EUCAIM
CANCER IMAGE EUROPE

European Cancer Images Federation

Ignacio Blanquer (Universitat Politècnica de València) on behalf of the EUCAIM consortium

Co-funded by
the European Union

Content

- EUCAIM background
- EUCAIM in a nutshell
- Actors in EUCAIM
- Architecture of EUCAIM
- Data in EUCAIM
- Sustainability
- Conclusions

Co-funded by
the European Union

Europe's Beating Cancer Plan

- **New EU approach to cancer prevention, treatment and care**
- Four key action areas

Prevention	**Early detection**	**Diagnosis and treatment**	**Quality of life**
Adressing key risk factors	*Improving access, quality, diagnostics*	*Ensuring intergrated, comprehensive care*	*Improving support, rehabilitation, integration*

- One of the objectives of the EBCP is to make the most of the potential of data and AI to combat cancer
- 10 flagship initiatives, including the **European Cancer Imaging Initiative**

Co-funded by the European Union

EUCAIM in the context of the Health Data Spaces

The European Health Data Space (EHDS)

1. Empowering individuals through increased digital access to and control of their electronic personal health data,.

2. Fostering a single market for electronic health record systems, relevant medical devices and high risk AI systems.

3. Providing a trustworthy and efficient set-up for the use of health data for research, innovation, policy-making and regulatory activities (secondary use of data).

Health Data Spaces Pilot
(https://ehds2pilot.eu/project/)

EU infrastructure ecosystem for the secondary use of health data for research, innovation, policy making and regulatory purposes

Testing and Experimental Facilities

Specialised large-scale reference sites open to all technology providers across Europe to test and experiment at scale state-of-the art AI solutions, including both soft-and hardware products and services in real-world environments.

Four areas: Agri-Food, Healthcare, Manufacturing & Smart Cities & Communities.

Testing and Experimentation Facility for Health AI and Robotics
(https://www.tefhealth.eu/)

It aims to test and validate innovative artificial intelligence (AI) and robotics solutions for the healthcare sector and accelerate their path to market.

Infrastructures in Flagship Areas

The Genomic Data Infrastructure
(https://gdi.onemilliongenomes.eu/)

The Genomic Data Infrastructure (GDI) project is enabling access to genomic and related phenotypic and clinical data across Europe. It is doing this by establishing a federated, sustainable and secure infrastructure to access the data.

European federated infrastructure for cancer images (EUCAIM)
(https://cancerimage.eu/)

A pan-European federated infrastructure of cancer-related images for the validation and development of AI tools, which will support and enhance the cancer diagnosis procedure, treatment and the identification of the need for predictive medicine.

European Commission

EUCAIM

EUROPEAN CANCER IMAGING INITIATIVE

#euCancerImaging

Co-funded by the European Union

European Cancer Imaging Initiative

Cancer imaging datasets exist for different cancer types, but are scattered and not easily accessible

What is the ECII trying to achieve?

- Foster innovation and **deployment of digital technologies** for improved clinical decision-making, diagnostics, treatment and prediction.
- Link resources and databases to establish an **open infrastructure of cancer images** for development and benchmarking, and piloting tools.
- **Showcase access and use of medical images**, while ensuring privacy, trust and security
- Make large amounts of cancer images and linked data easily accessible in line with the **European Health Data Space** and **EOSC.**

Co-funded by
the European Union

The project

- Research infrastructure developed by the EU-funded **EUCAIM project**
  - Coordinated by EIBIR, scientifically led by Prof. Luis Martí-Bonmatí (HULAFE, Valencia/ES)
  - Consortium: 76 partners from 14 countries
  - Runtime: January 2023 - December 2026
  - Budget: €35.6m
- Involving the major RIs in Health.

EURO BIOIMAGING

elixir BBMRI-ERIC eatris

Co-funded by the European Union

The project

Partners and Stakeholders

EUROPEAN INSTITUTE FOR BIOMEDICAL IMAGING RESEARCH

Instituto de Investigación Sanitaria La Fe

Beneficiaries: 65
Affiliated Entities: 10
Associated Partner: 1

Partners: 76

Stakeholders: 22

#euCancerImaging

European Commission

Vision and mission

Vision

Enhance cancer **diagnosis** and **treatment** through AI tools

Mission

- **Deploy a hybrid federated infrastructure** to power up AI & imaging to beat cancer
- Provide a research platform for the **development & benchmarking of AI tools** toward Precision Medicine
- **Address the fragmentation** of the existing cancer image repositories by building a distributed **Atlas of Cancer Images** *(>60m anonymised cancer images)* accessible to clinicians, researchers and innovators
- Create a federated data warehouse approach for deploying observational studies

Co-funded by
the European Union

The Atlas of Cancer Images

EUCAIM

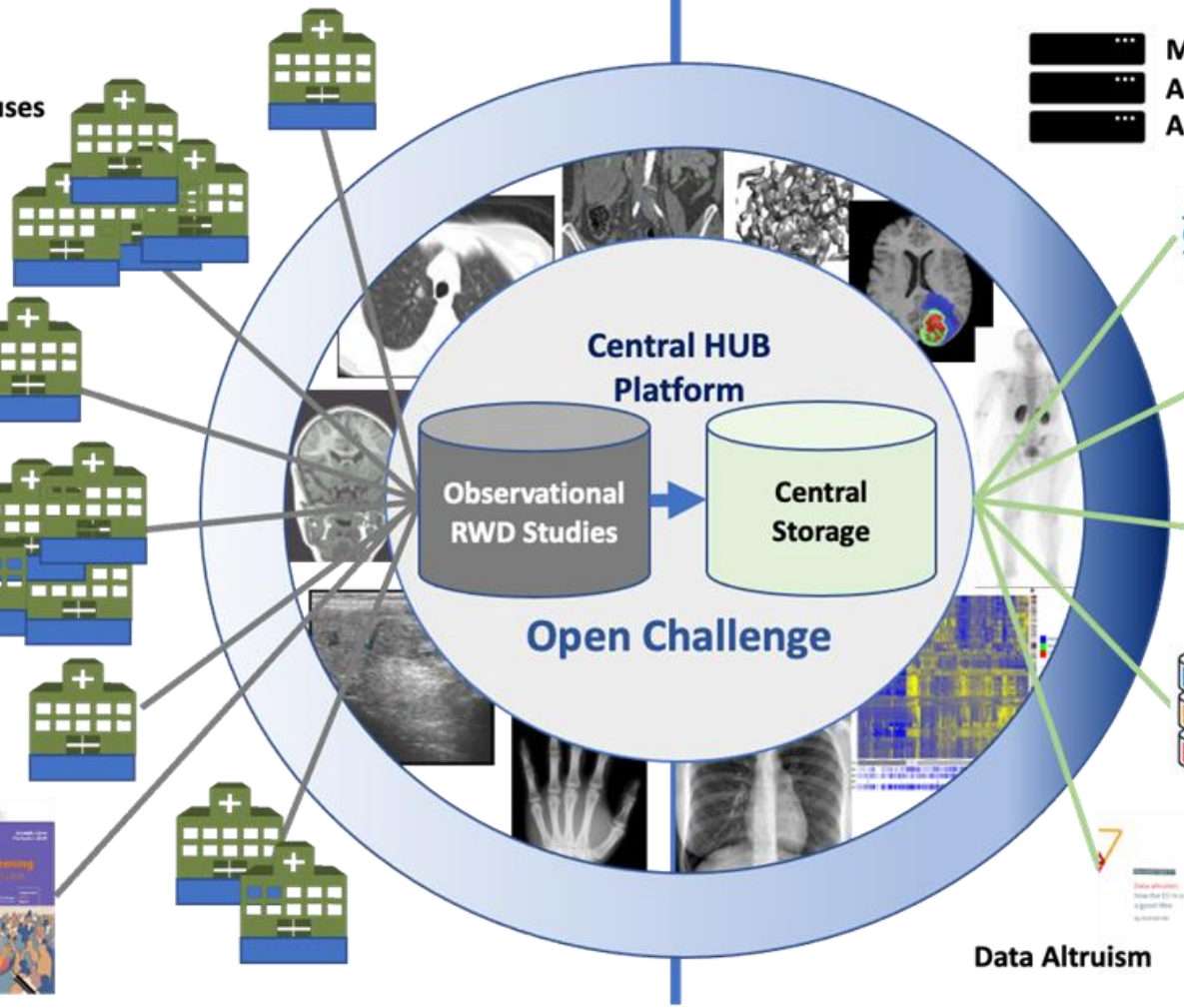
Hybrid Platform

Distributed RW Data Warehouses
Extract, Transform and Load
ML Federated Learning

CDM hyper-ontology

Primary & Secondary Used Area

Cancer Screening Programs



Metadata Catalogue
Annotated Structured Data
AI Experimentation Platform

AI4Health Imaging

Atlas of Cancer Images

ERIC
EUROPEAN RESEARCH INFRASTRUCTURE CONSORTIUM

CLINICAL TRIALS

Secondary Use Area

DICOM-MIABIS

Data Altruism

Co-funded by the European Union

Actors in EUCAIM

Co-funded by
the European Union

User roles

Data Provider/Data Holder/Data Controller

Definition: Any natural or legal person, including entities, bodies, and research organisations in the health or care sectors, as well as European Union institutions, bodies, offices, and agencies, who has the right, obligation, or capability to make certain data available, including registering, providing, restricting access, or exchanging the data.

Two options for joining the federation:

- Become a federated node
- Upload anonymised data to the central storage.

Tool Provider

Definition: Entity (startups, enterprises, research institutions, government agencies, non-profit organisations) that would like to contribute with processing tools, services, or applications they have developed to the EUCAIM's marketplace for use in the federated processing module of the platform.

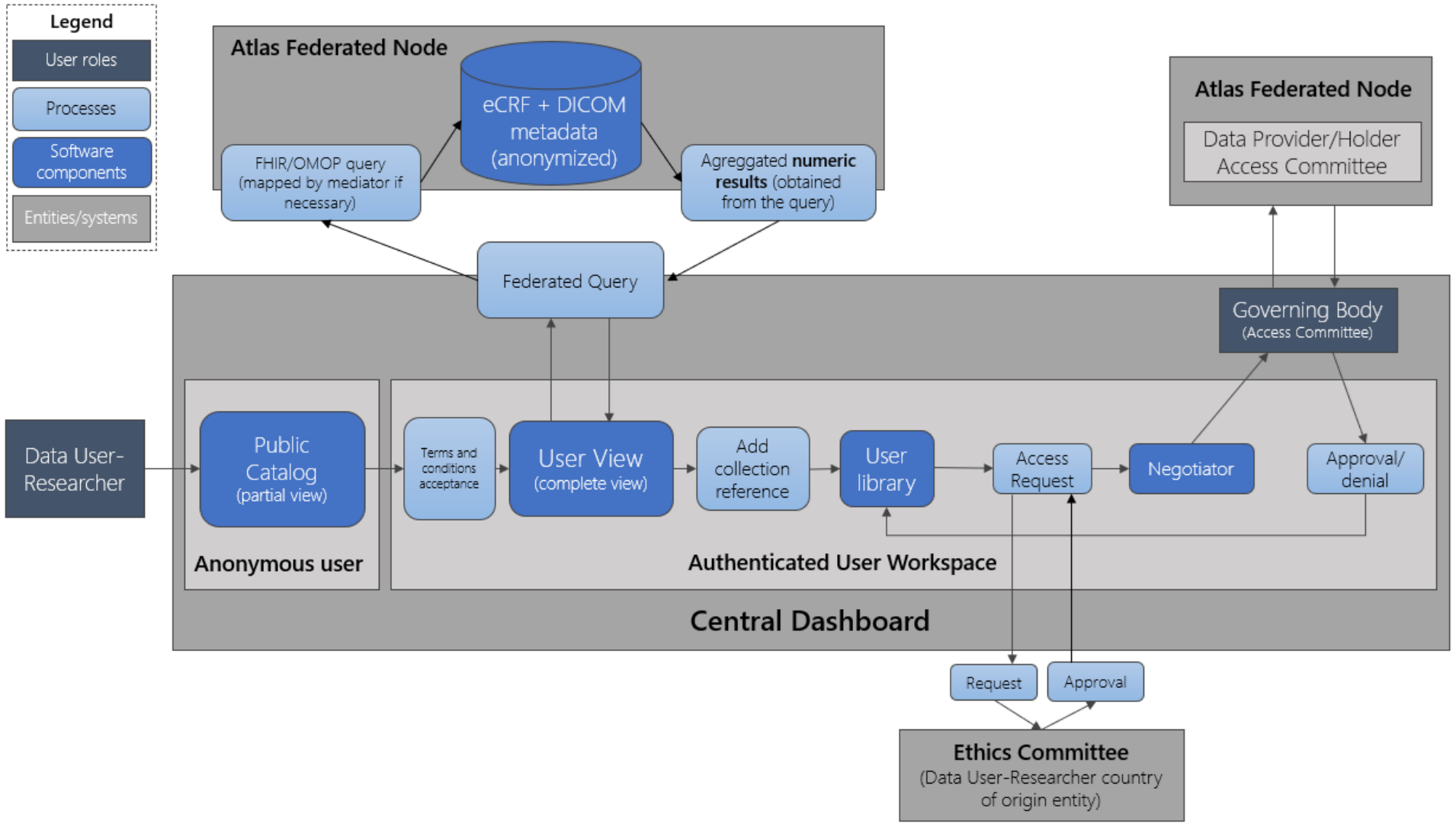
Both batch and interactive applications, following well-defined rules for participation and technical compliance guidelines.

Data User-Researcher

Definition: A person or entity that wants to explore the public catalogue and eventually request access to data and process them using either the tools available in the platform or their own AI tools to conduct studies, research, or analysis with the intention of generating new knowledge in the field of medicine and publishing the findings.

A data access request should be made through a Research and Development (R&D) project that will be evaluated by the Access Committee.

User's Journey

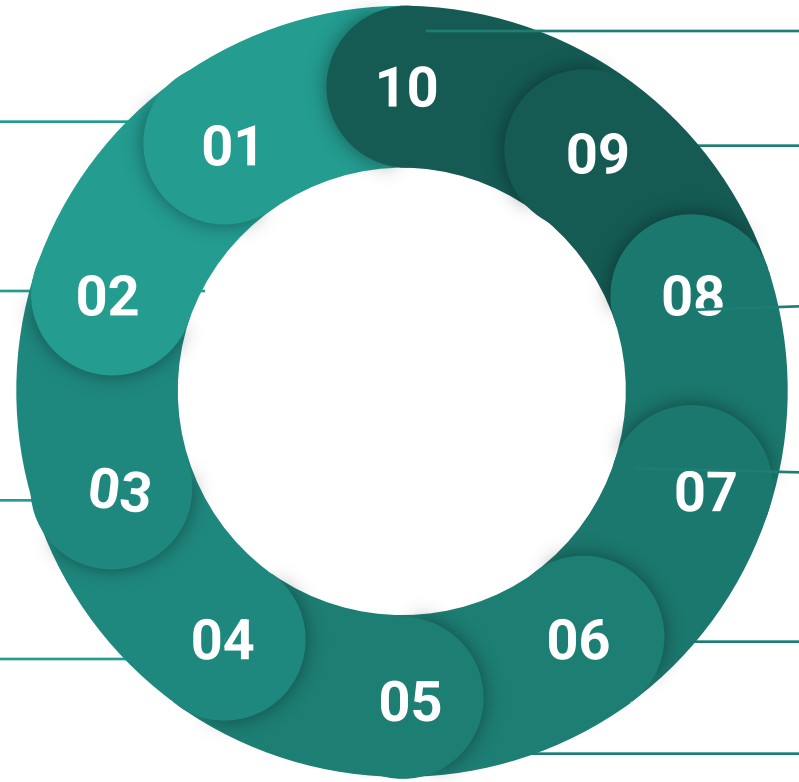


Co-fu
the E

EUCAIM Architecture

Co-funded by
the European Union

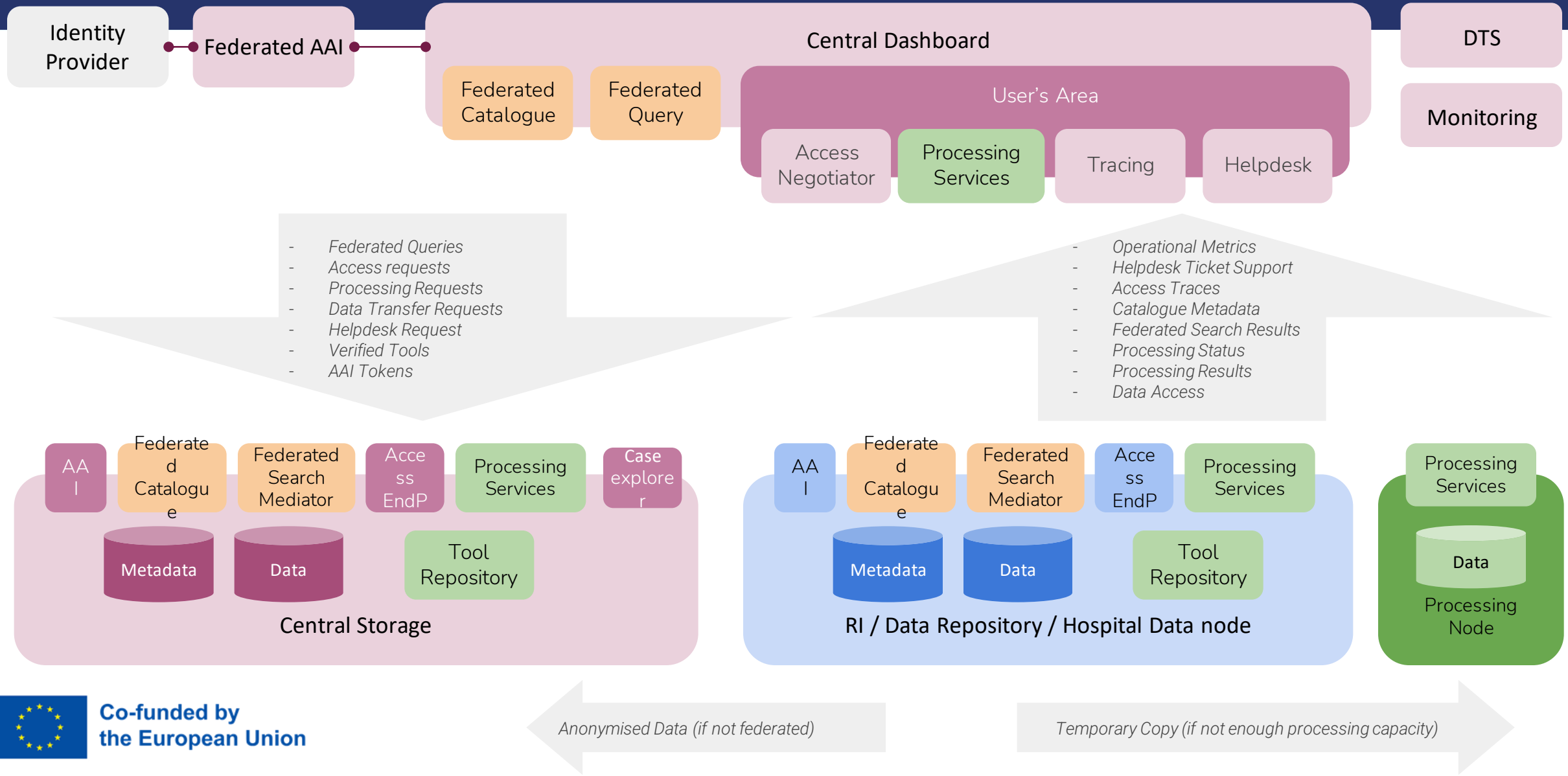
Architecture Components



- 01 Dashboard
- 02 Local Data warehouses
- 03 Authentication & Authorisation Infrastructure
- 04 Federated Catalogue
- 05 Federated Search
- 06 Access Negotiation
- 07 Distributed Processing
- 08 Data Transfer Service
- 09 Monitoring and Tracing
- 10 Helpdesk

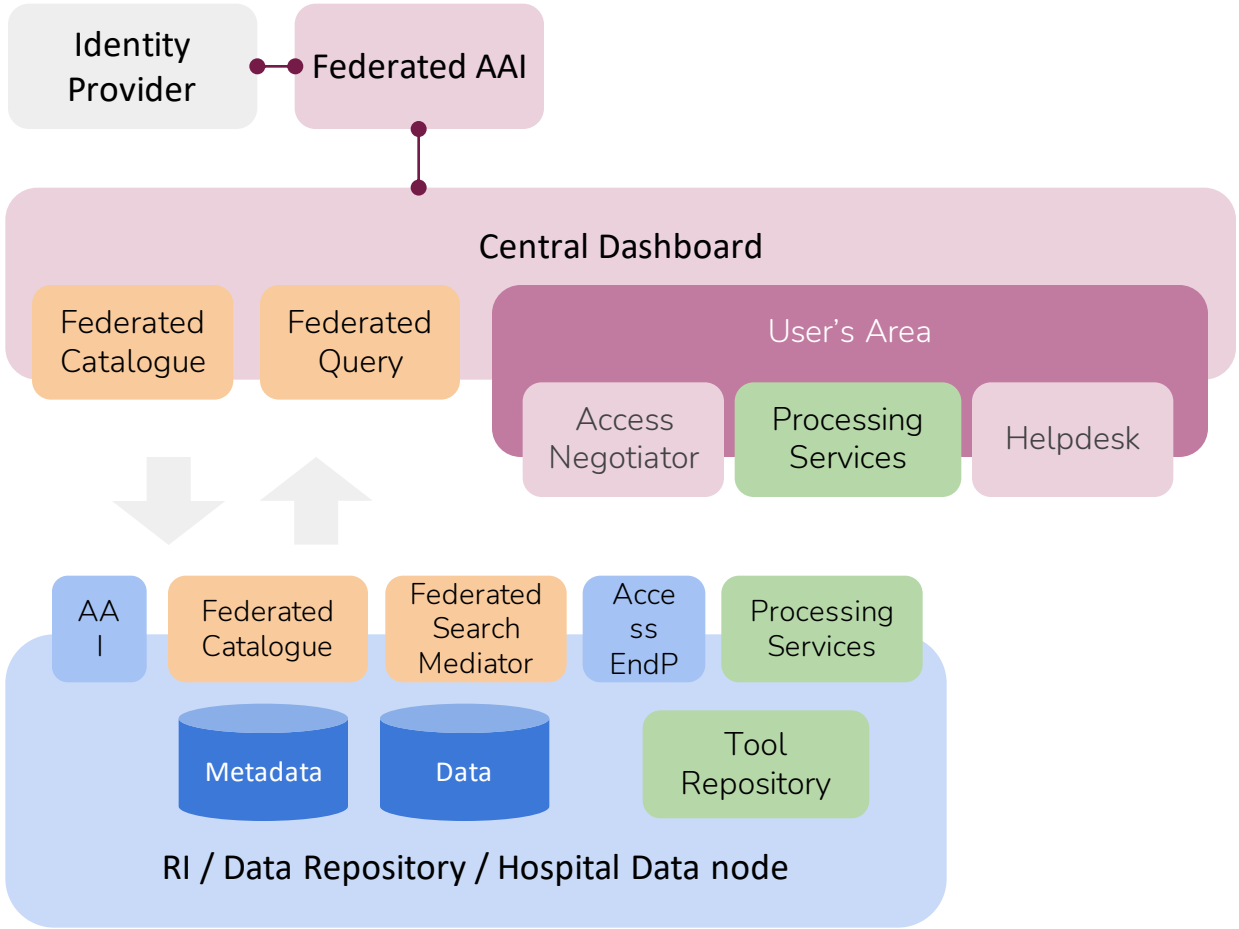
Co-funded by the European Union

Overview of the Architecture



Co-funded by the European Union

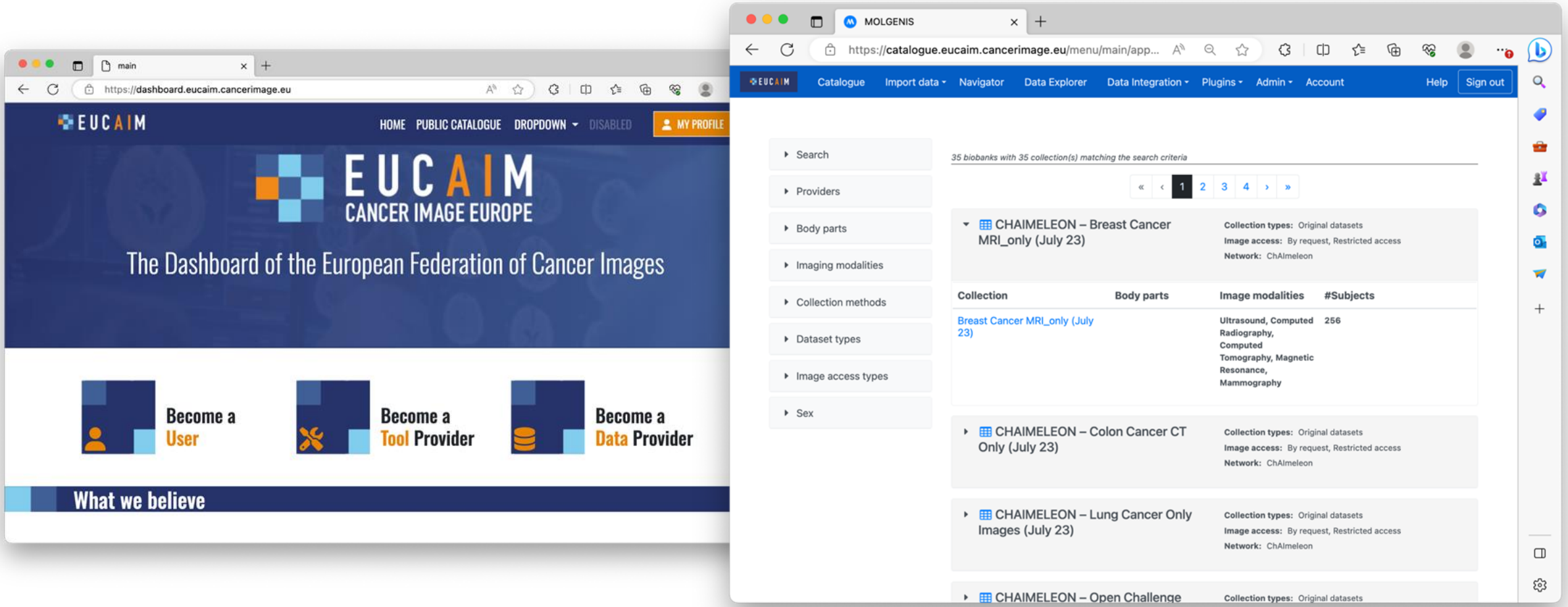
Demonstration Prototype



- A Dashboard with a personal area supporting LS AAI.
- A Catalogue with 35 dataset from AI4HI registered (>200K series of images).
- A Federated Query engine with limited functionality.
- An Access negotiation service.
- A separated distributed processing service.

Co-funded by the European Union

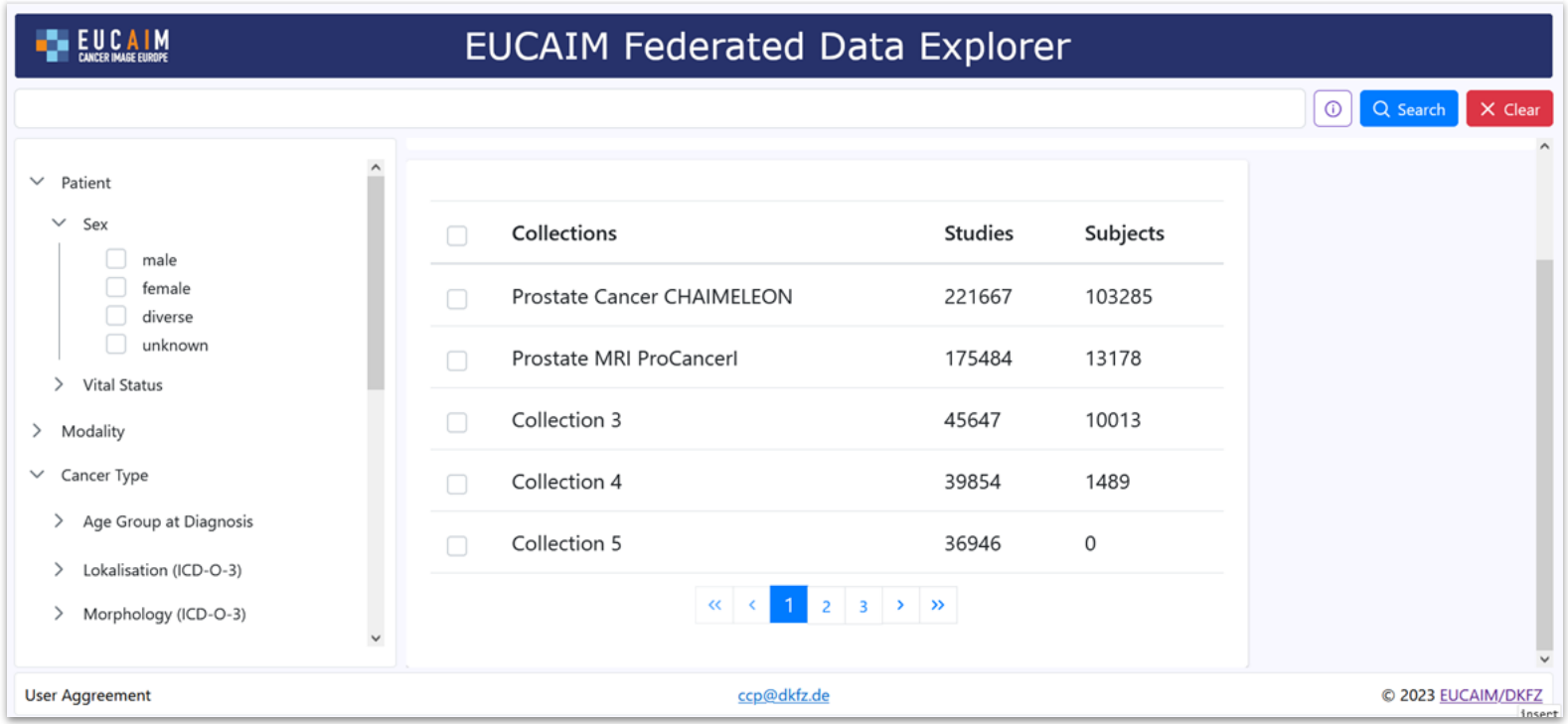
A Dashboard and a public catalogue



Co-funded by the European Union

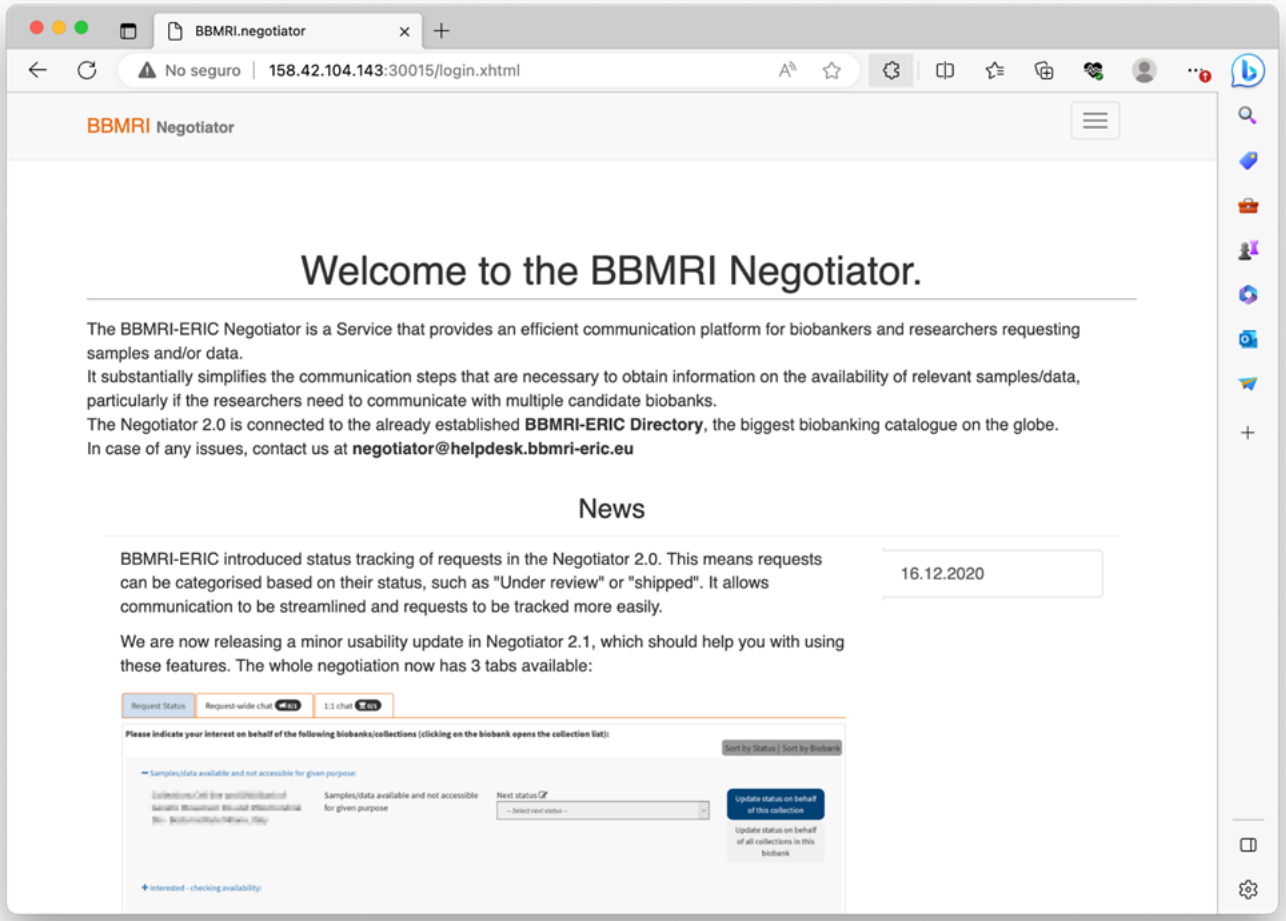
Federated Search

- Setup of piloting system based on:
  - Samply.Lens (https://github.com/samply/lens) for the search interface (same tech as for BBMRI Locator)
  - Samply.Beam (https://github.com/samply/beam) for network communication
  - Query language translation locally in the mediator component.



Co-funded by the European Union

Negotiator



- An access request is performed through a “negotiator” platform.
- Users provide with the necessary information to evaluate a request.
- The access level is negotiated among the requesters and the providers.

Co-funded by the European Union

EUCAIM Data

Co-funded by
the European Union

Data Contribution and Access

Central Repository & Federated Nodes:

a) **Central Repository (Data Transfer Agreement):** Provide information about the research project, metadata catalogue and software.

b) **Federated data providers** (Data Sharing Agreement): Provide information about the research project, metadata catalogue and software. Computational and storage information.

c) **Research Communities** (Data Transfer Agreement and Collaboration Agreement): Provide information about the research project, partners, metadata catalogue and software.

d) **Hospital data providers** (Collaborative Agreement, DSA/DTA per project): Provide information about how the hospital data warehouse is structured: CDM (OMOP/FHIR), HPC requirements, IT policies.

User Requirements:

a) Only a **Research Project** (from final undergraduate projects to large funded grants)

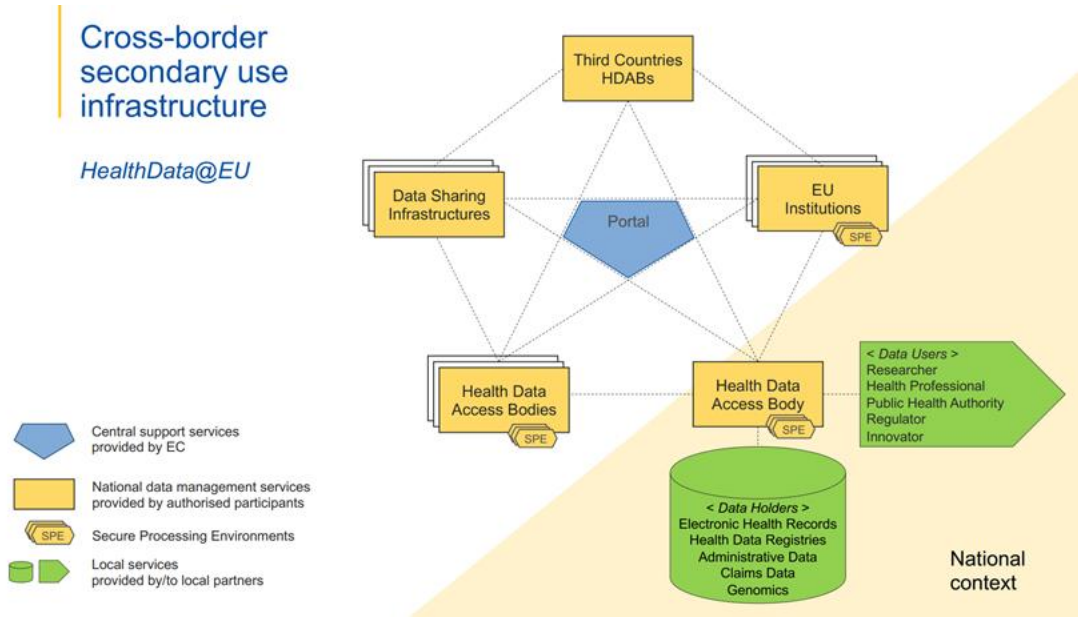
Co-funded by
the European Union

DATA in EUCAIM

EUCAIM Data Curation and Transformation:

Check and transform, if necessary, to our **Common Framework**:

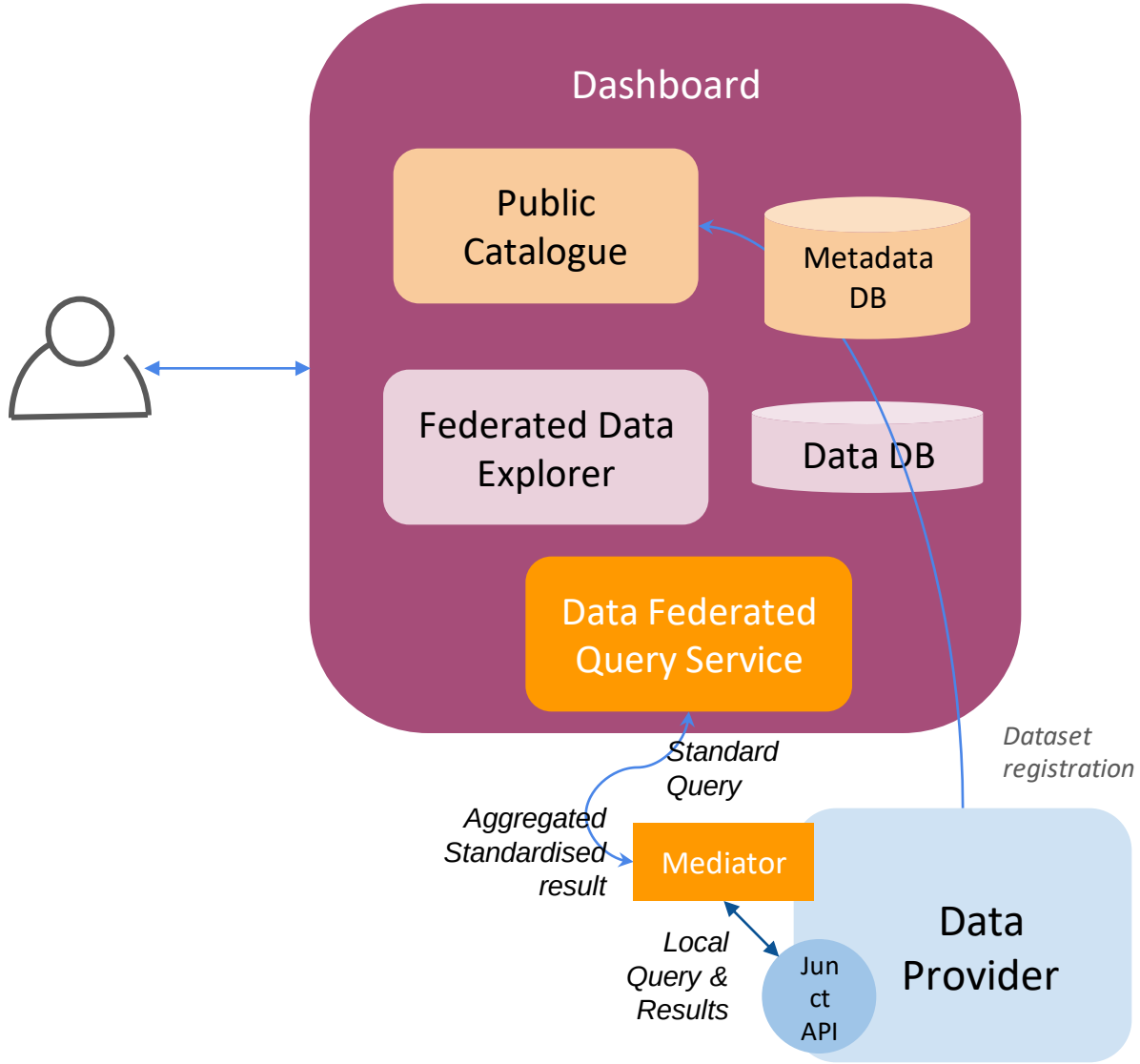
- Data anonymization
- CDM, standardized metadata compliance, data ontology
- DICOM metadata exploitation
- Data harmonization
- Data annotation and labeling
- Data cleaning, enhancing and integrating
- Time coherence
- Image quality assessment
- FAIR compliance
- Clinical endpoints definition



Co-funded by the European Union

Common ontology

- Standard common ontology at the level of the dataset metadata
- Higher flexibility when dealing with data models
  - Either data is transformed to a common ontology model or mediators are provided
  - Different levels of compliance are forecasted to facilitate onboarding.



Co-funded by the European Union

EUCAIM's Sustainability

Co-funded by
the European Union

EUCAIM's Sustainability

Sustainability of EU Funded Results

- EU Research Projects have developed their own data and tools repositories.
- They have also created a network of researchers on a specific topic.
- Sustainability of their repositories is a concern.
- Possible linking Options (Memorandum Of Understanding):
    - Metadata Catalog (as soon as possible) and Federated Node (as soon as available, **Open Challenge**).
    - Transfer Data set (images, related data and tools), to be use by EUCAIM while keeping the virtual research network (**Research Community** within EUCAIM-EDIC).

Co-funded by
the European Union

Conclusions

- EUCAIM aims at consolidating the Cancer Imaging research community through a sustainable infrastructure.
- Data is extremely sensitive and complex, requiring a widely accepted legal framework.
  - The federated model will be key to fulfil the legal constraints.
  - The construction of the federation is cumbersome and technically challenge.
- FAIR principles and interoperability with other actions in health, and interdisciplinary areas such as environment, social sciences, climate, etc. are key.

Co-funded by the European Union

Thank you for your attention

2023

Design completed

- Requirements analysis
- Design
- Collaboration mechanisms
- Early release of the data federation framework

2024

First version of platform

- Platform validated and populated for external production
- Data providers connected
- Prototype for federated learning
- Benchmarking platform

2025

Final release of platform

- Federated learning
- Final version of tools and services
- Federation of new cancer images databases through open calls
- Implementation of clinical use cases

2026

Full operation of federated repository

- Integration with other data infrastructures
- Piloting of the business model
- Legal and operation model finalised

2027

Expansion

Get in touch!

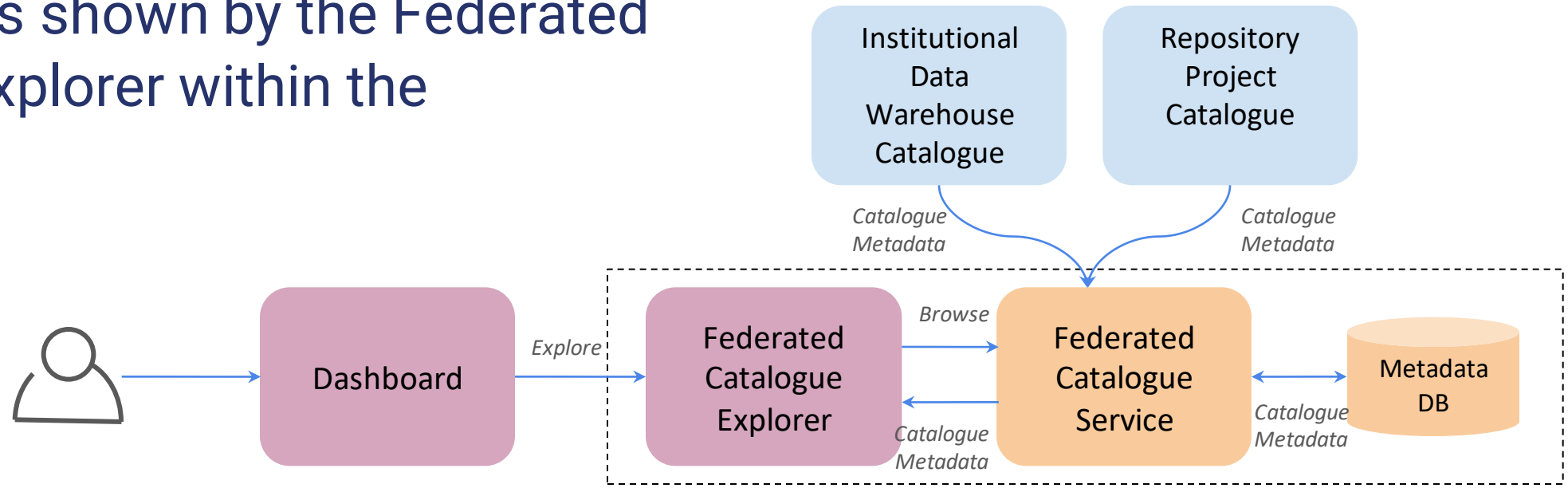
Co-funded by the European Union

Backup slides

Co-funded by
the European Union

A Federated Catalogue

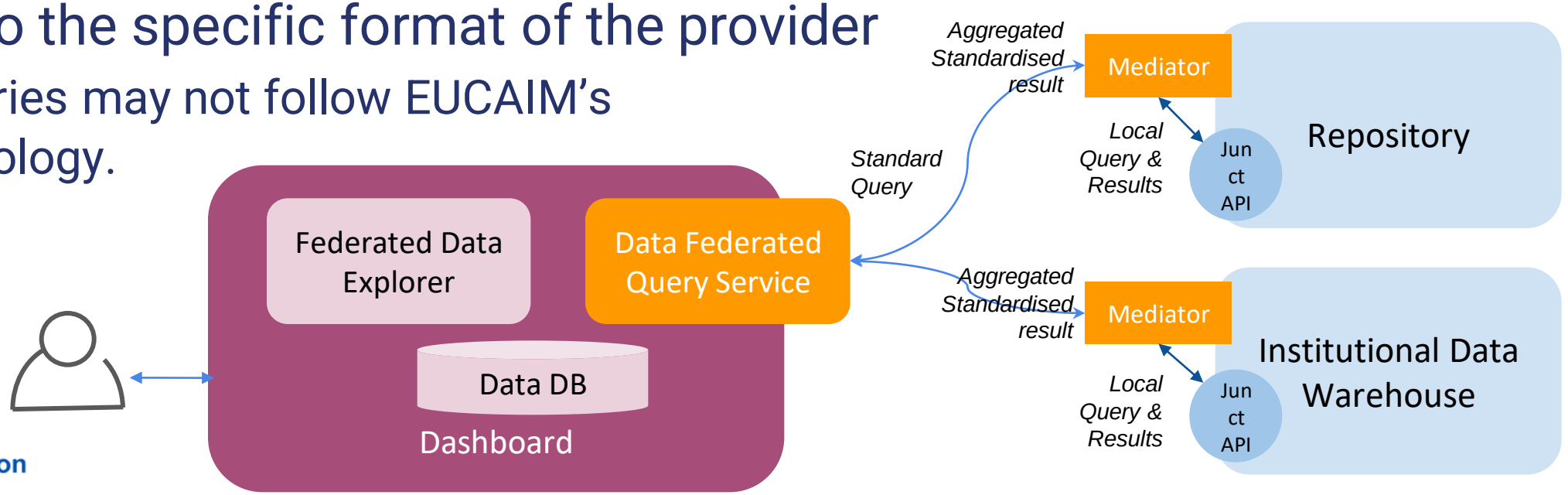
- A federated catalogue indexes the collections from the providers of the federation and the central storage.
- Collections' metadata is registered in the Federated Catalogue Service
  - Collections metadata follow a common metadata model.
- These data is shown by the Federated Catalogue explorer within the Dashboard.



Co-funded by the European Union

Federated Query

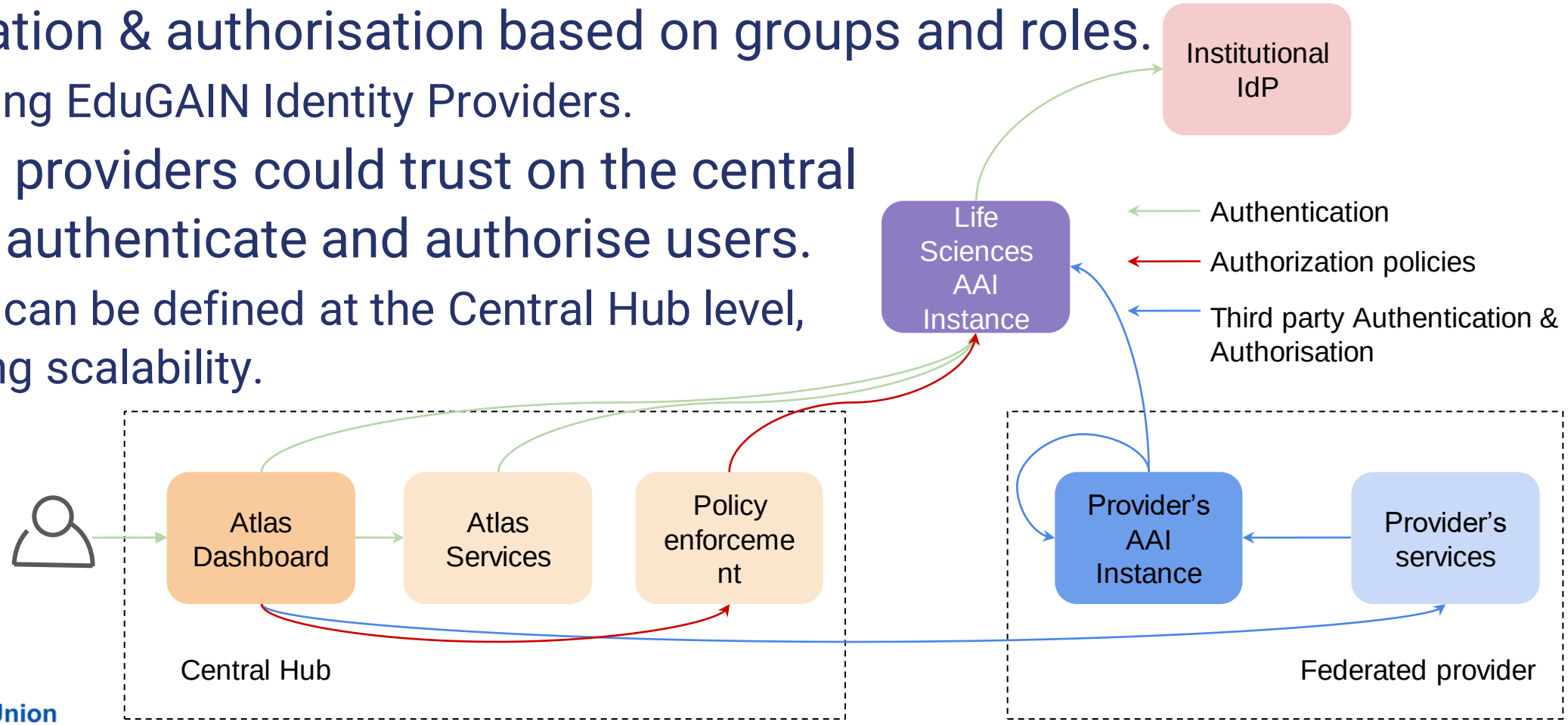
- A Federated Query service can retrieve the number of cases which fulfil a searching criterion
  - Searching criteria based on selected fields of the images metadata defined by the hyperontology (E.g. Body Part, Gender, Age range, …)
- This will require that providers have a mediator service that adapts the query to the specific format of the provider
  - Repositories may not follow EUCAIM's hyperontology.



Co-funded by the European Union

Authentication and Authorisation Infrastructure

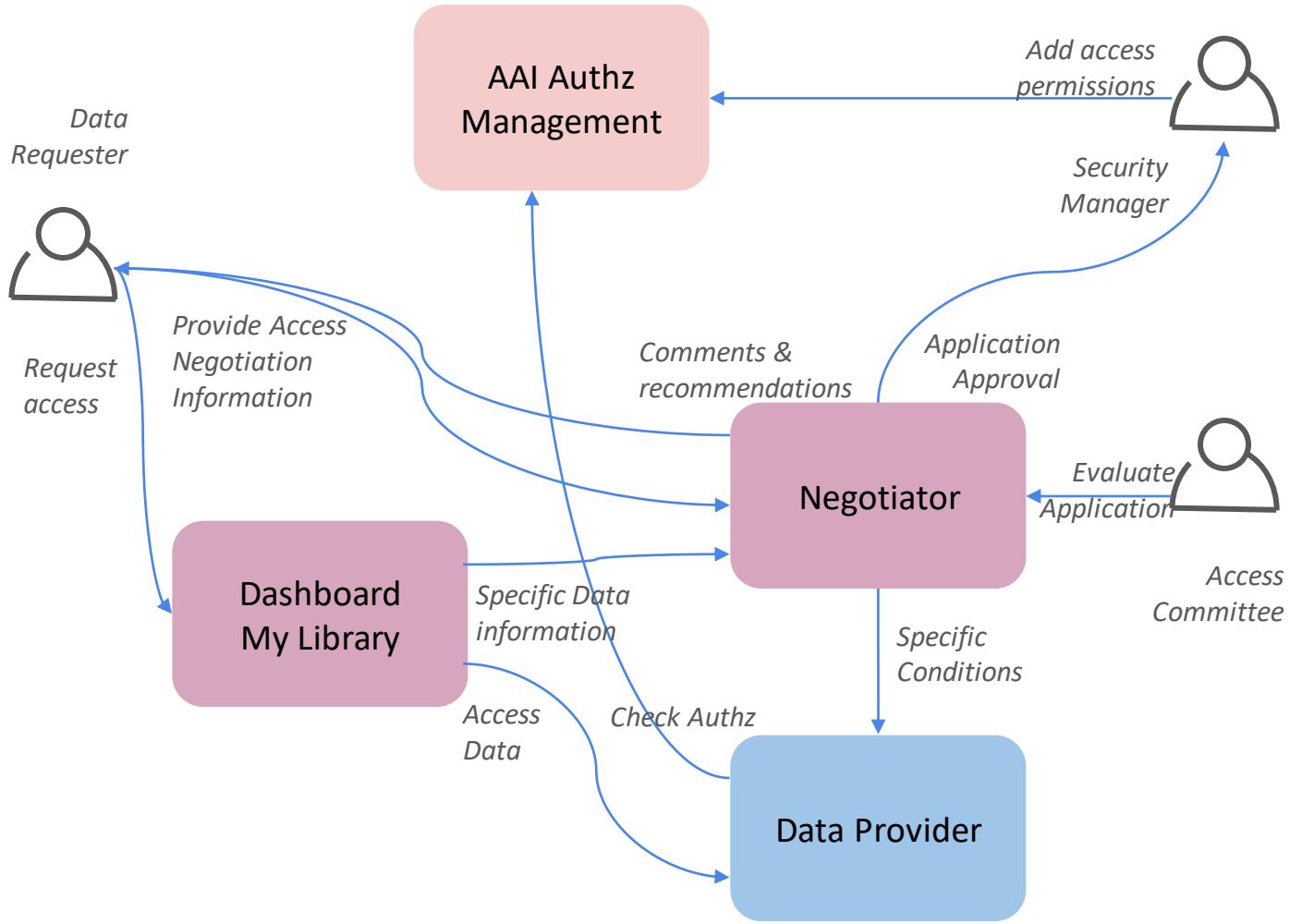
- A central AAI (Based on Life Sciences Login AAI) supports authentication & authorisation based on groups and roles.
  - Supporting EduGAIN Identity Providers.
- Federated providers could trust on the central service to authenticate and authorise users.
  - Policies can be defined at the Central Hub level, improving scalability.



Co-funded by the European Union

Access Negotiation

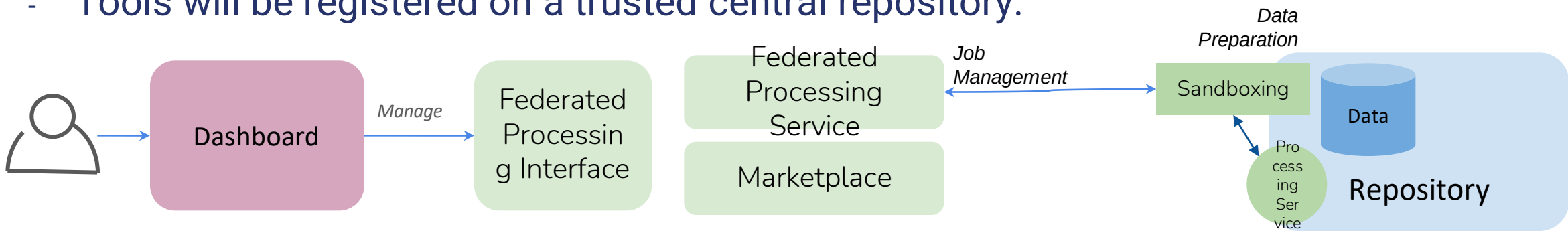
- Access to the collections will require submitting a project proposal, ethical evaluation and institutional mandates.
- Proposals will be evaluated by the Access Committee, involving the providers of the federation.
- The proposal could go through several iterations, until it gets accepted.
- Central authorisation is updated.



Co-funded by the European Union

Distributed Processing

- Federated processing will trigger the execution of jobs on the providers of the federation
  - Eventually it will require temporary data transfer to a trusted processing service.
- Data must be prepared in a platform-agnostic way so processing tools can run on different providers.
  - Either copied or linked to a sandbox area.
  - Tools will be registered on a trusted central repository.



Co-funded by the European Union