Clinical rare cancer research situation

Health data constitutes **30%** of the world's data

Rare cancers
- **Incidence** rate < 6 in 100,000 but the majority (74%): **< 0.5** in 100,000
- **200 types** of rare cancers exist **~20-25%** of all cancer diagnoses
- In Europe, rare cancers have a **5-year OS rate** of **47%** vs 67% for common cancers

IDEA4RC Goal

To establish a **Data Space for rare cancers** that will make possible the secondary use of existing multisource health data (cancer registry data, national registries, data from biobanks etc.) across EU healthcare systems leveraging emerging interoperability technologies and AI approaches.

11 pilot sites from EURACAN ERN

Clinical centers

- Centre de Lutte Contre le Cancer Leon Berard
- Assistance Publique Hopitaux De Paris
- Instituto Investigacion Sanitaria Fundacion Jimenez Diaz
- Västra Götalandsregionen
- Panstwowy Instytut Badawczy - Maria Sklodowska-Curie
- Fakultni Nemocnice V Motole
- Oslo Universitetssykehus Hf
- Masarykuv Onkologicky Ustav
- Fundacion Profesor Novoa Santos
- Universitaetsklinikum Essen - Ukessen
- Fondazione IRCCS Istituto Nazionale dei Tumori

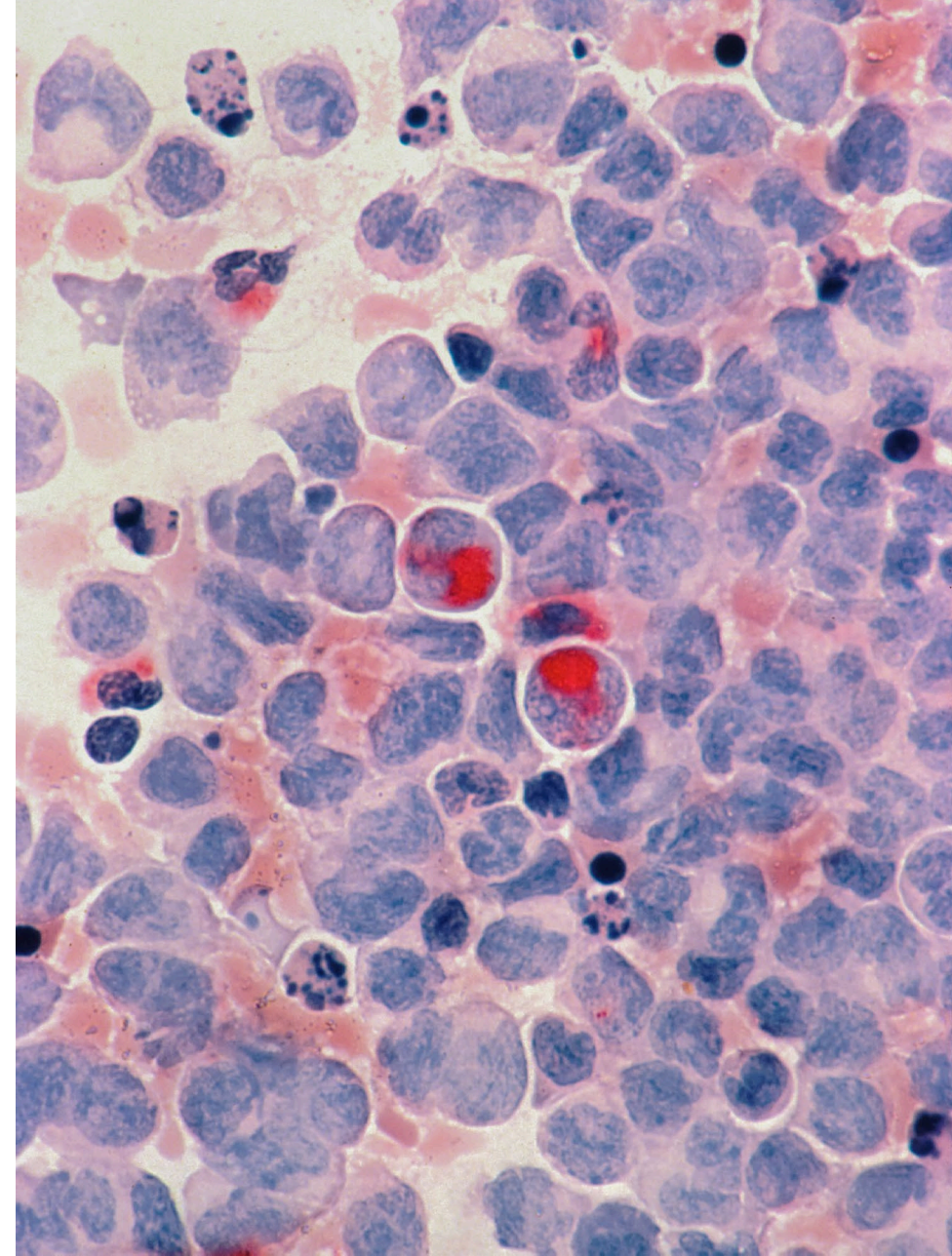**Sarcoma** and **Head and Neck Cancer** use cases

# Expected studies related to secondary use of data

RARE CANCER DATA ECOSYSTEM

- Advance research

- Increase knowledge

- Improve quality of care

- Ensure access to optimal treatment

- Underpin data economy towards the European Health Data Space

IDEA4RC ecosystem

The ecosystem will be designed in order to take into account the values of the different **stakeholders involved in the rare cancers care**. Social scientists will organize workshops to engage with different stakeholders and discuss with them their values and expectations.
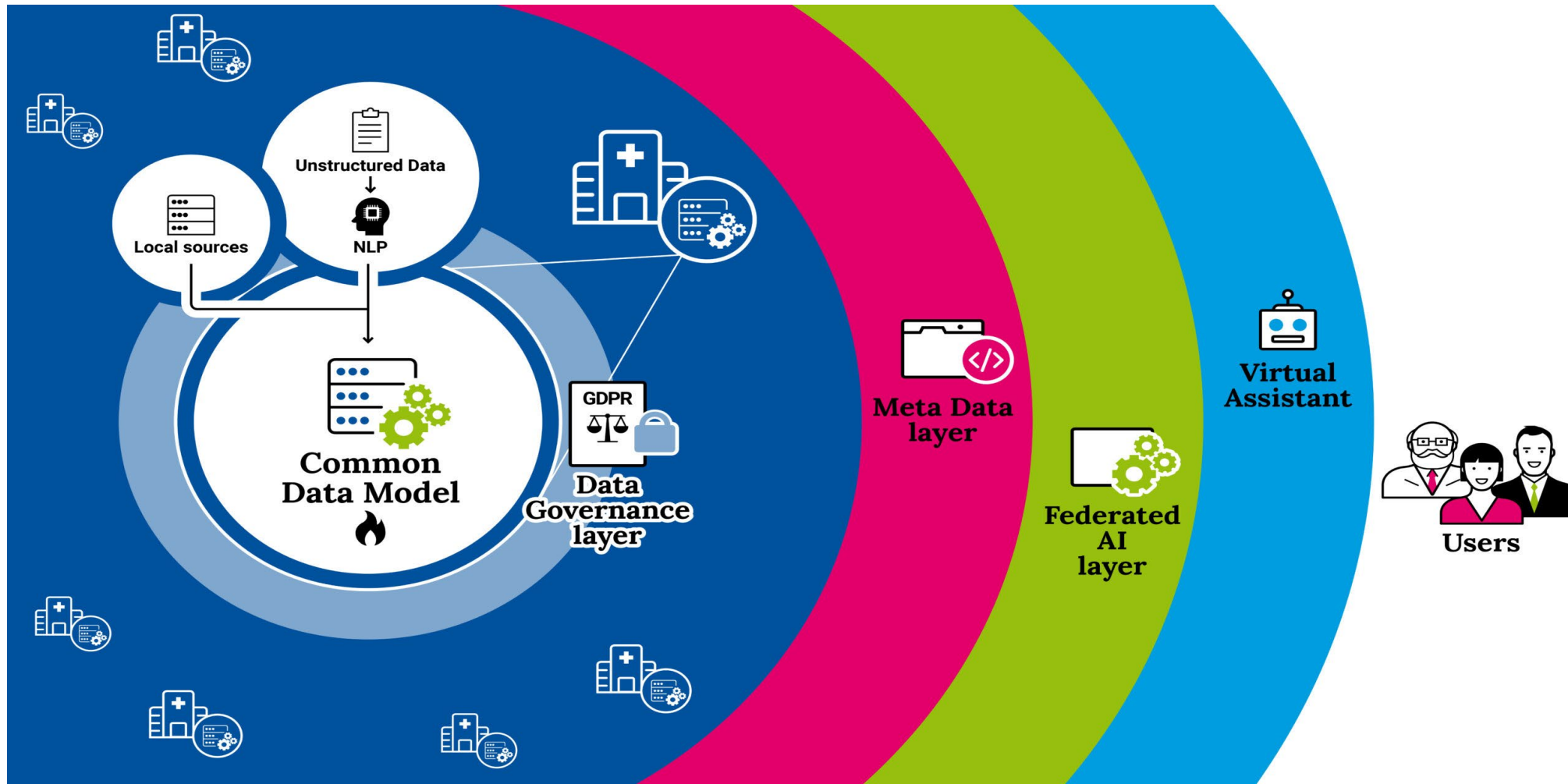
"OPEN DATA AS POSSIBLE AND CLOSED, WHEN NECESSARY,WITH LOCAL PROCESSING ALWAYS AVAILABLE"

Enable federated learning and analytics (IDEA4RC ORCHESTRATOR PLATFORM) to execute the expected data analysis in privacy preserving environments (IDEA4RC-CAPSULE).

ZERO-TRUST

OMOP/FHIR
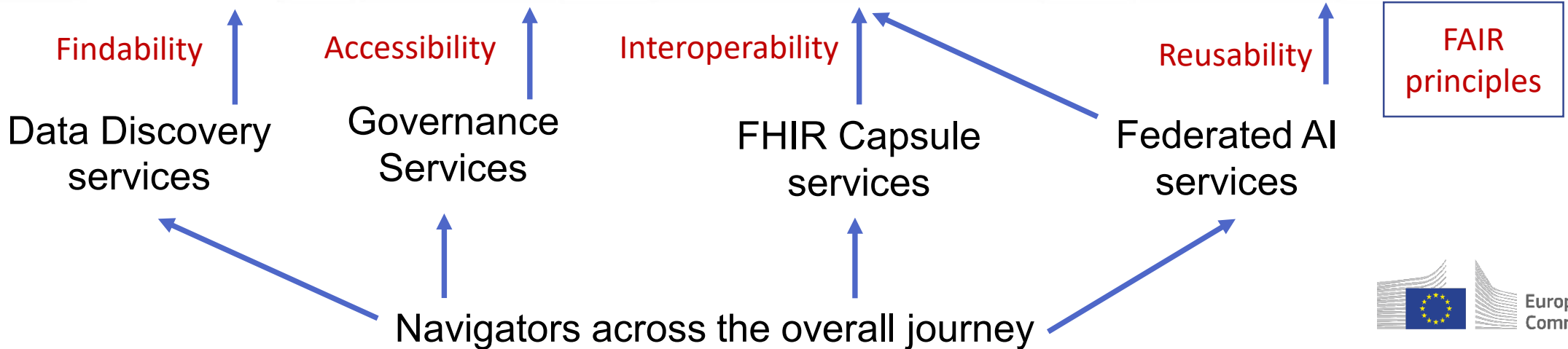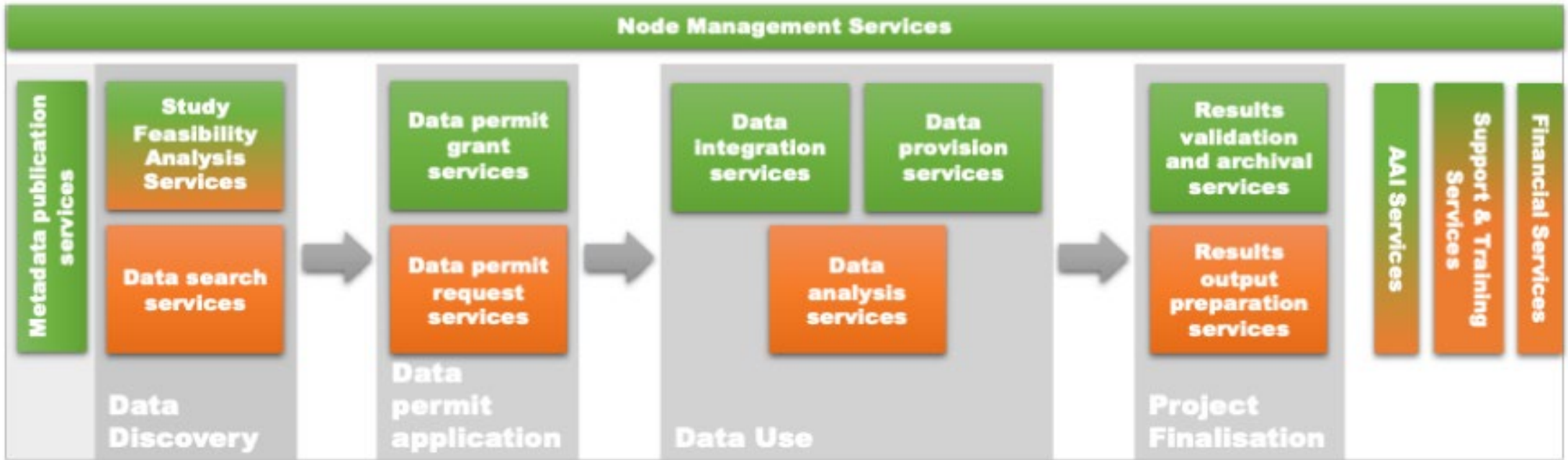
SERVICE MESHES

NLP

IDEA4RC and EHDS user journey support

■ Platform view (structural services capsule-to-platform harmonization) ■ User view (Virtual assistants specifications)

Findability — Data Discovery services

Accessibility — Governance Services

Interoperability — FHIR Capsule services

Reusability — Federated AI services

FAIR principles

Navigators across the overall journey

IDEA4RC - technical approach

IDEA4RC ORCHESTRATOR PLATFORM

Enhance use & reuse of data in different contexts

Suitable environment for local data processing

Moving **processing closer to the data**.

Federated learning & analytics

Decouple infrastructure from the system in use with **scalable deployment**

Pre-conditions for **high availability & computational capacity**

IDEA4RC CAPSULE PLATFORM

Zero Trust model to underpin data economy for the **EHDS**

Interoperability of data and services by using **HL7 FHIR and OMOP standard**

Enriching data with annotations (e.g., ontologies)

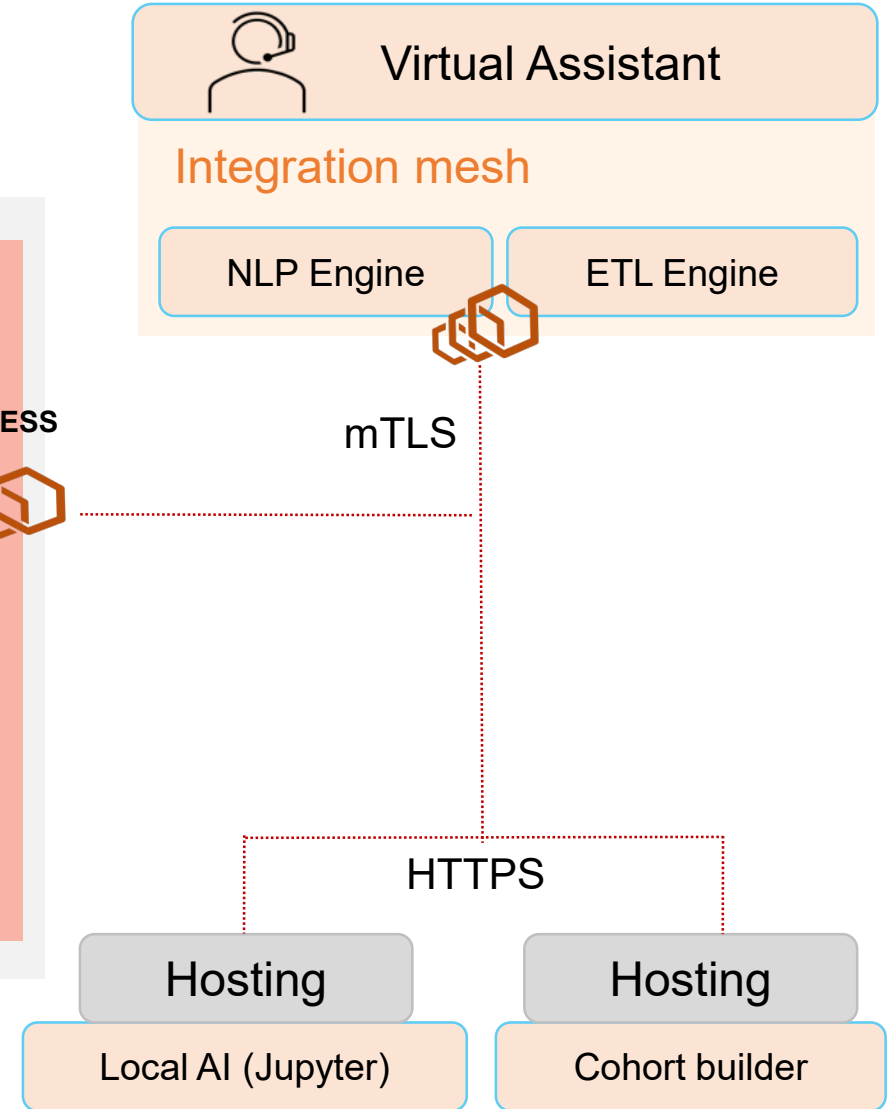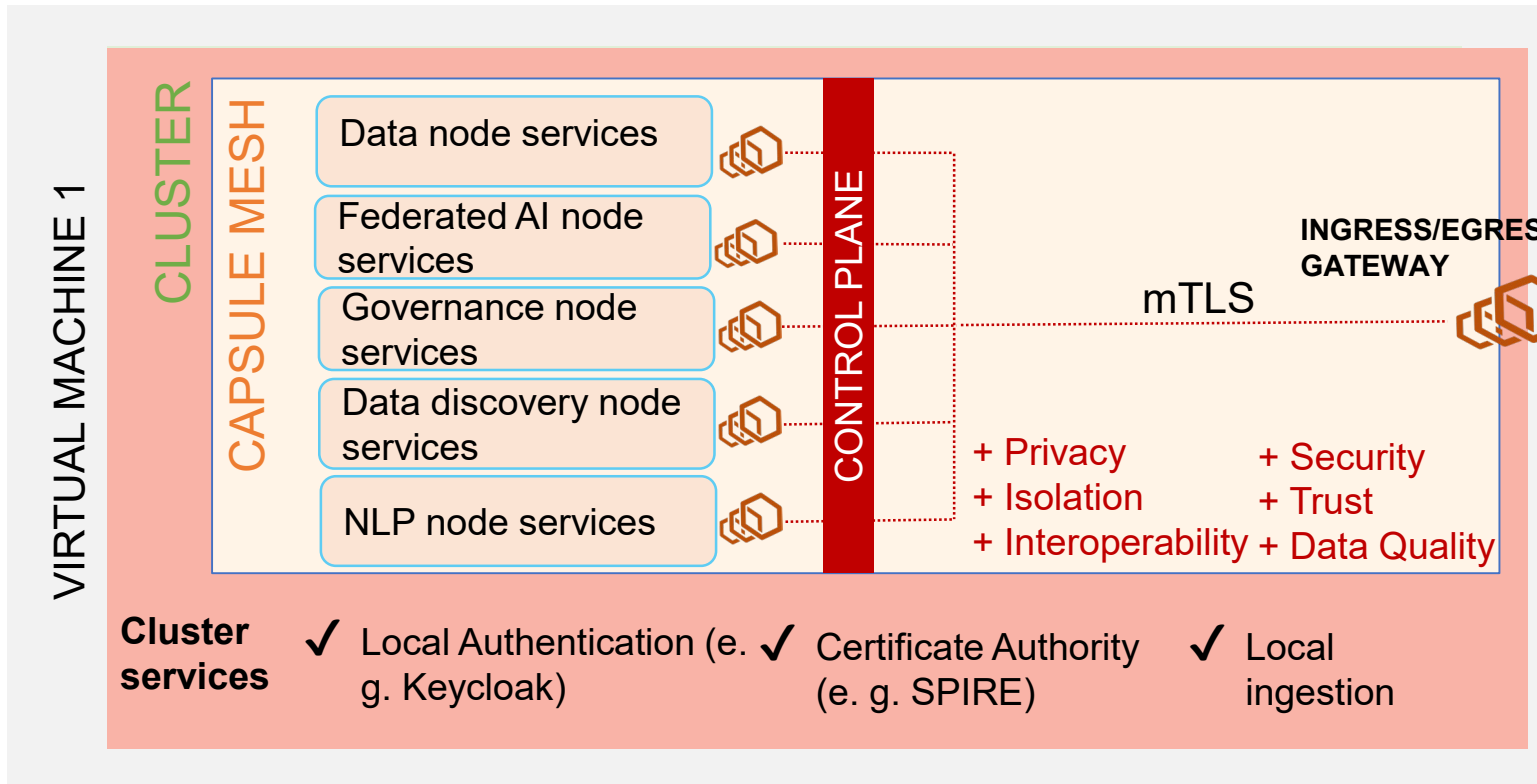Speed-up structuring of **unstructured data**

Support **federated learning** approaches

Trustiness and explainability of AI models

IDEA4RC CAPSULE ARCHITECTURE

Virtual Assistant

Integration mesh

NLP Engine | ETL Engine

mTLS

VIRTUAL MACHINE 1

CLUSTER

CAPSULE MESH

Data node services

Federated AI node services

Governance node services

Data discovery node services

NLP node services

CONTROL PLANE

INGRESS/EGRESS GATEWAY

mTLS

+ Privacy   + Security
+ Isolation   + Trust
+ Interoperability + Data Quality

Cluster services

✓ Local Authentication (e. g. Keycloak)   ✓ Certificate Authority (e. g. SPIRE)   ✓ Local ingestion

HTTPS

Hosting | Hosting

Local AI (Jupyter) | Cohort builder

IDEA4RC ORCHESTRATOR ARCHITECTURE
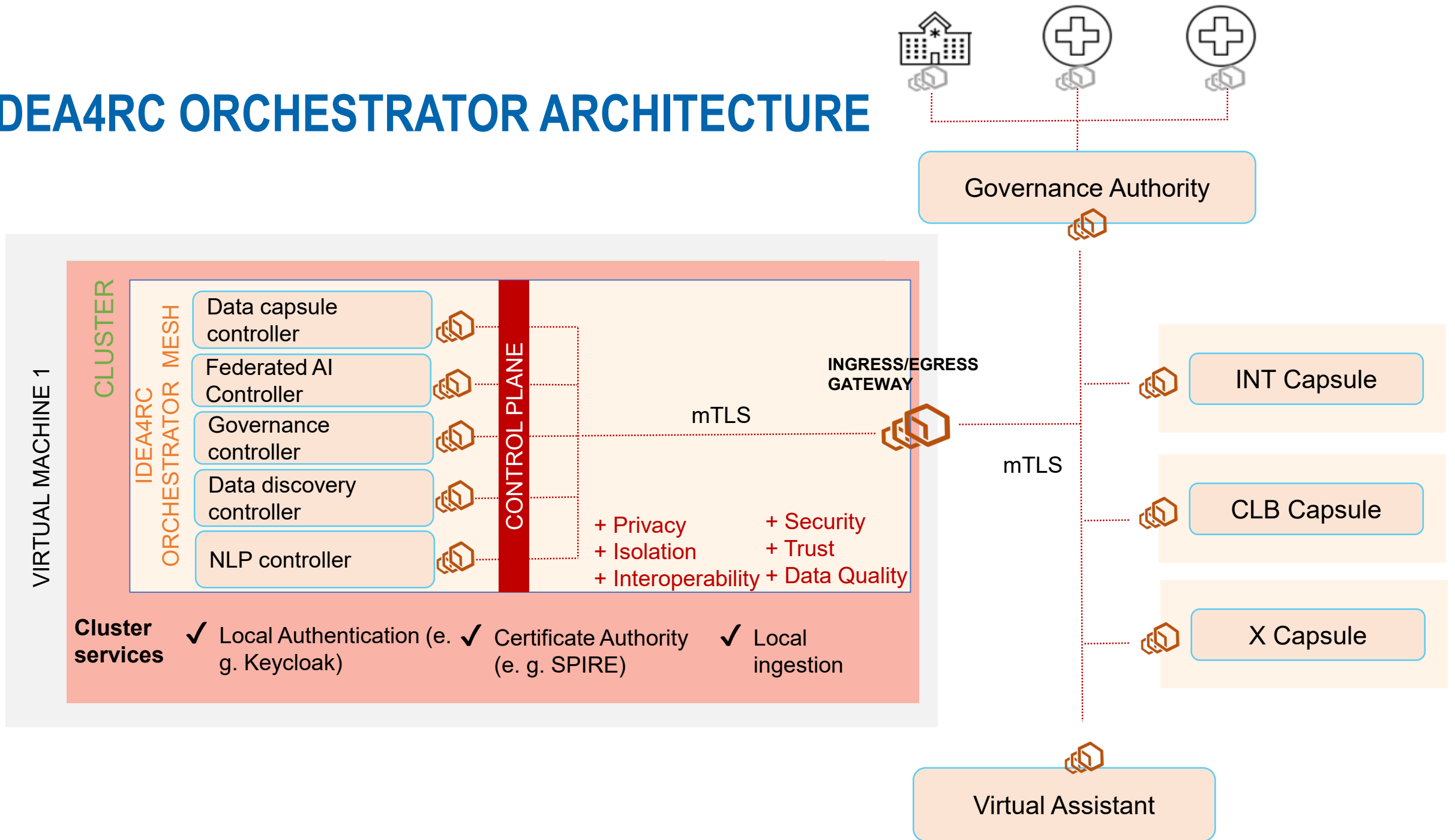
Data controller services

We have defined a common data model "**clinician friendly**" that hide the complexing of the OMOP and FHIR standards

We have optimized the data model to easy **query temporal evolution of cancer** (similar to OMOP episode model)

But

- we have mapped it into **OMOP oncology extension**
- we have mapped it to FHIR with an **IDEA4RC FHIR implementation guide**
- we have enabled **query the data** model in all **its possible representations**

Data discovery services

They are aligned with EHDS **DCAT-health profile specification**

But
- we have added the **codebook information**
- we have added the **terminologies used for every variable** of the codebook
- we have defined a set of quality checks that include:
    - **Plausibility**
    - **Conformance**
    - **Completeness**
    - **Validation of temporal evolution**

Data governance services

They are aligned with EHDS **data permit**

And

- enable to each center to evaluate documentation apported by researchers

- reduce the complexity of data permit submission to only required information

- allow personalization of process approval center by center

NLP services

They allow:

- extract **data model concept** from free text

- reconstruct **temporal evolution** of data

- help researchers to **convert textual query into SQL** query of IDEA4RC data model

- help researchers to understand platform functionalities with **AI chatbots**

Federated AI services

They allow privacy preserving analysis through federated computing

And

- enable federated analysis with OMOP, FHIR and Idea4rc data model

- align the federated analysis with the approved data permits

- cover all use cases defined by clinicians within the project

IDEA4RC – Zero Trust Model

Avoid cyber attacks against healthcare organizations in nowadays IT ubiquitous environments

Zero Trust

The network is distributed among many "castles." Authenticated users are granted granular, time-limited access to specific assets.
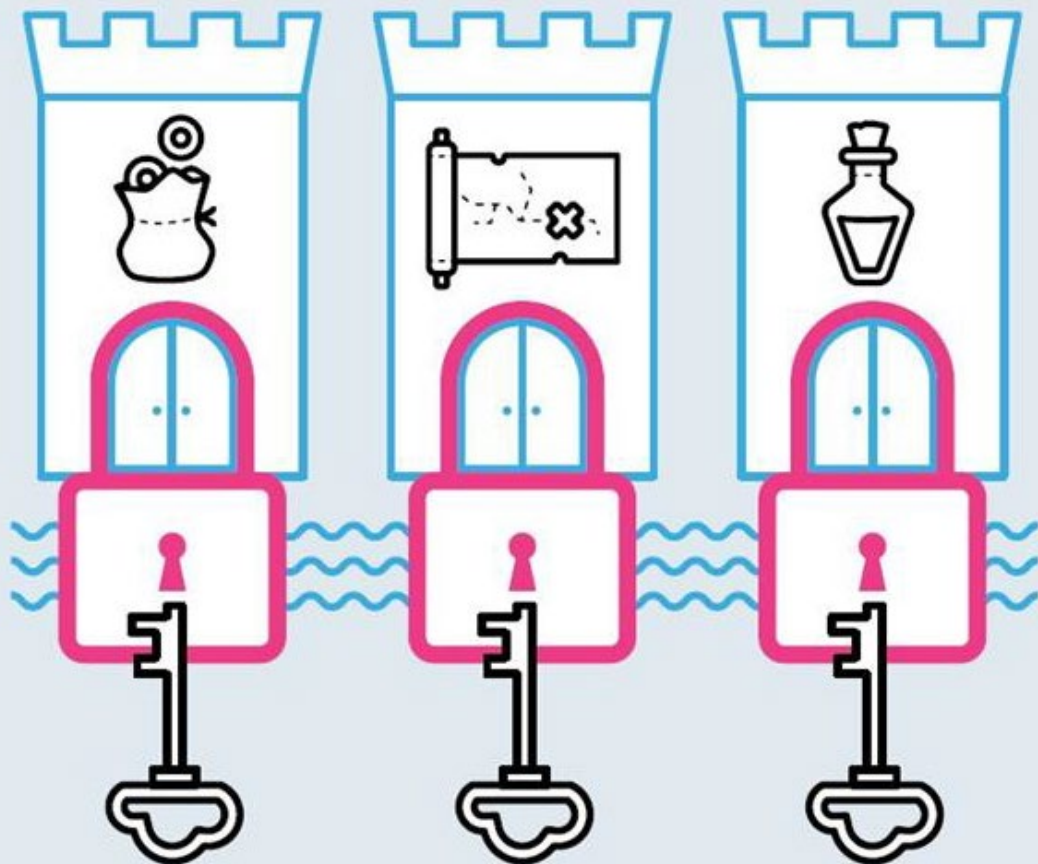
Zero trust PILLARS

"never trust, always verify."

Right peer

with

Right credential

on behalf of

Right user

Project evolution and current status

The IDEA4RC solutions are designed to solve specific needs of the users in the pilots aligned with the objectives expected by EHDS on secondary use of data.

Year 1

Pilots

Tools

Year 2

What's next

User Story Methodology
WHO, WHAT, WHY, SCENARIO

AI models and NLP

Bot Framework

Speech

Your Assistant
Your Brand
Your Personality

Dispatch

QnA

{ }
Language Understanding

Vision

USER IDENTIFICATION

USER REQUIREMENTS

USE CASES

USER SCENARIOS

INFRASTRUCTURE, SERVICES, VIRTUAL ASSISTANTS AND MULTIMODAL NAVIGATORS

PILOT IMPLEMENTATIONS