An Integrated Medical Imaging Archive and Secure Processing Environment

I. Blanquer, A. Alic, P. Lozano, D. Segrelles, P. Montoliu, D. Arce

Universitat Politècnica de València

Co-funded by the European Union

GENERALITAT VALENCIANA
Conselleria de Innovación, Industria, Comercio y Turismo

EUCAIM
CANCER IMAGE EUROPE

Mision and Vision

EUCAIM
CANCER IMAGE EUROPE

We want to bring together medical images that are currently stored in many different places, so that this information can be safely reused for research.

By connecting radiology and nuclear medicine image repositories (providing >60 million of anonymized cancer images) we can better study cancer and develop more accurate AI tools for diagnosis and treatment.

Accessible to doctors, researchers, and innovators.

Enhance cancer diagnosis and treatment through Imaging Biomarkers and AI Tools

Development & Benchmarking of AI tools and Models toward Precision Medicine

Create a Federated Secure Processing environment for deploying observational studies



95 partners, including established major European Research Infrastructures (Euro-BioImaging, BBMRI, EATRIS, ELIXIR), cutting-edge hospitals, universities, and private companies in the med-tech industry.

Co-funded by
the European Union

Main Achievements

EUCAIM
CANCER IMAGE EUROPE

Data & Infrastructure

104 imaging datasets connected

129,646 subjects available (in public catalogue)

19 cancer types represented

3 compliance tiers for data holders

1 Central Hub in operation for Research & Innovation

2 Reference Nodes as central storage with SPE

AI & Innovation

94 software tools in the EUCAIM toolbox

3 federated processing demonstrators released

47,000+ studies searchable through the federated search service

Federated processing environment established

User Community

315 registered users with access rights to EUCAIM services

Users from **16 countries**

Growing cross-border collaboration between researchers, clinicians and innovators

Strategic Progress

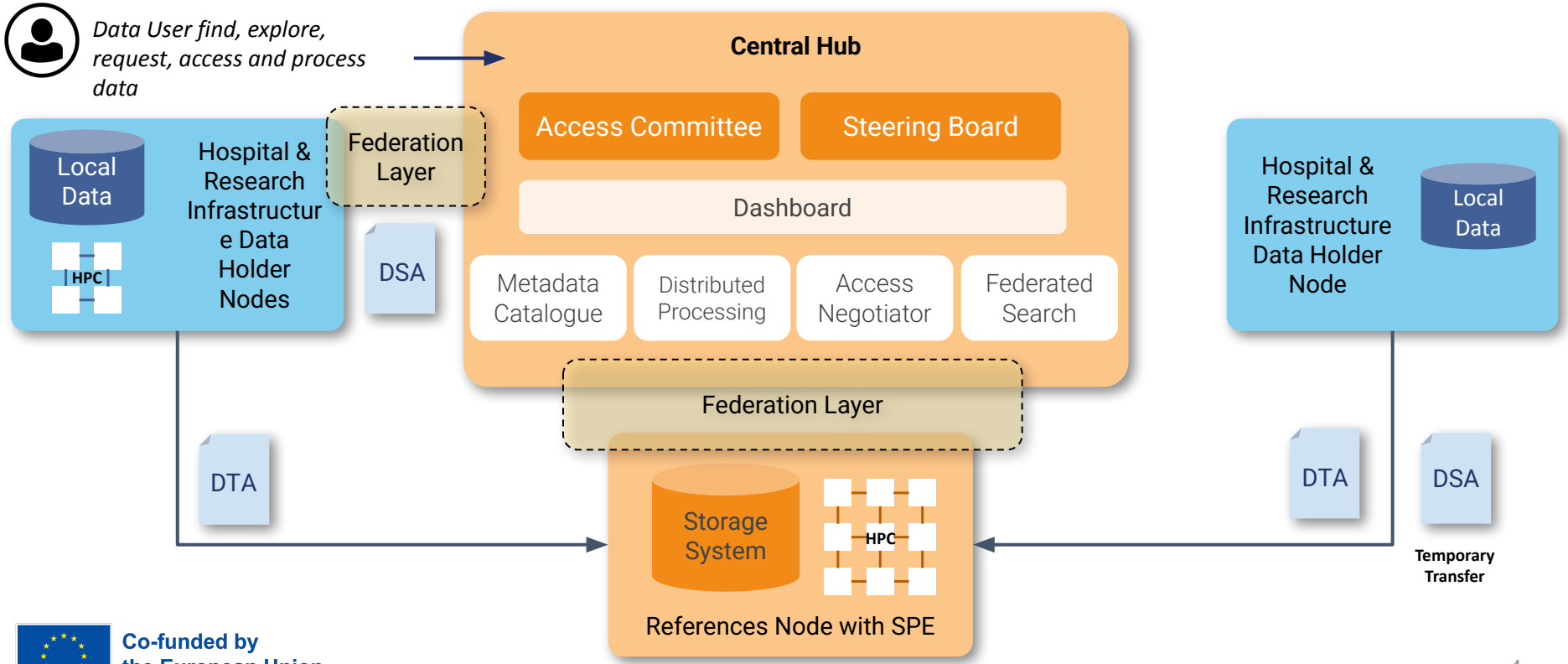
Central Hub Services operational:

- Registry
- Federated Search
- Access Negotiation
- Helpdesk
- Training
- Monitoring
- LS-AAI
- Dashboard

Platform continues to expand towards long-term sustainability

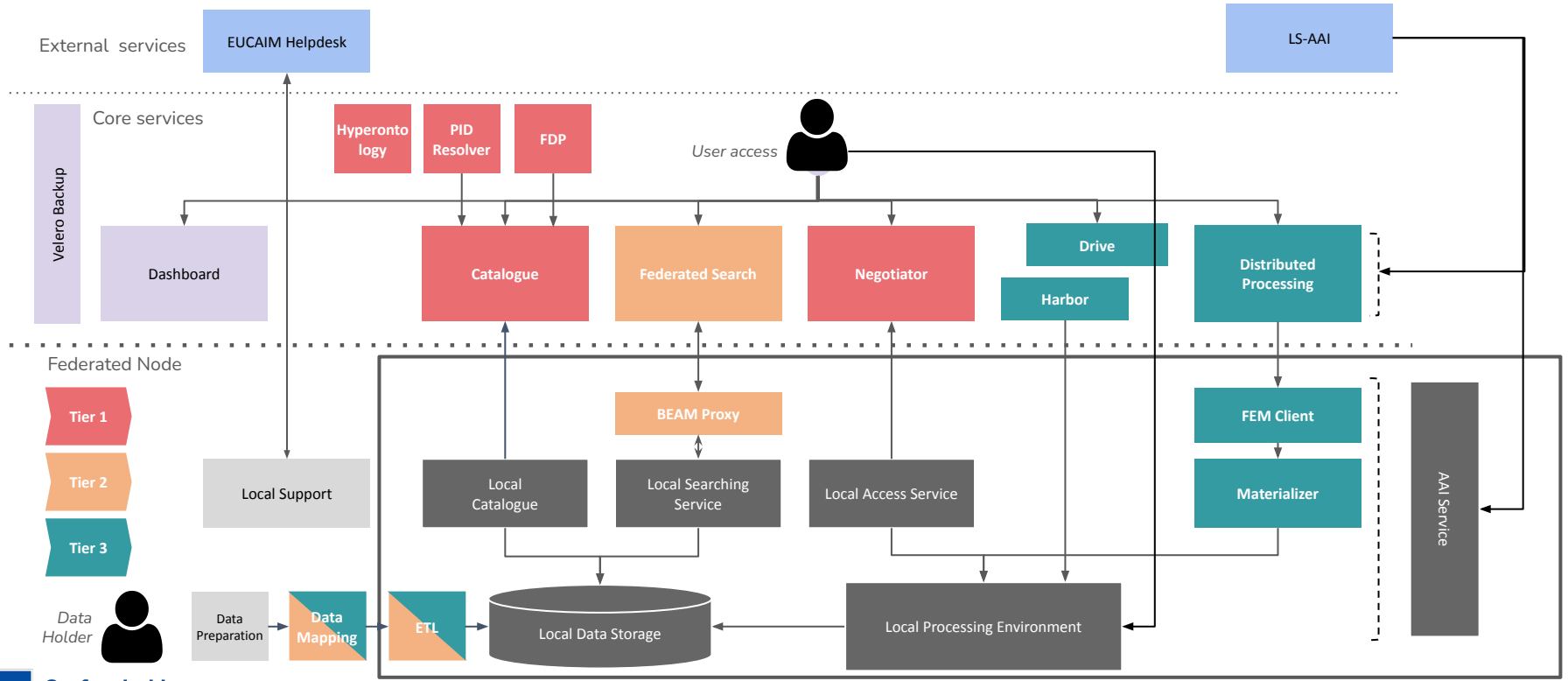
the European Union

EUCAIM Architecture



Co-funded by the European Union

Architecture of the Federation



Co-funded by the European Union

- A SW release using Open Source components
  - Installation and usage instructions in this document.
- It installs several components:



Own Components

- Dataset collections management (Dataset Explorer and Dataset Service).
- Secure Virtual Environments management.
- Batch jobs manager (Jobman).
- DSWS Operator (workspace management).
- Routing components for minikube.
- Base SW container for batch jobs.

Third-party components

- Keycloak (Authentication and Authorisation with OIDC)
- Guacamole VRE
- Harbor (Container registry SW) and Kubeapps (Kubernetes Helm application registry)
- Kubernetes Dashboard, Traefik, Cert-manager
- ORTHANC/OHIF DICOM PACS and workstation
- EUCAIM Federation components (Focus, Beam Proxy, FEM Client).

Co-funded by
the European Union

Measures to be considered

The European Health Data Space (EHDS)(*) defines the measures that a SPE should feature:

a) Restrict access to authorised persons listed in the respective data permit;

b) Minimise the risk of the unauthorised reading, copying, modification or removal of electronic health data hosted in the secure processing environment through state-of-the-art technological means;

c) Limit the input of electronic health data and the inspection, modification electronic health data hosted in the secure processing environment to a l authorised identifiable individuals;

d) Ensure that data users have access only to the electronic health data data permit, by means of individual and unique user identities access modes only;

e) Keep identifiable logs of access to the secure processing environment time necessary to verify and audit all processing operations in that environment;

f) Ensure compliance and monitor the security measures referred to mitigate potential security threats.


EUROPEAN COMMISSION

Strasbourg, 3.5.2022
COM(2022) 197 final
2022/0140 (COD)

Proposal for a

REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

on the European Health Data Space

(Text with EEA relevance)

{SEC(2022) 196 final} - {SWD(2022) 130 final} - {SWD(2022) 131 final} - {SWD(2022) 132 final}

EN

EN

Co-funded by the European Union

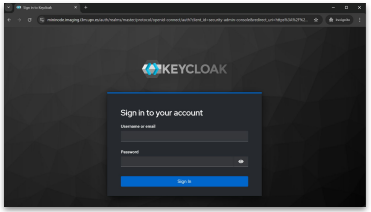
(*) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52022PC0197

Local Node

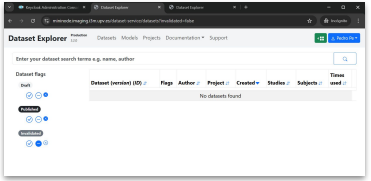
demo-node.imaging.i3m.upv.es

EUCAIM CANCER IMAGE EUROPE

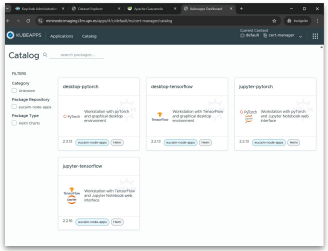
Final applications



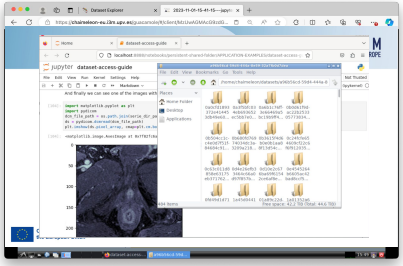
Authentication and Authorisation platform



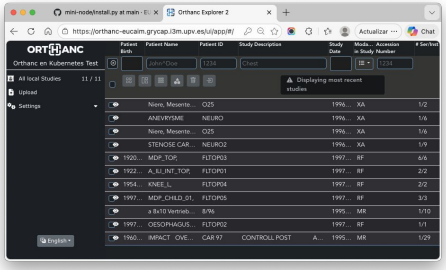
Dataset Catalogue



Interactive Applications catalogue

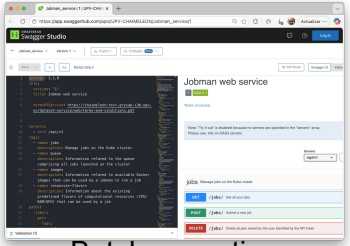


Secure VRE

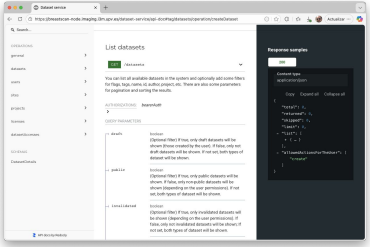


DICOM PACS

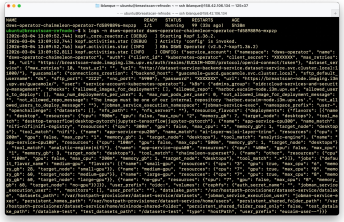
Internal services



Batch execution service



Dataset Service



Data Storage Operator

Co-funded by the European Union

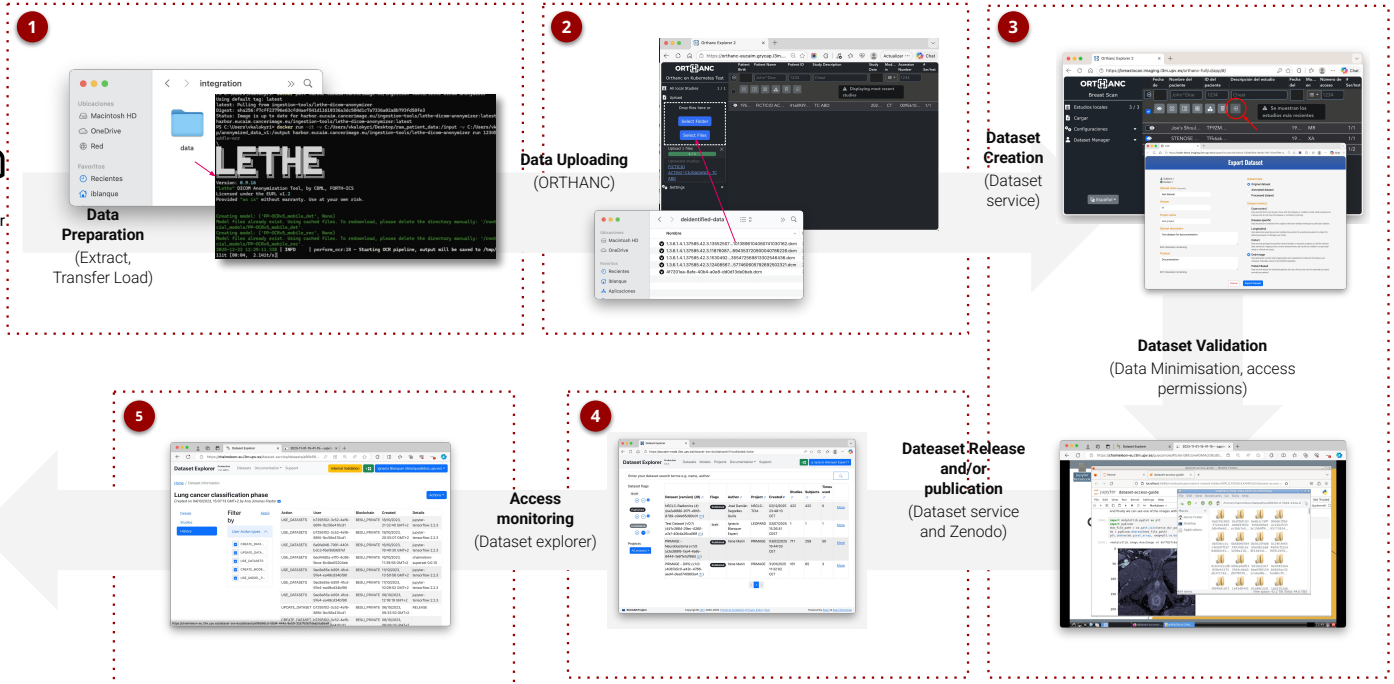
Data Holder's Journey in IMASPE

EUCAIM CANCER IMAGE EUROPE

1. Extract the Imaging and clinical data, annotation, de-identification, risk of re-identification and Data Quality assessment
2. Data Uploading in the Federated Node, second pseudonymisation in case of longitudinal data.
3. Dataset creation and validation
4. Dataset Release or Publication, registering metadata in Zenodo and defining external users access policies.
5. Data Access records evaluation.



Data Manager

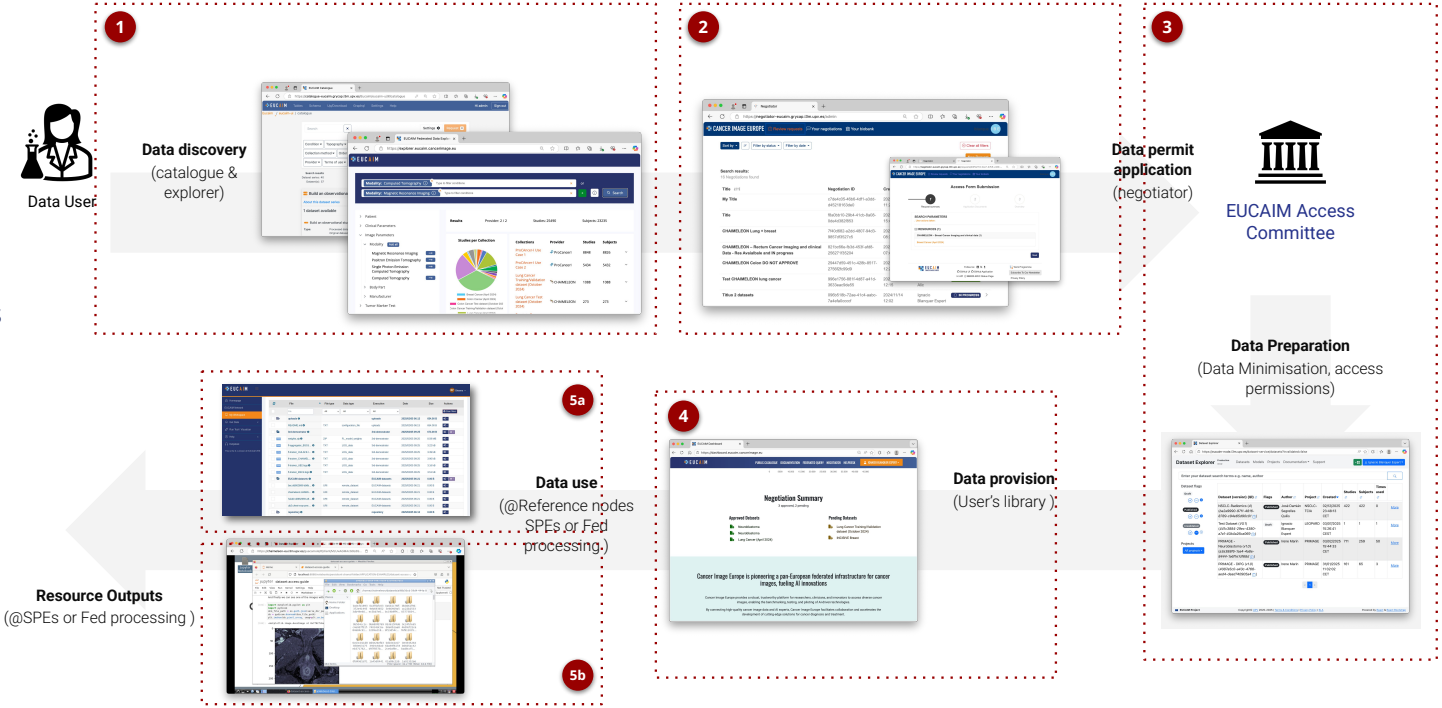


Co-funded by the European Union

User's Journey in IMASPE

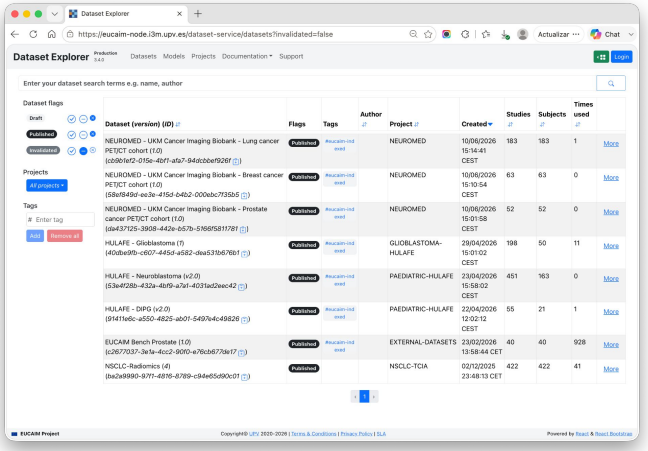
EUCAIM CANCER IMAGE EUROPE

1. A data user accesses the Catalogue and/or the Federated Search to discover the datasets that fit their research.
2. The data user selects the datasets from the catalogue and requests access through the Negotiator application.
3. The Access Committee evaluates the suitability of the request.
4. The data user is notified and the dataset appears in the users' library.
5. The data user accesses the dataset through the provider's SPE or the federated processing system.



Co-funded by the European Union

Dataset Service of IMASPE

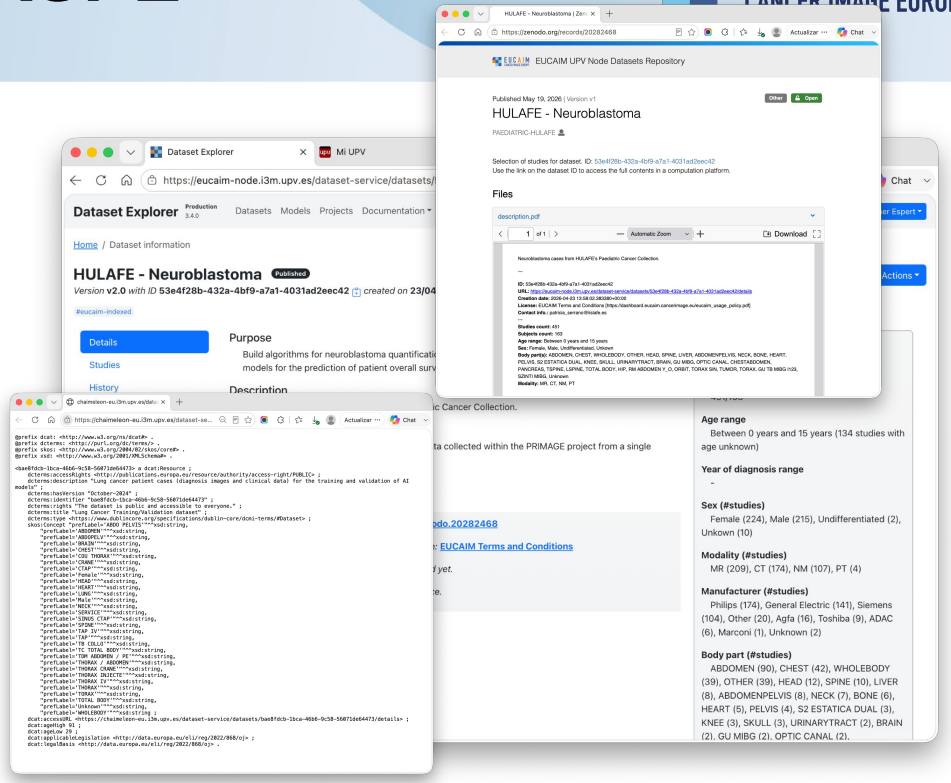


- The **Dataset Service of IMASPE** manages the lifecycle of data providing solutions for data ingestion, data exploration, dataset creation, metadata collection and publication and access management.
- Only users **with permissions** can **upload** data or **create datasets**.
- Datasets could have **four statuses**:
  - Draft, only accessible by the user who created it (dataset validation and preparation).
  - Released. A PID is assigned and it becomes accessible by the users in a specific project.
  - Published. Metadata is registered in Zenodo and access is managed through ACLs (even to users in other projects).
  - Invalidated. Used to make datasets inaccessible without deleting them.
- A dataset is a **virtual grouping** with **additional metadata** of the data available in the platform, and it is the **authorisation unit**.

Co-funded by
the European Union

Dataset Service of IMASPE

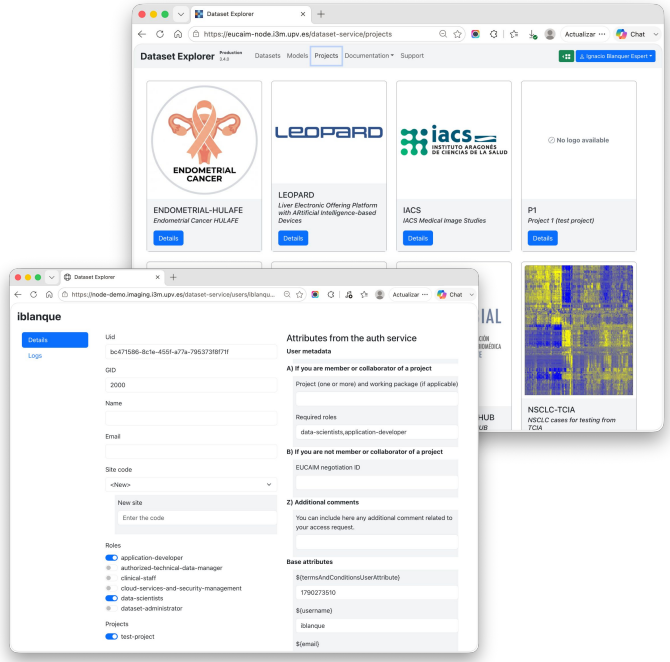
- IMASPE extracts **metadata** directly from DICOM Headers.
- It provides a **landing page** for each dataset's metadata and an **RDF FAIR** Data point for each dataset.
- It can register the dataset's metadata in **ZENODO** and retrieve a **DOI**.



Co-funded by
the European Union

Access control

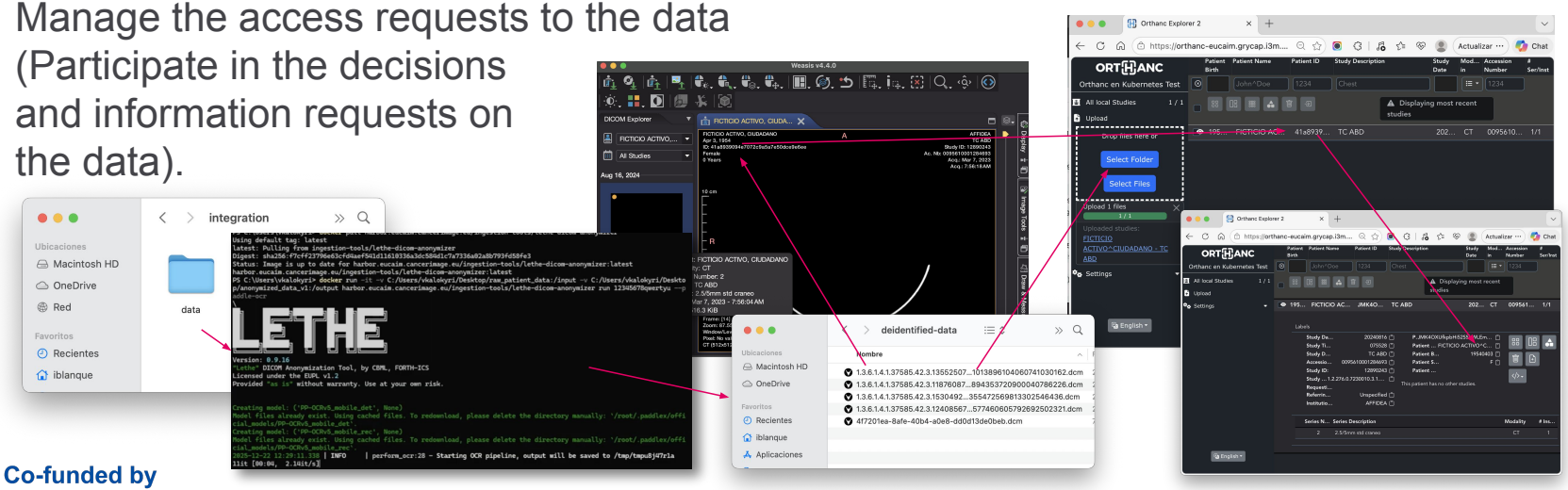
- Based on **Projects, user roles and ACLs.**
- Datasets are created within **projects**.
- Users can belong to **projects** and have different **roles** (upload data, create dataset, access data, access services, admin)
- Users in a project **can access** datasets in status “**Release**” in the project.
- Users **can** access “**Published**” datasets if the Platform admin setups the appropriate **ACLs** (individually at the dataset level).



Co-funded by
the European Union

Data Ingestion

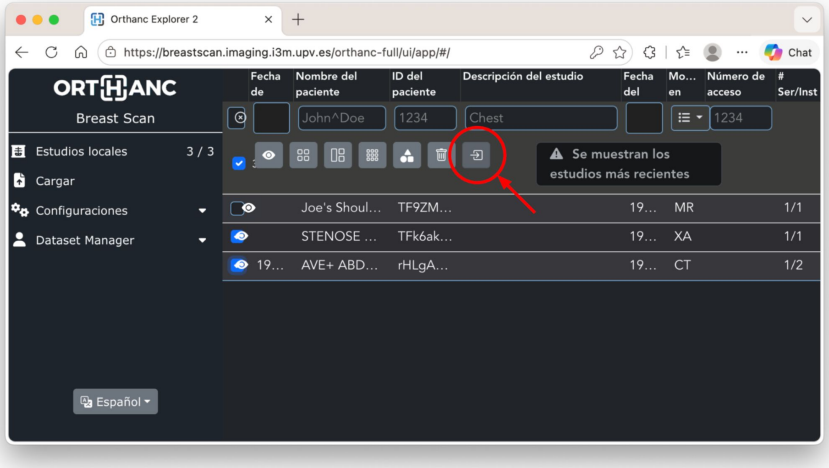
- Prepare the data (Extract, transform and anonymise)
- Upload the data into the federated/reference node (Validate and check anonymisation and quality).
- Create the dataset (Define the collection of data and group them into a citable object)
- Manage the access requests to the data (Participate in the decisions and information requests on the data).

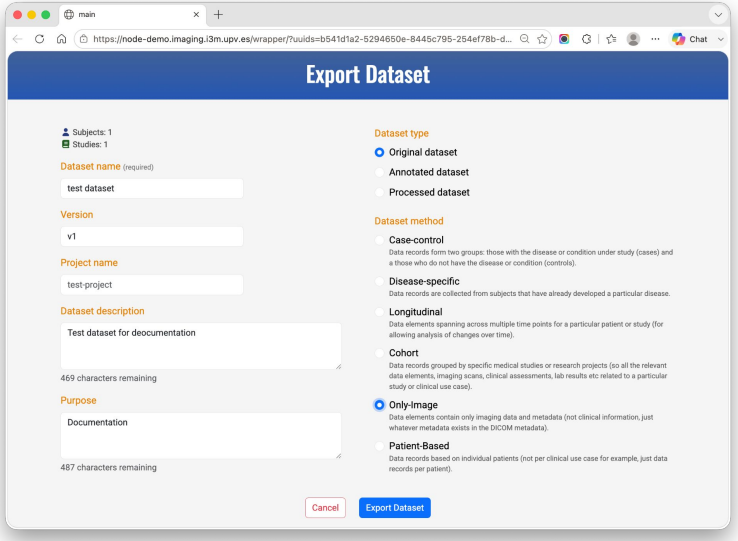


Co-funded by
the European Union

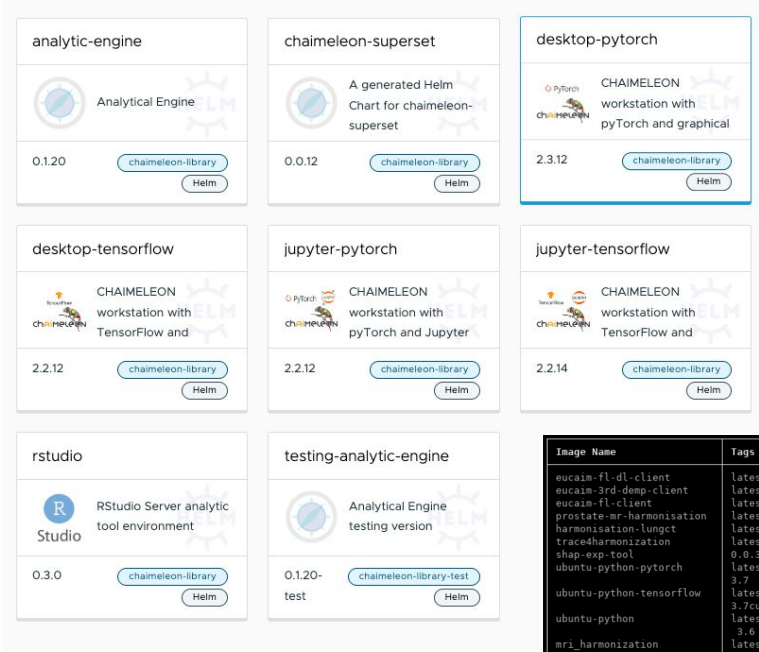
Dataset creation

- A dataset is the basic access unit
  - It comprises several subject's data.
  - Part of the metadata is directly obtained from the data itself, but part of the metadata should be provided in the creation.
  - Dataset service creates the dataset under "Draft" status.





Applications



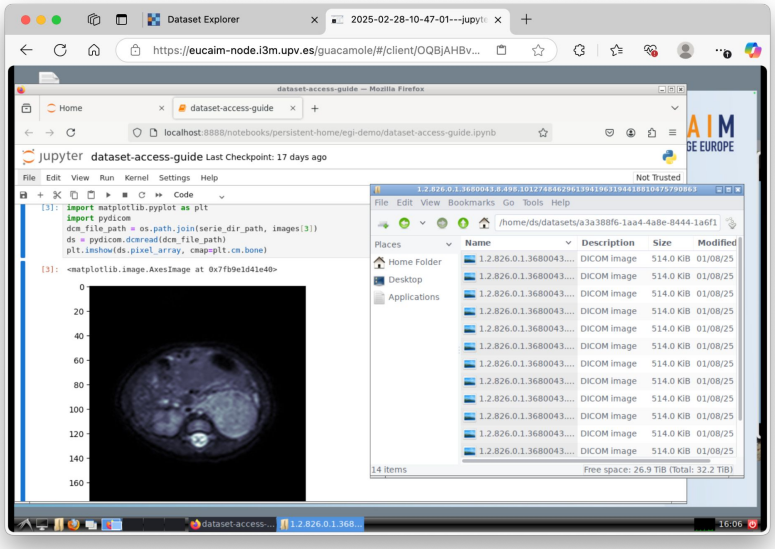
- IMASPE can run Interactive (Desktop) applications and batch jobs.
- Both are based on containers following EUCAIM's guidelines and uploaded on a trusted Container Registry (although a local repository can be used).
- Desktop applications are coded as Helm charts and batch applications are single containers.
  - NextFlow pipelines currently validated, integration in progress.

Image Name	Tags List
eucaim-fl-dl-client	latest 0.2
eucaim-3rd-demp-client	latest 0.2
eucaim-fl-client	latest 0.2
prostate-mr-harmonisation	latest 1.0
harmonisation-lungct	latest 2.1 2.0
trace4harmonization	latest 1.0-2 1.0
shap-exp-tool	0.0.3-testing latest
ubuntu-python-pytorch	latest-cuda 3.10cuda11 3.9cuda11 latest 3.9 3.8cuda11 3.8 3.7cuda11 3.7 3.6cuda11 3.6
ubuntu-python-tensorflow	latest-cuda 3.10cuda11 latest 3.10 3.9cuda11 3.9 3.8 3.8cuda11 3.7cuda11 3.7 3.6cuda11 3.6
ubuntu-python	latest-cuda 3.9cuda11 latest 3.9 3.8cuda11 3.8 3.7cuda11 3.7 3.6cuda11 3.6 3.5cuda11 3.5
mri_harmonization	latest-cuda 1.41cuda latest 1.41 1.40 1.39cuda 1.39 1.38 1.37 1.36cuda 1.36 1.35cuda 1.35 1.34cuda 1.34 1.33 1.32 1.31 1.30 1.29 1.28 1.27 1.26 1.25 1.24 1.23 1.18cuda 1.18-gpu 1.18
privacypreserver	1.0
harmonisation_lungct	1.0
deepfakes-detector	latest 1.0

Co-funded by the European Union

Interactive applications

- Access is limited to the **Virtual Research Environment** and the batch job execution endpoint.
- A Keycloak instance manages the authorisation and eventually the authentication (it can be configured easily to rely on an external IdP).
- An execution endpoint is exposed to enable remote execution through federated or distributed environments.
- No data can be extracted from the environment (it requires the manual intervention of the platform manager).
- No Internet access is allowed from the platform (IPs can be whitelisted, though).

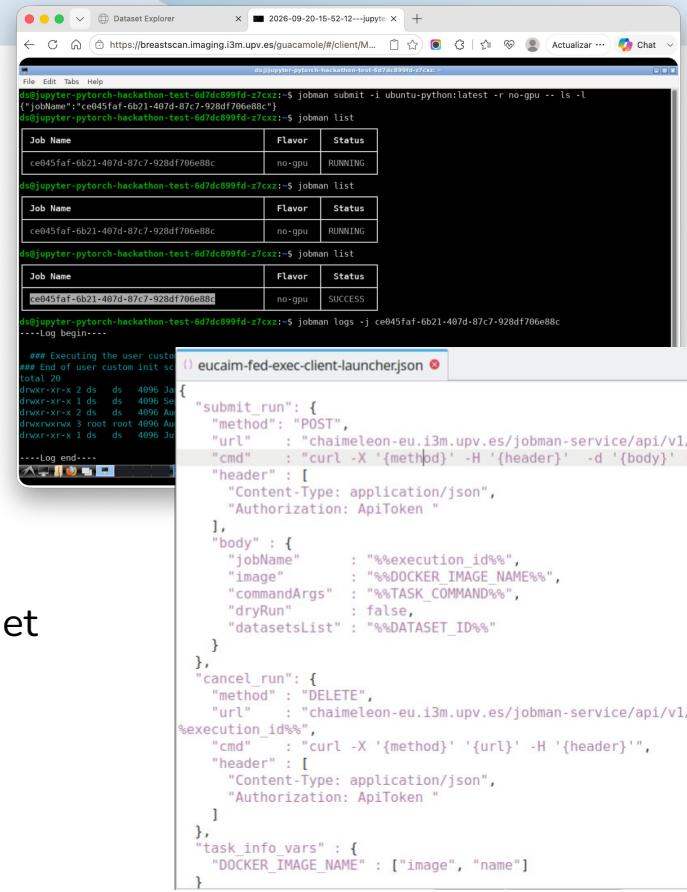


Data cannot be downloaded due to the network policies and the restricted configuration of the access proxy - accesses are traced and preserved.

Co-funded by
the European Union

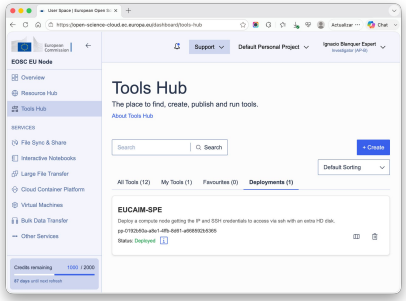
Batch execution

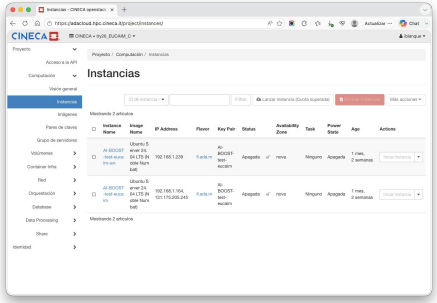
- Jobman v2 - Both a CLI and a web service - offers:
  - Rich interface to manage batch job lifecycle
    - An extended Kubernetes job type that manages volumes automatically.
  - Dual authentication & authorization
    - OIDC and ApiToken (for internal use)
  - jobs management from outside the cluster
    - https://chaimeleon-eu.i3m.upv.es/jobman-service/api/v1/
  - Protection of batch apps (images) contents
    - Users do not have direct access to the jobs
  - Usage through local client and/or tools like curl/wget
- Deployed on EUCAIM at:
  https://eucaim-node.i3m.upv.es/jobman-service/api/v1
- OpenAPI specs at:
  https://app.swaggerhub.com/apis/UPV-CHAIMELEON/Jobman_service/1#/

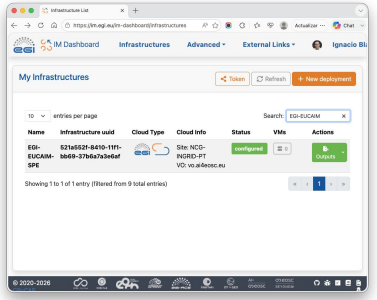


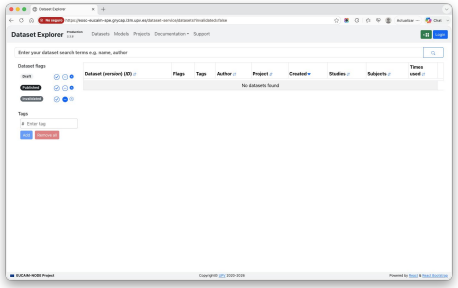
Co-funded by
the European Union

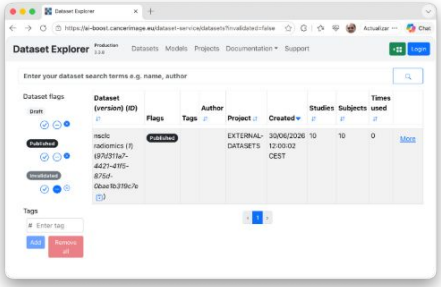
Deployments outside EUCAIM

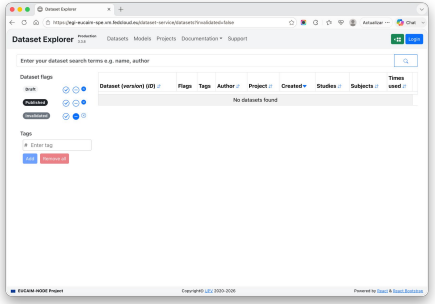












EOSC.eu node

AI-BOOST CINECA

EGI FedCloud

Co-funded by the European Union

Conclusions

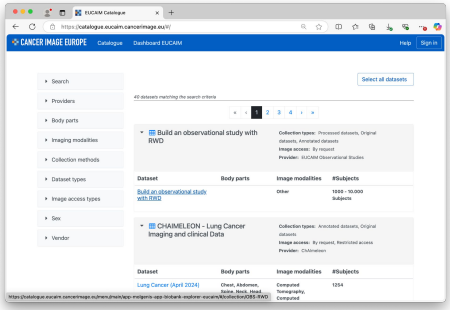
- **EUCAIM's IMASPE provides a secure, federated infrastructure** for cancer imaging data, aligning with EHDS.
- The platform supports both **interactive and batch workloads**, offering flexibility for diverse research needs.
- **Security and compliance are embedded** in the architecture through access control, audit trails, and isolated environments.
- It provides most features needed to fulfil the SPE requirements from TEHDAS2 (D7.4), which are the basis for the national implementation acts of EHDS.
- Missing requirements are on prototypes and planned to be incorporated.
  - Improved auditing of output results, encrypted storage backend, audited workflows.

Co-funded by
the European Union

Interoperability Layers supported

EUCAIM CANCER IMAGE EUROPE

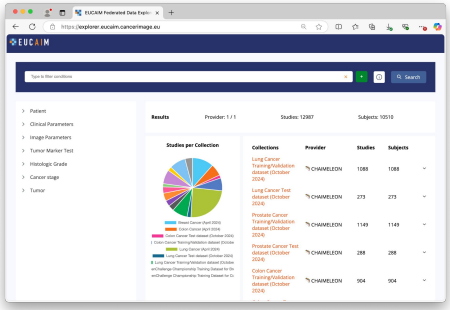
Tier 1: At the Catalogue level



Optional local catalogue, registration in the general catalogue

EUCAIM HealthDCAT-AP

Tier 2: At the level of the Data Discovery

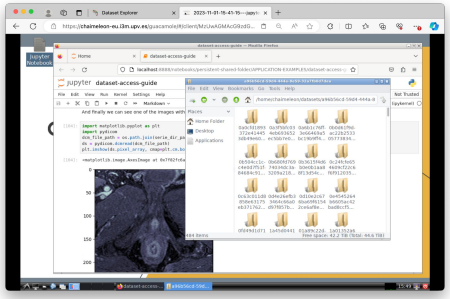


Registration in the federated catalogue

Local deployment of Beam and Focus

Connection to the databases created by the ETL

Tier 3: At the level of the Federated Processing



VRE for local processing through interactive services or batch processing through jobman.

Jobman service to connect to Federated Execution Management.

Co-funded by the European Union