Project title: European Federation for Cancer Images

Project acronym: EUCAIM

Grant Agreement: 101100633

Call identifier: DIGITAL-2022-CLOUD-AI-02

D4.7 First EUCAIM Dashboard

Author(s):	Esther Bron (Health-RI, Erasmus MC), Mark Janse (Health-RI), Ignacio Blanquer (UPV)
Reviewers (WP)	Hanna Leisz (DKFZ) & Irene Marin Radoszynski (HULAFE)
WP2	
WP3	
WP4	
WP5	
Date of delivery:	30/06/2024
Version:	V0.1
Due date:	Month 18
Actual delivery date:	
Type:	Demonstrator
Dissemination level:	Public

Table of contents

Abbreviations

Terms	Definitions
AAI	Authentication and Authorisation Infrastructure
API	Application Programming Interface
DCAT-AP	Data Catalogue vocabulary Application Profile
DICOM	Digital Imaging and Communication In Medicine
EUCAIM	European Federation for Cancer Images
FAIR	Findable, Accessible, Interoperable, Reusable
FDP	FAIR Data Point
Guacamole	A clientless remote desktop gateway
GUI	Graphical User Interface

IdP	Identity Provider
LS-AAI	Life Sciences Authentication and Authorisation Infrastructure
MOLGENIS	A modular web application for scientific data, initially focused on molecular genetics research (molecular genetics information system) but expanded to other disciplines.
Negotiator	BBMRI-ERIC service for structured negotiator for biomedical resources
OpenID	Open standard and decentralised authentication protocol
PACS	Picture Archiving and Communication System
QUIBIM	Spanish company on AI applied to Image Biomarkers
RIS	Radiological Information System
UPV	Universitat Politècnica de València (Valencia University of Technology)
VO	Virtual Organisation, referring to the EUCAIM Virtual Organisation in the LS-AAI

Disclaimer

1. Introduction

This report complements the information presented in the demonstration video (https://dashboard.eucaim.cancerimage.eu/D4-5_7-video.mp4) and jointly constitutes the Deliverable *D4.7 First EUCAIM Dashboard*. The users of EUCAIM access the platform through the Dashboard. The Dashboard integrates all the applications of EUCAIM targeting the data users (researchers) and provides them a portal to manage their requests. The Dashboard includes the landing pages with the instructions and the link to the public catalogue. Through the Dashboard the user can explore the data using the federated search and request access through the negotiator system. Problems can be reported via the Helpdesk. In addition, instructions for other types of users (i.e. data holders, software providers) are included on the landing page.

2. Dashboard Component Description

The Dashboard is a website that integrates the Graphical User Interfaces of the different components in a seamless environment with a common design. This website links to the core services as described in *D4.5 First Federated Core Services*: the Public Catalogue, Federated Search, Negotiator, the Authentication and Authorisation Infrastructure (AAI), Data Population Monitoring and the Helpdesk.

2.1. Functionality

The dashboard integrates a number of applications under a coherent interface. The main set of applications is depicted in Figure 1. This figure shows 6 different workflows:

- In blue, workflows that can be done by an anonymous data user.

- In purple, workflows that can be done by a data holder and a software provider.

- In green, workflows that can be done by an authenticated data user.

- In yellow, workflows related to access and provisioning requests.

- In orange, workflows related to data access.

- In black, workflows related to data processing.

The figure also considers three realms:

- The anonymous public area, accessible by anonymous users.

- The authenticated area, accessible only to authenticated users.

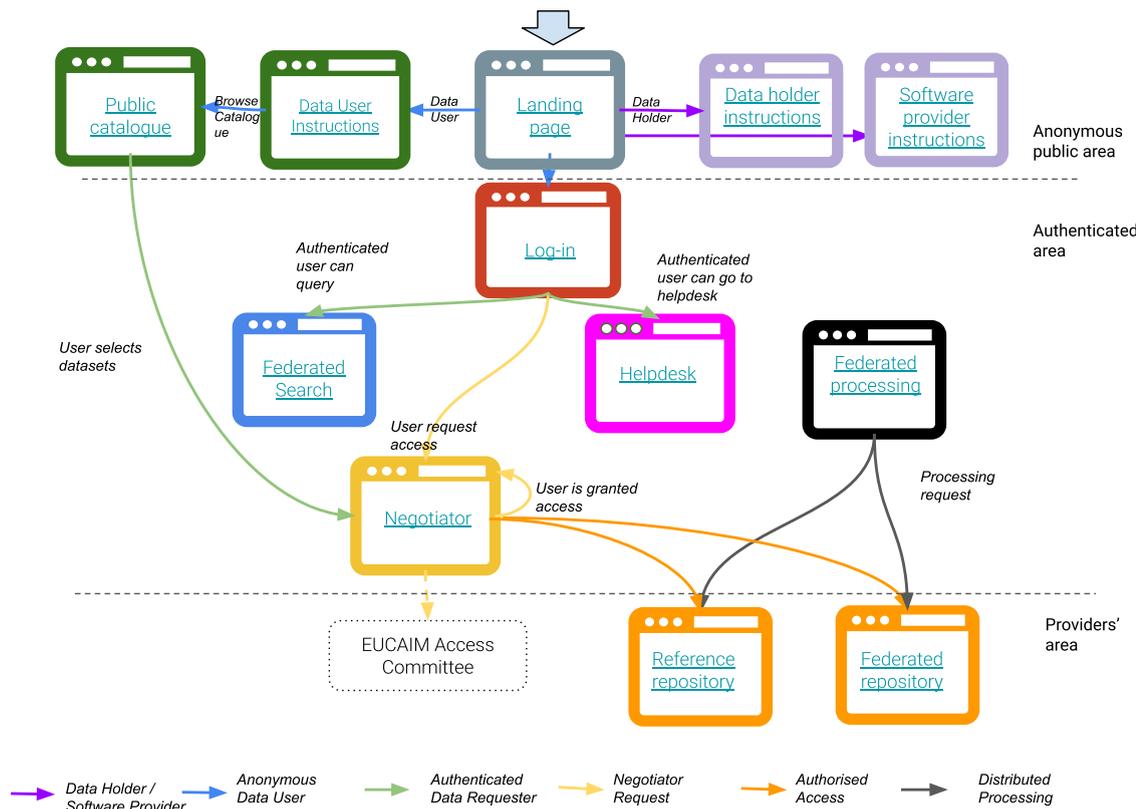- The provider's area, accessible only to users authorised to access the provider's data.

Figure 1: Dashboard workflows in the current implementation.

2.2. Roles supported

The current focus of the dashboard implementation is on data users (e.g. researchers and innovators) who would like to access the data registered in EUCAIM for data analysis activities including the training and validation of AI software. Additionally, the dashboard serves as the entry point for data holders and software providers who would like to contribute to the platform. In this context, the dashboard guides data holders and software providers to the documentation pages explaining the general process required for becoming a data holder in EUCAIM and listing applications provided by EUCAIM, e.g. for data search and request.

This section describes the interactions of the different user roles through the dashboard. However, not all of the user stories, as have been defined in *D4.1 First EUCAIM Operational Platform*, will be implemented through the dashboard. Actions related to the preparation of the local node and the data will need to be performed on premises.

All dashboard interactions are grouped into three main roles (data holder, software provider, data user). A colour coding is used to reference the corresponding workflow in Figure 1. Actions (verbs) are underlined.

1. As a data holder, I want to go to a website to understand the processes involved in data preparation and provision and find instructions. The instructions include the process of registration of datasets in the catalogue and uploading data to the Central Repository or setting up a Federated node. While instructions are there, currently doing this requires switching of platforms as this functionality is not yet implemented

in the dashboard. In addition, seeing information about the usage of datasets is not yet implemented.

2. As a software provider, I want to go to a website to find instructions and descriptions as well as links to supporting tools for interfacing with the repositories. The instructions include information about uploading software to the Marketplace. While instructions are there, currently doing this requires switching of platforms as this functionality is not yet implemented in the dashboard. In addition, seeing information about the usage of software is not yet implemented.

3. As a data user (researcher), I want to go to a website and understand the terms of usage and the access conditions, as well as an overview of the available data.
   a. I want to see the metadata of datasets in a public catalogue.
   b. I want to be able to register and log into the platform.
   c. I want to search the datasets based on specific criteria (disease, imaging modalities, age groups …).
   d. I want to request/negotiate access to that data (if needed and possible).
   e. I want to go to the helpdesk to report problems.
   In addition, the instructions include how to apply processing software to the data that is in a federated or central repository and getting the results. While instructions are there, currently doing this requires switching of platforms as this functionality is not yet implemented in the dashboard. In addition, informing the providers of interesting results obtained thanks to the use of their data for their consideration is not yet implemented.

Based on these user stories, the following functionality is implemented in the dashboard
   1. Registration and login of users. Registration is performed at the Dashboard and login is performed at all the services. (3.b)
   2. Dataset registration is performed at the level of the Catalogue (1)
   3. Data uploading is performed at the level of the reference storages (1)
   4. Information about the usage of datasets (1) and software (2) is performed at the level of the reference storages. (2)
   5. The visualisation of the metadata is provided by the catalogue. (3.a).
   6. The searching of the data is provided in the explorer federated search (3.c), as well as the selection of studies for the request (3.d)
   7. The request of the data as well as the monitoring of the progress of a request is performed through the negotiator. (3.d)
   8. Life-cycle management of the processing will be performed through the distributed processing application (2, 3).

The actions related to the points 1-7 are described in the implementation section. The distributed processing will be described in the deliverables of the WP6 activity.

3. Implementation

The website's main page uses the React framework and is deployed on a Node.js server. A MongoDB database is used to store data. All the components are embedded in Docker containers and run on a managed platform based on Kubernetes. The deployment of the

application is described in the deliverable *D4.9 Central Core Infrastructure set-up*. The source code of the Dashboard is available in GitHub[1].

The Dashboard presents a summary of the imaging datasets present, direct access to the public catalogue and sections for each one of the user roles. Figure 2 shows the landing page of the Dashboard.
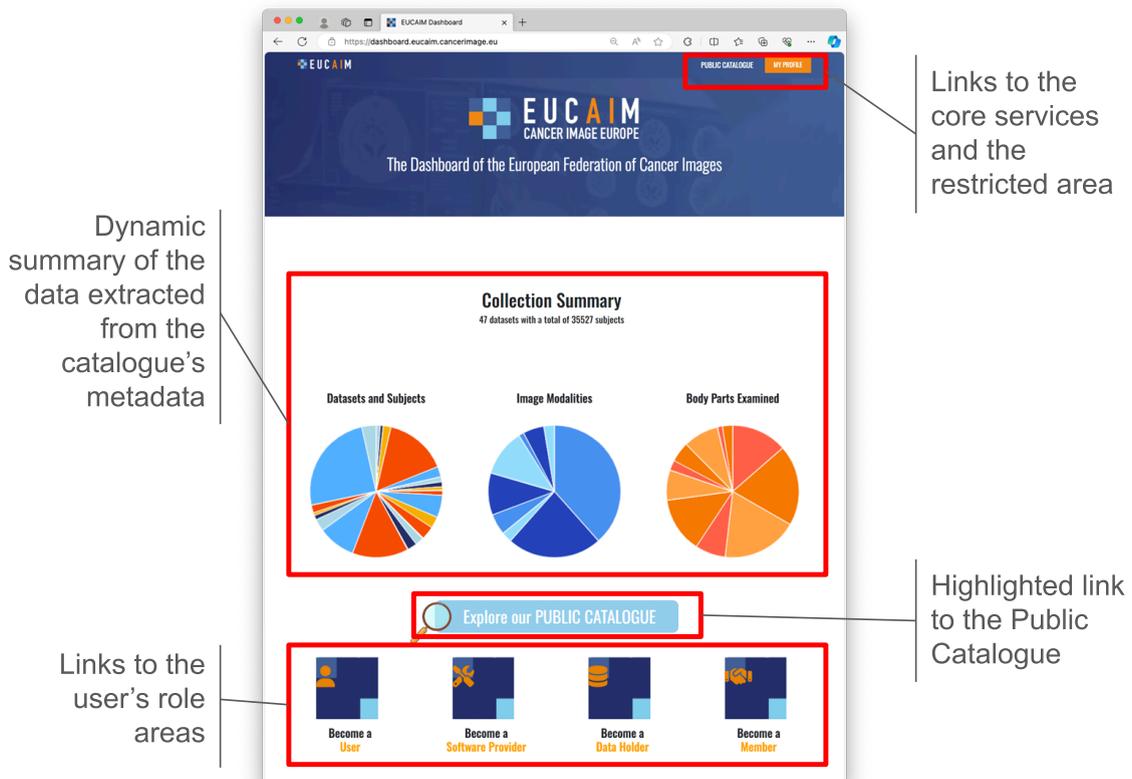

Figure 2: Main components of the dashboard landing page.

The following subsections describe the implementation of each one of the grouped requirements.

3.1. Registration and log-in of users

The registration process is described in the deliverable *D4.13 End-user guide to the system - Annex 1*, which describes the registration and authentication process in detail. In summary, the process involves two steps:

- Creating an LS-AAI account, linked to your institutional account.
- Enrolling into the EUCAIM Virtual Organisation (VO) group, which requires manual intervention by the VO administrator.

Figure 3 shows a few snapshots of this functionality.
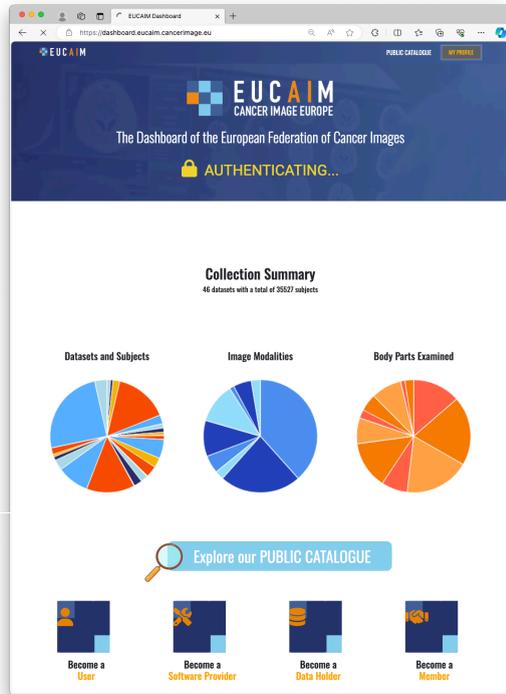
---

[1] https://github.com/EUCAIM/dashboard
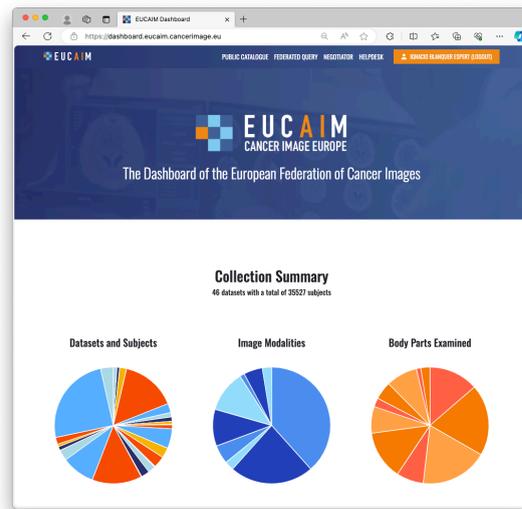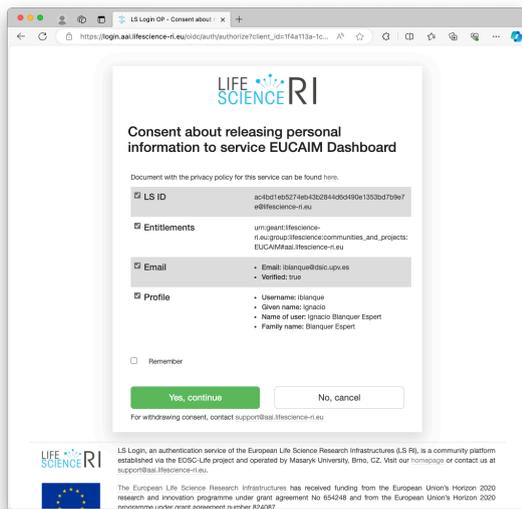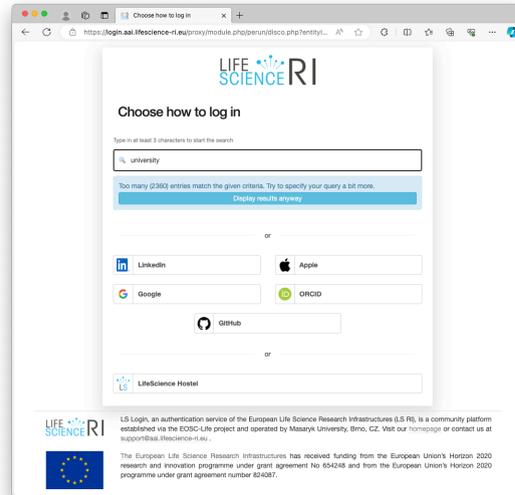
Figure 3: Snapshots of EUCAIM's Dashboard. From top to bottom and left to right: a) anonymous access, clicking on my profile; b) selecting the Identity Provider; c) account information once it has been registered; d) authenticated dashboard (see top right corner).

3.2. Dataset registration

Datasets in the EUCAIM registry can come from trusted federated repositories or the reference node. EUCAIM has set up an initial manual registration process through the catalogue and a proof of concept for the automatic synchronisation of datasets through FAIR data points. Details on the latter are provided in Deliverable *D4.5 First Federated Core Services*. The manual registration requires that the datasets are described using the agreed dataset metadata in EUCAIM as described in the Deliverable *D5.2. The EUCAIM CDM and*

hyper-ontology for data interoperability: initial version and the EUCAIM Hyperontology[2]. The collections are described through a hierarchy that includes three levels:

- The Network, which corresponds to the project, institution or network of institutions in which the dataset has been collected. It does not need to be a juridical person.
- The imaging biobank, representing a logical aggregation of data according to a cancer type, or other relevant grouping criteria. Each imaging biobank is represented by a juridical person that can act on behalf of the data holders of the data. If this mandate cannot be granted to a single actor, it is advisable to create separate biobanks for each one of the data holders.
- The dataset (imaging collection), which constitutes the minimum entity to request access and includes the aggregated metadata of the images and clinical data stored.
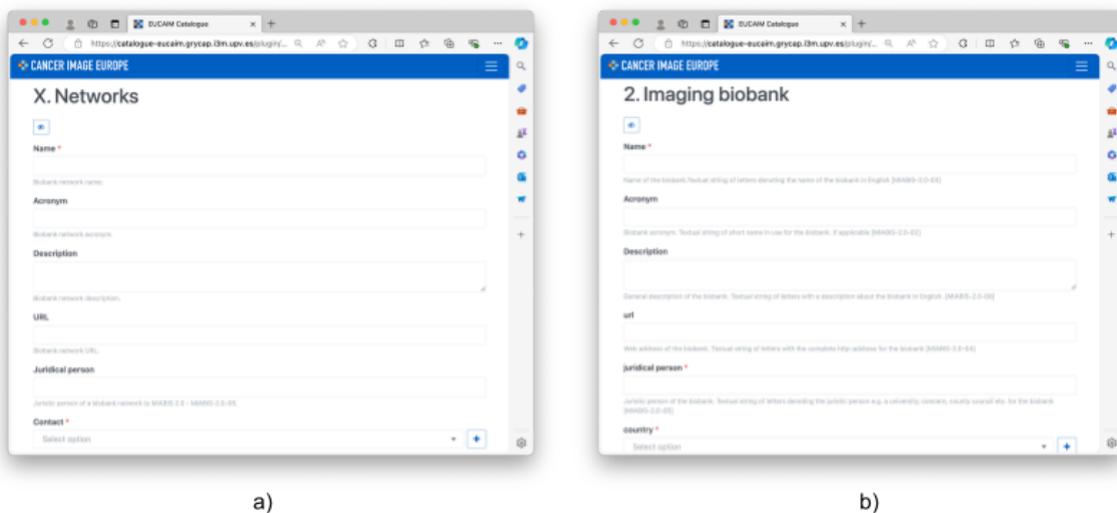
Figure 4. information of a Network (a) and an imaging biobank (b)
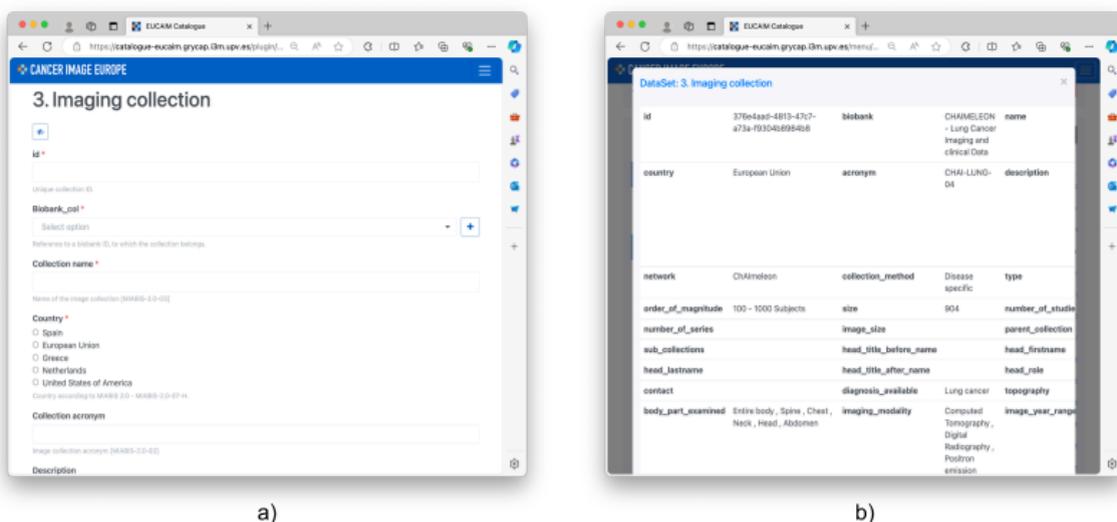
Figure 5. Form for creating an image collection (a) and an example of the information of an existing collection (b).

---

[2] EUCAIM Hyperontology https://zenodo.org/records/11109765

Figure 4 shows the information of a Network and an imaging biobank. Figure 5 shows a snapshot of the form to fill in the metadata and an example of one of the datasets.

3.4. Data access information

The actions performed in the platform are registered in a logging system. This logging system has different levels of information, including in some cases the individual processing and access actions for the users that already have been granted access to a resource. Figure 6 presents the access information records in the case of the UPV reference storage.

Figure 6: Access and processing request records in the UPV reference storage.

3.5. Visualisation of the dataset metadata

A user can visualise and search the metadata in the public catalogue. Figure 7 shows some screenshots as an example. The information can be shown at the level of the Network, the imaging biobank and the dataset.
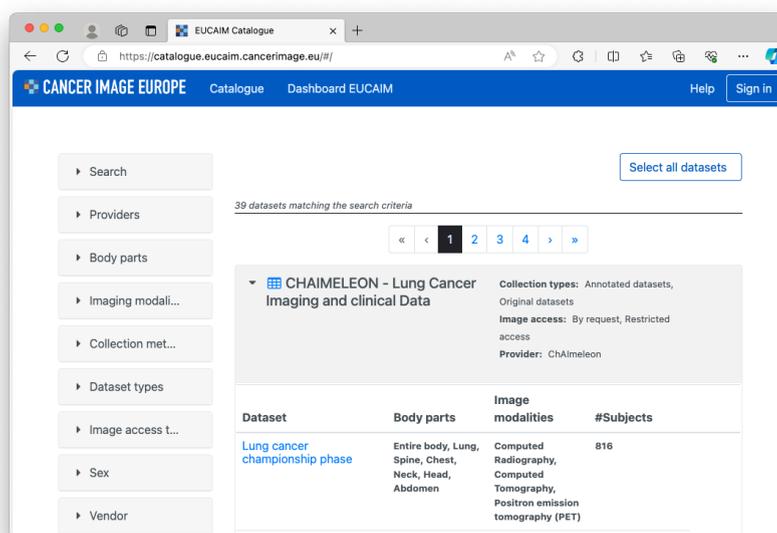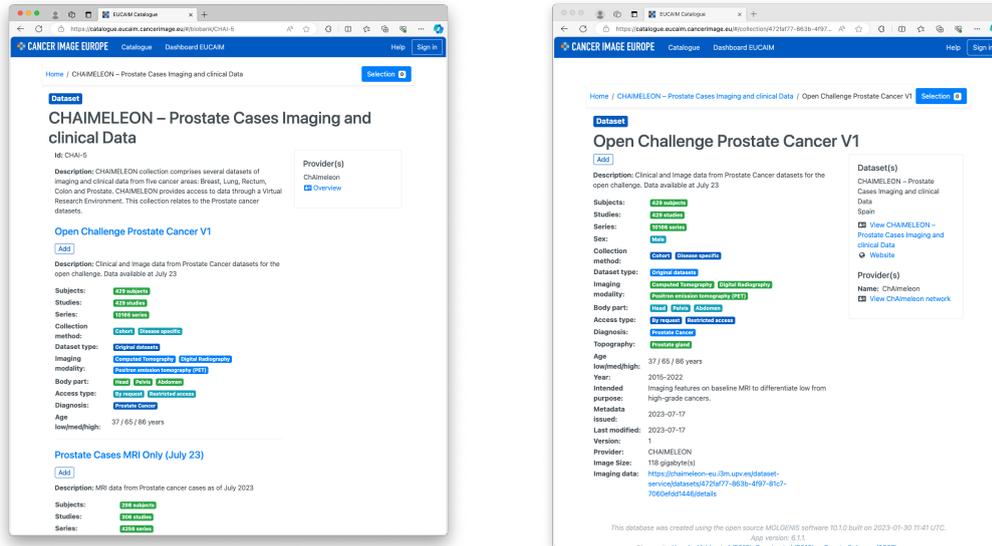
Figure 7: snapshots of the catalogue application.

3.6. Data Searching

More detailed queries on imaging datasets can be performed using the federated search functionality. See screenshots below:
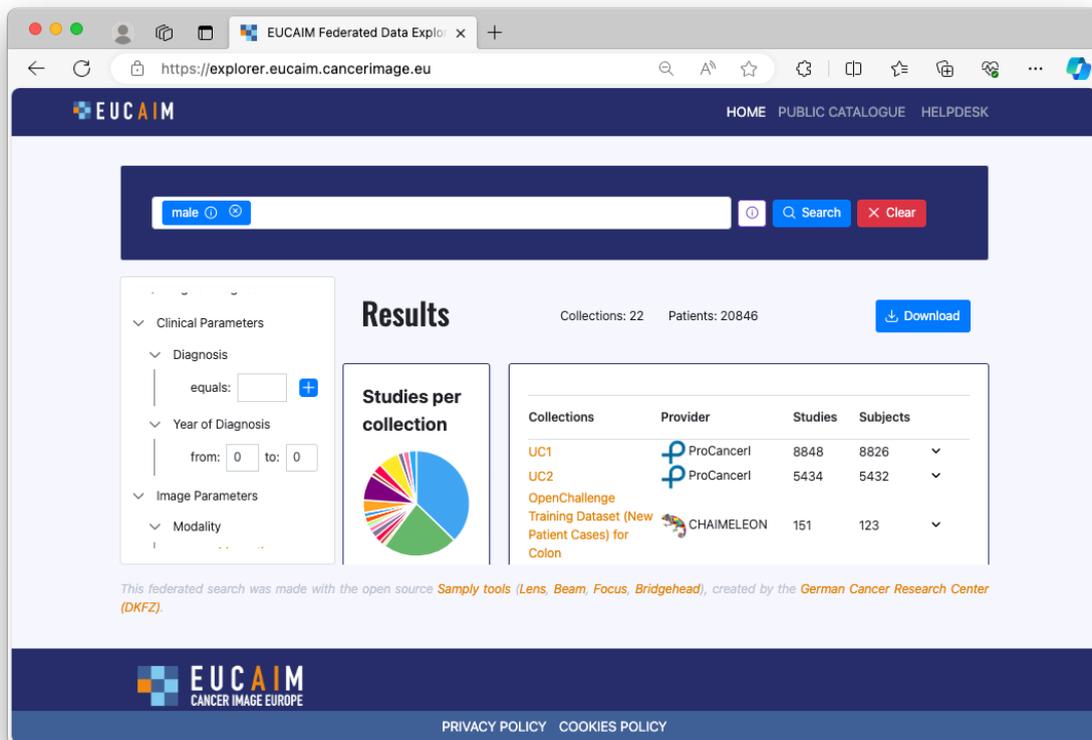

Figure 8: Snapshot of the federated search.

3.7. Data Request

The access to data requires the submission of an application. The application must indicate the access purpose, the needs, the materials and methods, the expected outcome and the

approval of an ethical committee. The description of the items and procedures is described in the Dashboard, in the "become a user" section 2.

The process of access request is sparked from the catalogue, which interacts with the negotiator service for collecting the information. Figure 9 shows the interaction from the catalogue and the dataset access form in the negotiator.
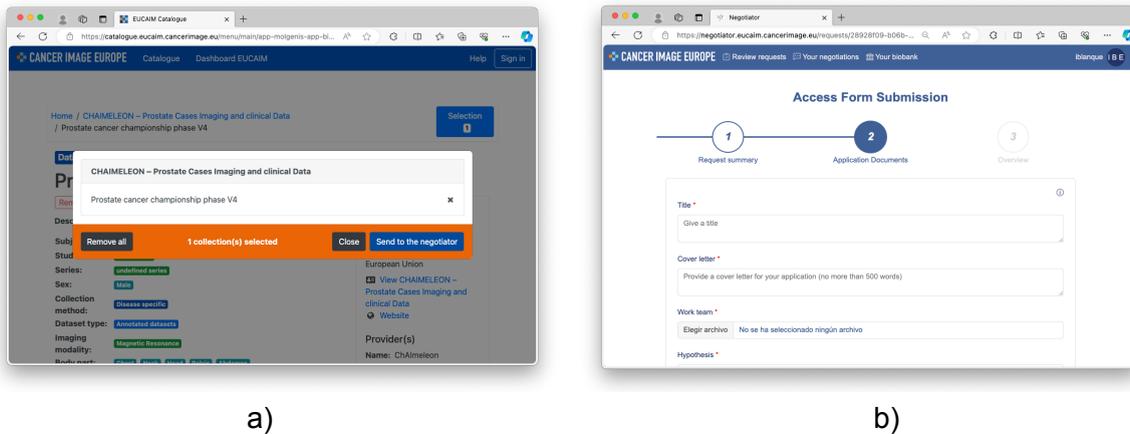

| a) | b) |

Figure 9: Request access to a dataset: a) request initiated from the catalogue, and b) access form.

After logging in and confirming the contents of the request, the data user is asked to enter additional information into the data access form. This information will be used by the access Committee to decide on the request. After filling in all required information, the data user is shown an overview of the request and can submit it. Once submitted, the application enters in the process of evaluation. Figure 10 shows the application submission confirmation and the dataset request status view.
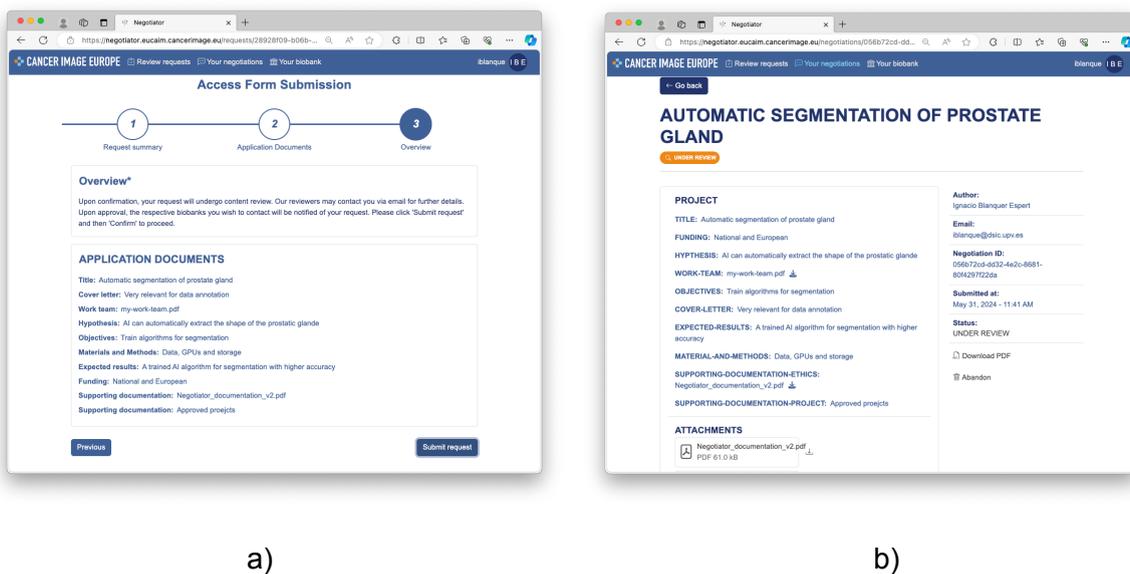

| a) | b) |

Figure 10: Confirmation of the access request submission a) and follow-on b).

In the negotiator dashboard, the contents and status of the request are shown. The process of access request involves three actors: the data user, the EUCAIM negotiator manager, and the image dataset responsible at the level of the access committee or the data holder responsible when needed (i.e. in case the request does not comply with the agreed

licences). These actors check the suitability of the request and the technical viability of the request. Figure 11 shows the data user and the imaging dataset representative views.
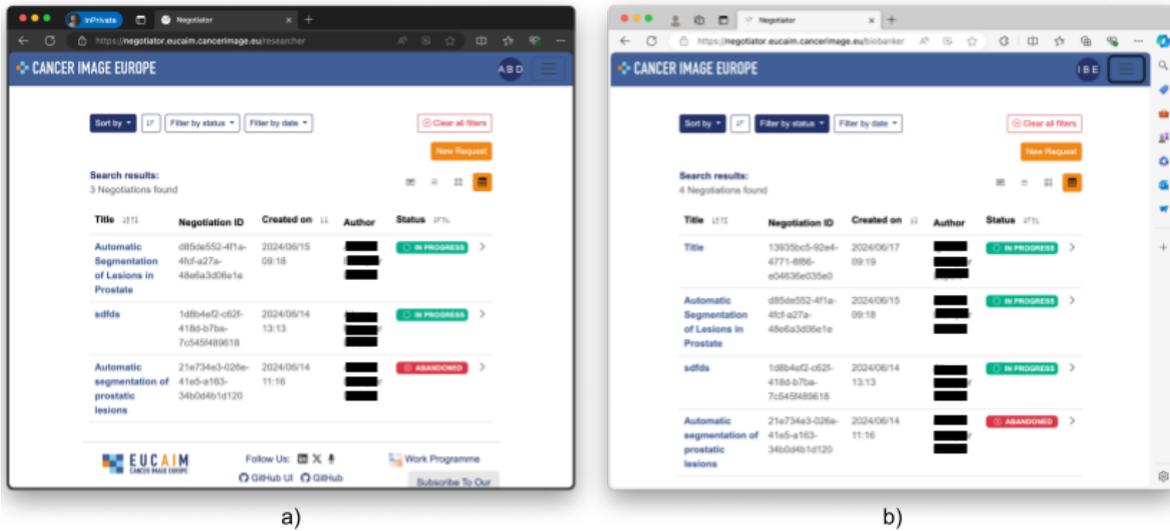

Figure 11: The data requester panel (a) and the imaging dataset responsible panel (b).

The access request lifecycle goes through several stages described in detail in Deliverables *D4.5 First Federated Core Services* and *D4.13 End-user guide to the system*. Figure 12 shows the interactions among statuses and sub-statuses.
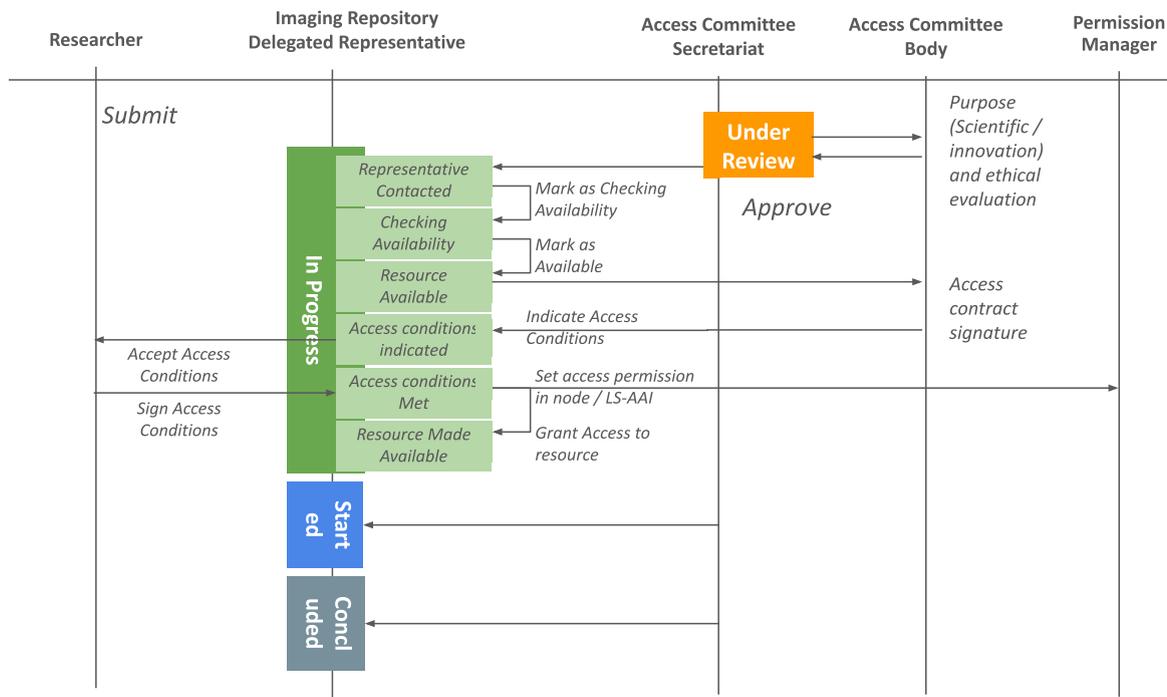

Figure 12: lifecycle of the access request negotiation and the different transitions among different status (solid colours) and sub-status (green-framed white boxes).

4. Demo video

A demo video, combined with the one describing the Core Services has been prepared (https://dashboard.eucaim.cancerimage.eu/D4-5_7-video.mp4). Details on the video are extensively provided in deliverable *D4.5 First Federated Core Services*.