Image based on an original idea by HARMONY (https://www.harmony-alliance.eu/).

HARMONY / HARMONY PLUS

Healthcare Alliance for Resourceful Medicines Offensive against Neoplasms in Hematology

- De facto anonymized data
  - GDPR does not apply, but the de-facto anonymization procedure assumes that the technology providers are ISO 27001 certified
- AgID measures (Standard and some of the Advanced)
- IaaS Service Model
  - HARMONY is the Data Controller
  - We are responsible for the infrastructure

https://www.harmony-alliance.eu/

Alleanza Contro il Cancro - ACC

The National Oncology Network founded in 2002 by the Ministry of Health, joined by 51 IRCCS, ISS, AIFA, INFN and Politecnico di Milano and several patients' associations to perform translational research in the field of cancer research.

- Genomic pseudonymized data
- GDPR applies
- AgID measures apply
- Italian Data Protection Authority rules apply

https://www.alleanzacontroilcancro.it/en/

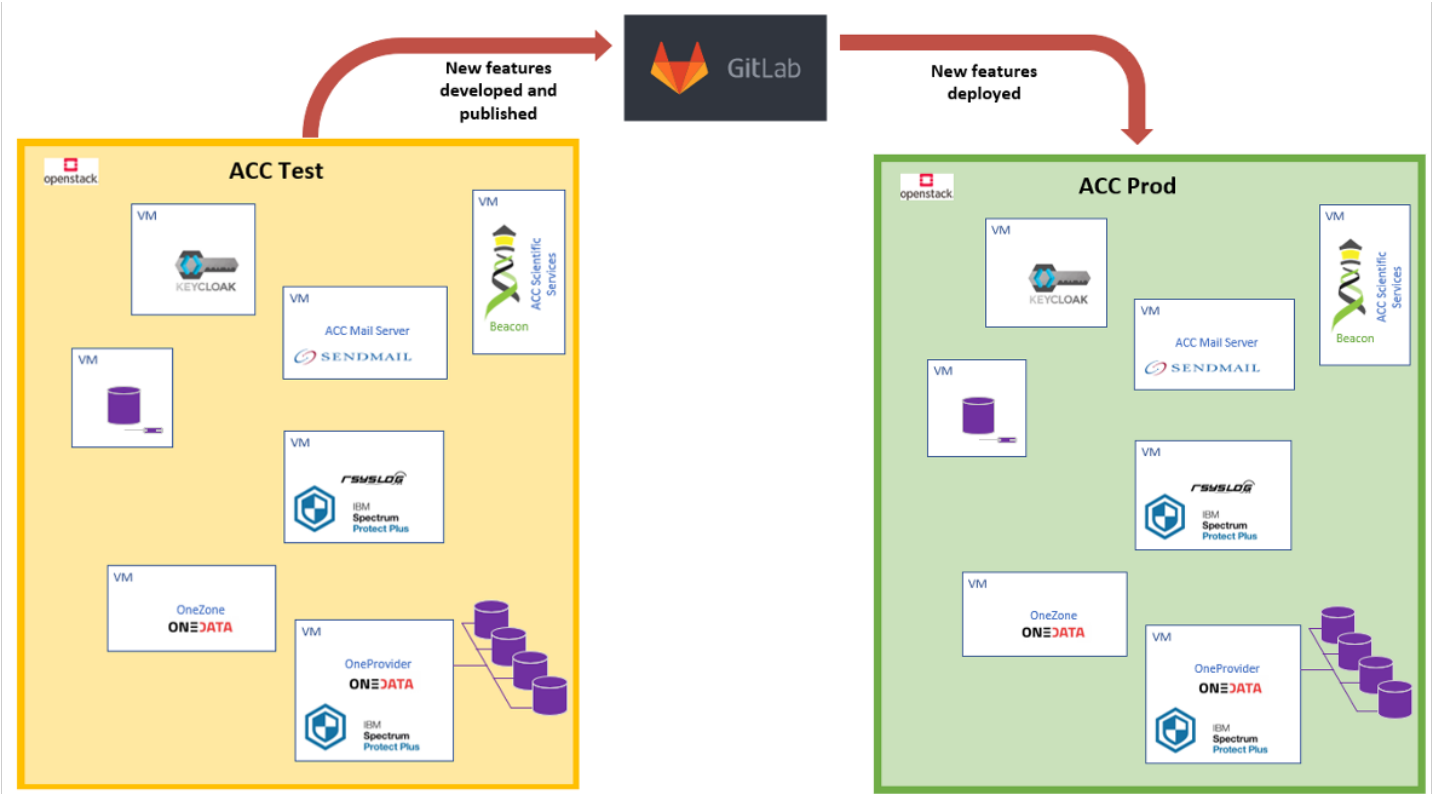
ACC - Services

Alleanza Contro il Cancro

Two separate OpenStack projects:

- **ACC-Test**: services have been configured and tested and every change in configurations has been validated
- **ACC**: where services have been configured and tested and every change in configurations has been validated

Each VM is hardened according to ISO 27001 OpenSCAP profile



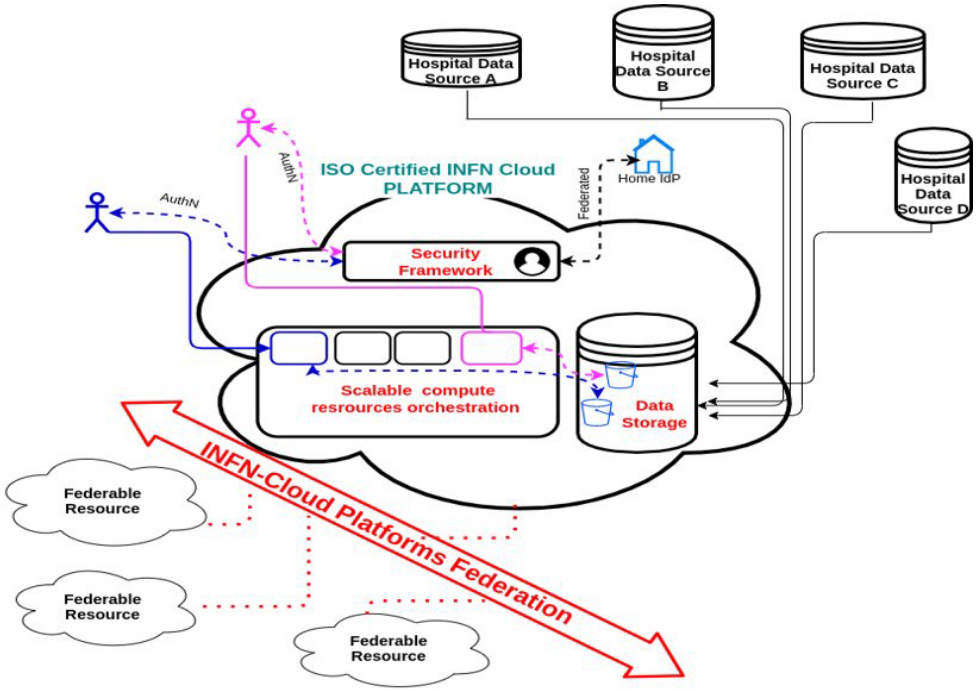
https://www.alleanzacontroilcancro.it/en/

PLANET

Pollution Lake ANalysis for Effective Therapy

An 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

- Data:
  - Pseudonymized clinical data (Covid-19 and Electronic Health Records from several hospitals)
  - Atmospheric data, population density, urban vs rural environment, mobility, socio-economic conditions
- Regulated by GDPR, Italian Data Protection Authority and ISS-INFN Convention
- Originally deployed on Cloud@CNAF



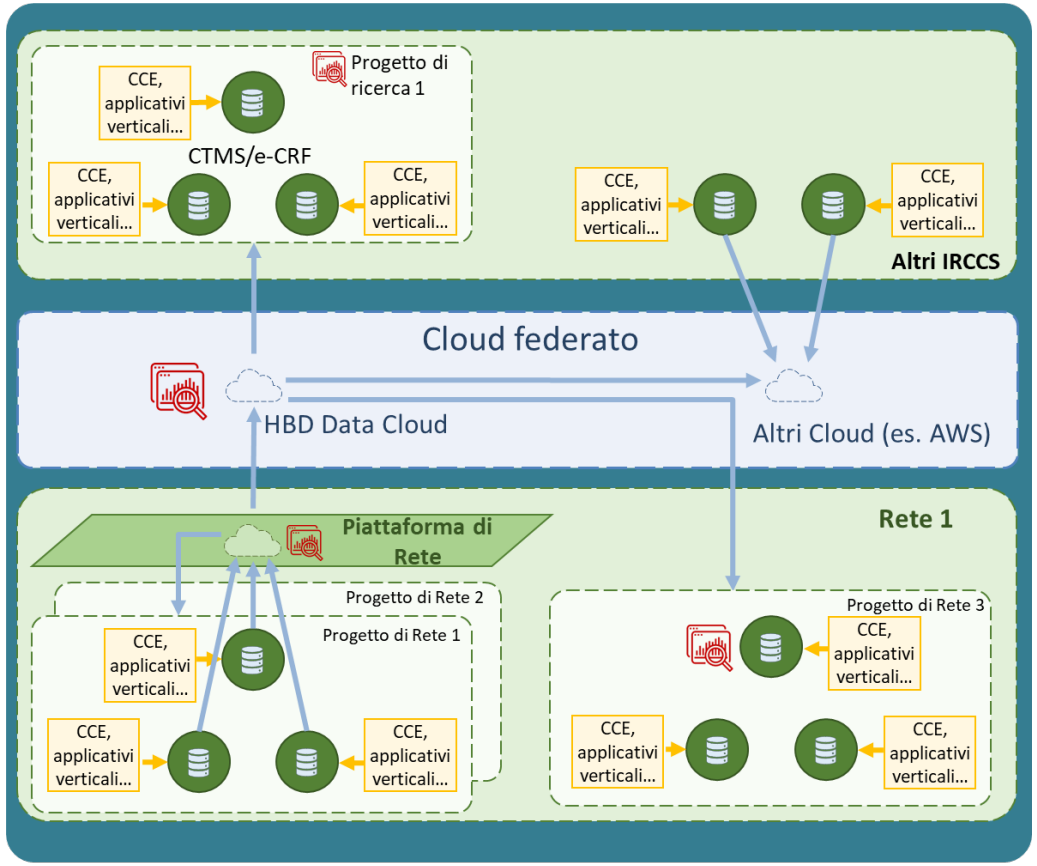
Our aim is to make this use case the first EPIC SaaS certified service (ambition/need to expand the scope of the ISO certificate to include SaaS services)

Health Big Data (HBD)

- Health Big Data is a 10-years project funded by the Italian Ministry of Health aiming at the creation of a federated and integrated big data platform for the health research at national level
  - 4 research networks: ACC, RIN, Cardio, IDEA
  - Research objectives: preventing diseases, personalizing treatments, improving the quality of life of patients
  - Budget: 55M€

Health Big Data (HBD)

- INFN is in the managing board of HBD. Its tasks include the definition of an integrated national platform and contributions to several Work Packages.
- The HBD architecture will provide solutions for several scenarios:
  1. Central harvesting of data collected remotely
  2. Edge anonymization, followed by central ingestion and analysis of data
  3. Edge feature extraction, followed by central ingestion and analysis of features
  4. Federated learning based on edge-based training, followed by publishing of the trained methods and by inference performed either centrally or at other edge locations



INFN-IRCCS Sant'Orsola Collaboration

Joint research agreement with the following objectives

- secure applications for genomic data
- GPU -based solutions for genomic analysis methods
- federated and integrated cloud platforms for homics data
- adaptation of genomic pipelines to cloud and data lake architectures based on microservices
- Integration of homics data and other clinical data like Electronic Medical Records (EMR)