



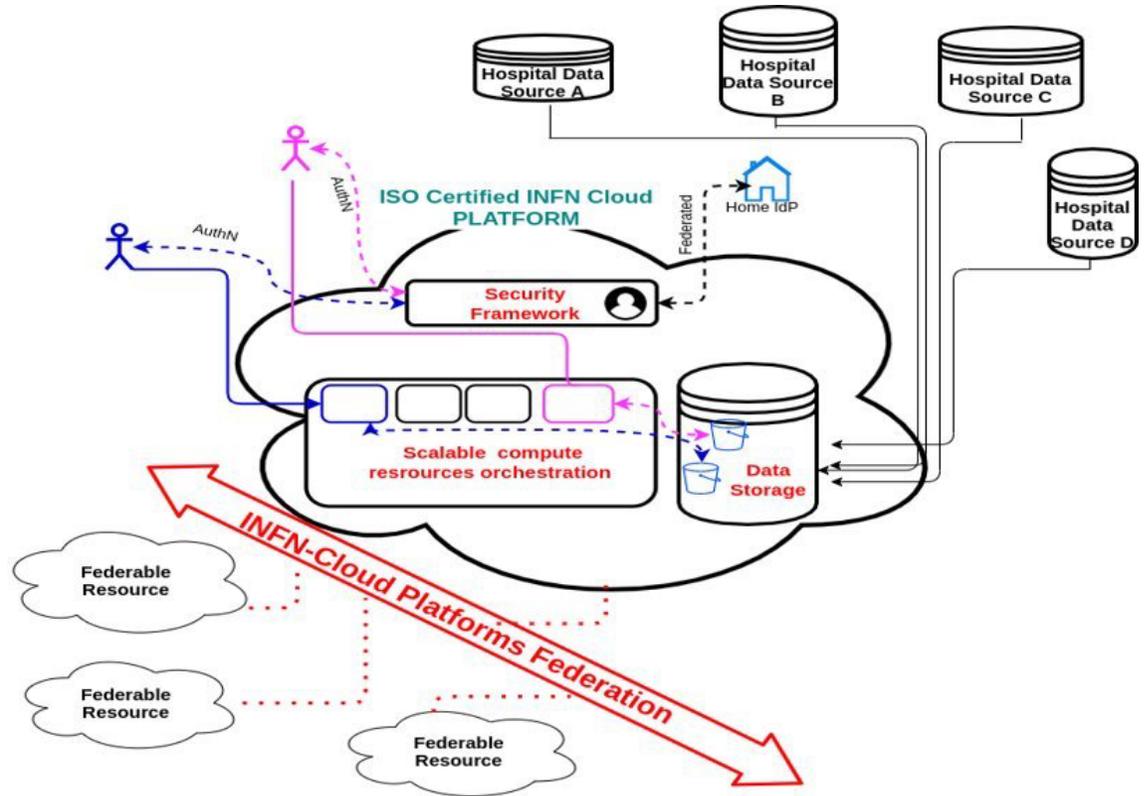
EPIC
Stato attività e
certificazione multi-sito

Elaborata la visione generale sul DataLake e servizi di EPIC Cloud Multi-sito



Servizi SaaS da rendere disponibili:

- Una piattaforma di genomica computazionale basata su BOSCO di IRCCS AOU di Bologna (Sant'Orsola)
- Una piattaforma Molecular Tumor Board sviluppata da Health Big Data
- Una piattaforma di radiomica da sviluppare in collaborazione con CSN5 (Next-AIM e nuova sigla VITA)
- Prototipi di HPC-Bubble certificate per utilizzare l'HPC via cloud
- Prototipi di piattaforme IoT per la raccolta di dati da sensori e dispositivi edge



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

Piattaforme integrate nella federazione INFN Cloud e ICSC mutuando le tecnologie di federazione INDIGO IAM, PaaS Orchestrator e RUCIO-FTS

Aspetti tecnici



Costituito **gruppo inter-progetto** (DARE/ICSC-Spoke8), inter-organizzazione (INFN, CINECA, UniBO) con obiettivo “**sviluppo della piattaforma certificata di DARE/ICSC**”

- Macro obiettivo: costruire la SaaS **AlmaHealthDB as a Service** su piattaforma cloud ICSC
 - si parte da quanto prodotto in Spoke0 e si aggiungono eventuali feature di sicurezza necessarie per gestire dati medici
- Primo task: deployment di RedCAP con autenticazione/autorizzazione gestita in keycloak

In prospettiva l’authN/Z EPIC basata su **Keycloak** dovrà **convergere con** quella sviluppata in Spoke0 basata su **IAM**

- Chiesto al gruppo di sviluppo **IAM** di definire una **roadmap**

In prospettiva le componenti della piattaforma (RedCAP, keycloak, tool per sviluppo di pipeline di analisi) potrebbero essere rese disponibili in PaaS

- Piano di sviluppo dell’**orchestratore in attesa di approvazione**

In prospettiva I dati potrebbero essere gestiti su datalake INFN (RUCIO/FTS)

- Chiesto a **WP6-DataCloud** di definire una **roadmap**

Plenaria EPIC del 30 Maggio <https://agenda.infn.it/event/41950/>

Non avendo certezza dello scope, si è lavorato nell'ipotesi di certificare interamente Bari e Catania e di certificare al CNAF unicamente i rack dedicati alla life-science (**3-10 rack**).

Su un totale di **25 processi** ne sono stati **pubblicati 23** e **discussi 12**

- necessario schedulare una call per discutere i processi rimanenti

Pronti il SoA preliminare e l'elenco dei rischi

- necessario schedulare due call per approvarli

Impostato il sistema documentale con workflow PowerAutomate su SharePoint

- necessario schedulare una call per presentarlo

Emerse molte domande per lo Steering EPIC che riflettono **incertezza di contesto**

Stime costi di certificazione

- Costo per estendere la certificazione **CNAF** ai siti di **Bari, Catania e AC**, includendo **IaaS PaaS e SaaS** in ambito DataCloud sarebbe di circa **13k Euro nel 2024** e **7k Euro/anno nel 2025 e 2026 ≈ 27k Euro**
 - E' possibile che **non** sia **necessaria l'ispezione di persona presso** i siti di **Bari e Catania**, ma sia sufficiente la visita ai siti del CNAF (Berti Pichat e Tecnopolo Manifattura)
- Costo per una libreria di 50 standard ISO fruibile da 50 colleghi INFN sarebbe di circa **7k Euro/anno**
 - spesa necessaria perché attualmente gli standard ISO sono disponibili solo a chi ha account UniBO

Risorse hardware disponibili

- **Storage:**
 - 1.5PB HDD, 600TB SSD
- **CPU**
 - 1440 core
- **Memoria**
 - 10 TB RAM
- **GPU**
- **6 schede Nvidia A100**
- **In attesa dell'hardware DARE acquistato con gara TeRABIT**
- **Attualmente abbiamo **terminato le GPU****
 - Varie soluzioni possibili per il partizionamento di GPU e l'allocazione dinamica delle partizioni di GPU attraverso container, batch system (da studiare), VM (da studiare)
 - **Questo sarebbe un tipico problema indirizzabile in DataCloud, partendo da un disegno architetturale basato sui requisiti espressi dalle comunita' mediche**

Problema su gestione GPU

- Le GPU non sono facilmente virtualizzabili
 - Necessaria una licenza specifica
 - Anche con la licenza, non siamo attualmente in grado di fornire questa caratteristica via cloud
- Ogni richiesta di GPU prevede **allocazione statica** della stessa su una specifica vm
 - Riassegnare la GPU richiede la distruzione della vm precedentemente creata
- Stiamo ipotizzando un approccio diverso basato su batch system per la gestione GPU, ma siamo ancora solo a livello di studio di fattibilità

Decisioni da prendere

Definizione del perimetro della certificazione:

- Interi siti o solo porzioni di essi (in particolare per il CNAF)
- Solo IaaS o anche PaaS e SaaS
- Tutto il processo di sviluppo software DataCloud oppure definiamo un processo di sviluppo esterno a DataCloud
- Includere nella certificazione IAM e Orchestratore oppure considerarle componenti esterne ad EPIC

Queste decisioni definiscono lo scope del certificato ISO e determinano il contenuto dei processi. Risultano quindi **bloccanti** per l'avanzamento dell'intero processo di certificazione.

Decisioni da prendere

- **Deadline per la certificazione**
 - considerando che dal momento dell'approvazione dello *scope*, sono **necessari almeno 6 mesi** di lavoro (Gantt del consulente e piano di lavoro presentato da Antonacci a novembre '23)
 - e considerando che il CNAF è attualmente impegnato nel trasloco del datacenter
- Decidere se certificarci solo ISO/IEC 27k, oppure includere anche ISO 9001
- Definire l'*owner* del processo «Resource Management»
- Decidere chi alloca le risorse e descrivere qual è il processo di allocazione:
 - **Steering EPIC**, **C3SN** ed i **responsabili dei fondi di progetti esterni** o una combinazione lineare di questi
- Definire quali risorse gestisce il processo *resource management*.
 - Se ci limitiamo alla 27001: solo le risorse per la security
 - Se utilizziamo anche la 9001: tutte le risorse

Criticità

- Necessaria una **comunicazione scritta delle decisioni prese**, la comunicazione dovrebbe arrivare direttamente dal management
- Necessario definire la data di certificazione, **ormai appare inverosimile certificarsi prima di febbraio '25**
- Necessaria un'azione di comunicazione ed engagement che **«sensibilizzi» sull'importanza delle attività**