Project title: European Federation for Cancer Images

Project acronym: EUCAIM

Grant Agreement: 101100633

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

D4.5 First Federated Core Services

Author(s):	Ignacio Blanquer (UPV), Alexander Harms, Mahlet Birhanu (Erasmus MC), Esther Bron (Erasmus MC, Health-RI), Mark Janse (Health-RI), Diego Silveira (ITI), Tobias Kussel, Enola Knezevic (DKFZ).
Reviewers (WP)	Paris Laras (MAG), Diego Silveira (ITI)
WP2	
WP3	
WP4	
WP5	
Date of delivery:	30/06/2024
Version:	V0.3
Due date:	Month 18
Actual delivery date:	
Type:	Demonstrator
Dissemination level:	Public

Table of contents

Abbreviations and terms

Terms	Definitions
AAI	Authentication and Authorisation Infrastructure
AARC	Authentication and Authorisation for Research and Collaboration
AI4HI	Artificial Intelligence for Health Imaging

API	Application Programming Interface
Auth	Authentication
BBMRI-ERIC	Biobanking and Biomolecular Resources Research Infrastructure - European Research Infrastructure Consortium
DBMS	Database Management Service
DCAT-AP	Data Catalogue vocabulary Application Profile
EduGain	Global interfederation service that interconnects multiple identity federations
ELK	Elasticsearch, Logstash, Kibana
EUCAIM	EUropean Federation for CAncer IMages
FAIR	Findable, Accessible, Interoperable, Reusable
FDP	FAIR Data Point
GDPR	General Data Protection Regulation
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
OHIF Viewer	Open Health Imaging Foundation Viewer
PACS	Picture Archiving and Communication System
POSIX	Portable Operating System Interface
Postgres	Object-oriented relational Database management system
RIS	Radiological Information System
SU	Support Unit
Vault	Hasihcorp identity-based secrets and encryption management system
VO	Virtual Organisation
VRE	Virtual Research Environment
WP	Work Package
XNAT	Extensible Neuroimaging Archive Toolkit

Disclaimer

The opinions stated in this report reflect the opinions of the authors and not the opinion of the European Commission.

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 *D4.5 First Federated Core Services*. EUCAIM Core services are those that support the main functionalities of the federation, and comprise: the Authentication and Authorisation, public catalogue, federated search, data access request, monitoring and helpdesk. Additional core services will be made available for data transfering, distributed processing and the storage of data.

The document briefly describes those components, updating the descriptions provided in deliverables *D5.1 Early release of the Data Federation Framework* and *D4.3 First rules for participation report*. The deliverable is not an extensive description of the architecture, which is kept continuously updated in an online document[1].

This deliverable is complemented by four other deliverables:

- *D4.7 First EUCAIM Dashboard*, describing the interoperability of the Dashboard.
- *D4.9 Central Core Infrastructure set-up*, which describes the resources and recipes used to deploy the core services.
- *D4.10 Technical evaluation of the Platform*, describing the validation of the services at the technical and user-level.
- *D4.13 End-user guide to the system*, a user manual of the core services.

2. Updated Architecture

Although infrastructure components have been continuously updated and improved, no major conceptual changes have been applied to the original architecture, although components have been continuously updated and improved. As before, the EUCAIM architecture focuses on a federated model in which nodes connect to the central core services, while still maintaining a reasonable degree of independence and autonomy. The technical requirements for the integration of local nodes with in the federated infrastructure relies on the model of tiers:

- Tier 1: Compliance with the metadata model for the datasets.
- Tier 2: Direct (through adoption) or indirect (through a mediator component) compliance with the data model for **searching** purposes.
- Tier 3: Direct (through adoption) or indirect (through a mediator component) compliance with the data model for **processing** purposes.

Figure 1 shows a simplified diagram of the architecture explaining the interactions among the components according to the tier model.

[1] EUCAIM Architecture: https://eucaim.gitbook.io/architecture-of-eucaim/

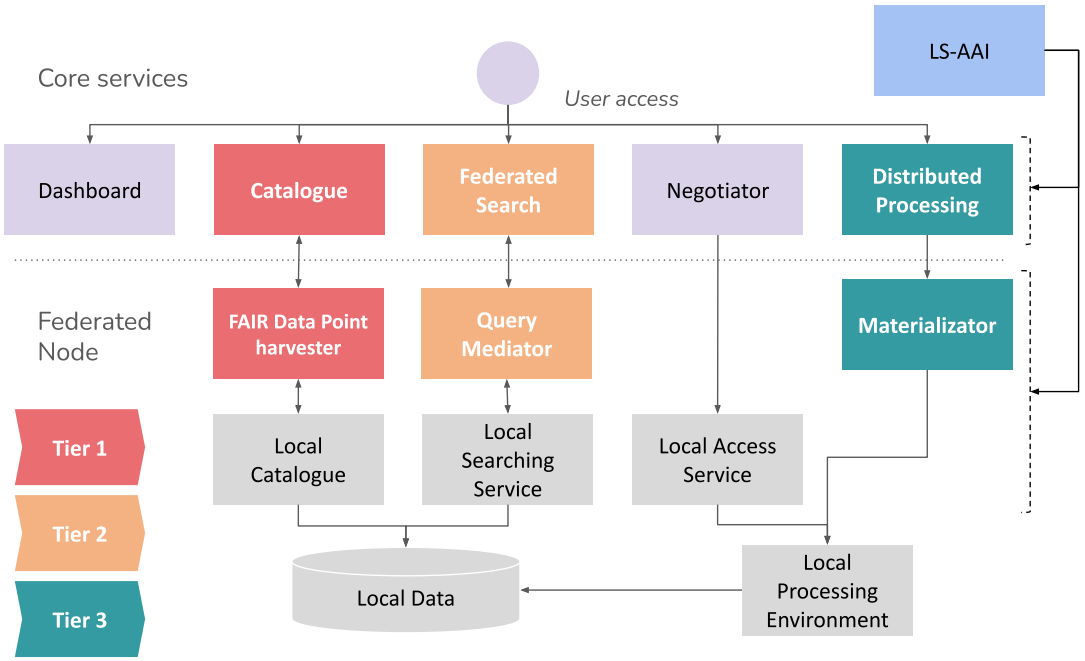


Figure 1: Simplified architecture diagram of EUCAIM.

The core services for the federation are shown in the upper part of the figure. The lower part represents a node which connects to the federation. The three tier levels described above are related to the following federation concepts:

- Tier 1: The datasets hosted by the federated node are registered in the central catalogue. Ideally, this is done through the exposure of FAIR Data Points that are used by the central catalogue to harvest the dataset's metadata.
- Tier 2: The data of the federated node is searchable through its local searching service, which is queried by the federated search system through a Query Mediator component that transforms the query from EUCAIM's model to the local model and vice-versa for the results.
- Tier 3: The federated node deploys a materialisation component that makes the data available to the federated processing, according to EUCAIM's model.

Components in grey relate to the node's pre-existing local services and components. Components in purple are tier-independent components of EUCAIM core services, and the LS-AAI component is an external entity that manages the common authentication model.

The integration in the federation of a node requires the data holder organisation to:

- In every case, LS-AAI should be supported to facilitate the seamless authentication of users. Authorisation policies are defined at the level of the local node.
- Tier 1: Report the datasets to be registered in the central catalogue, once adapted to the EUCAIM DCAT-AP specific definition. Ideally, expose them as a FAIR Data Point so it can be automatically harvested by the catalogue.

- Tier 2: Develop and deploy a query mediator component for the translation of the federated queries from the core services. An example is provided here[2]. The EUCAIM Technical Support Team will prepare guidelines, specifications and examples to facilitate the development of adaptations.

- Tier 3: Develop and deploy a materialiser component that makes the data accessible as a POSIX volume, following the model of EUCAIM. An example is provided here[3]. Similarly as above, the EUCAIM Technical Support Team will give support to facilitate the development of specific materialiser components.

3. Description of the Components

This section briefly describes the functionality and structure of each one of the components of EUCAIM. As described in deliverable *D4.9 Central Core Infrastructure set-up*, EUCAIM has deployed a production and a development environment to test new versions of core services and developments before rolling them into the production version.

3.1. Authentication and Authorisation Infrastructure (AAI)

Some services in EUCAIM allow anonymous access (Dashboard and the Catalogue). Those services provide access to general information, onboarding processes and aggregated data. Fine-grained searching and gaining data access require authentication and authorisation.

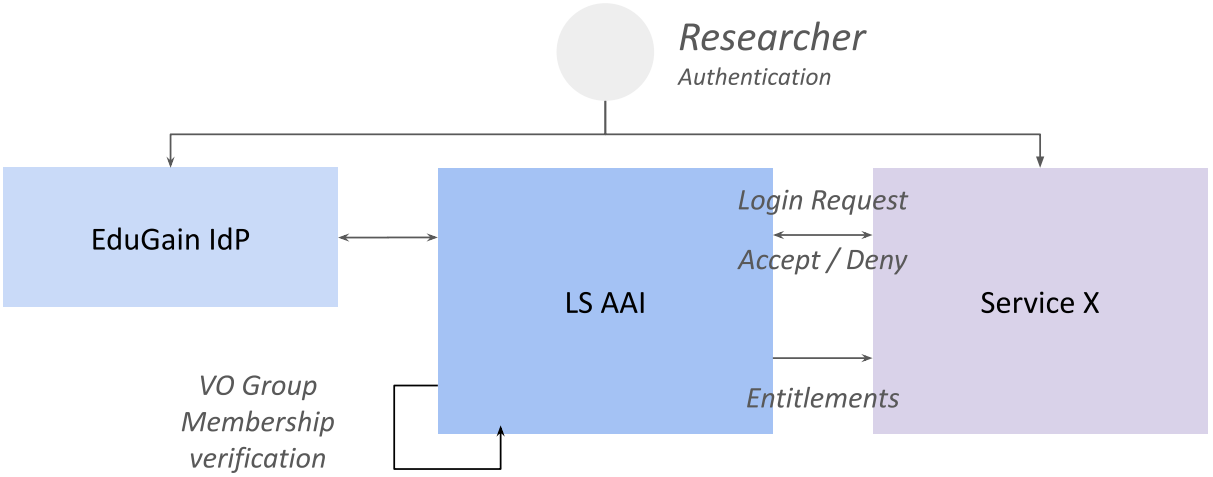


Figure 2: Basic interaction of the Core Services with the LS-AAI.

As shown in Figure 2, each service interacts with the LS-AAI endpoint which delegates the authentication to the institutional IdPs. Along with the acceptance of an authentication request, the LS-AAI service performs the validation of the membership of the users to the EUCAIM VO Group). The LS-AAI service also returns the Entitlements (additional authorisation attributes) of the user including the VO groups membership and roles, which are used in the authorisation process by the Service.

Authentication and Authorisation in EUCAIM relies on the Life Sciences AAI[4]. LS-AAI is the commonly agreed AAI framework for Life Sciences Research Infrastructures. It relies on the

[2] https://github.com/search?q=repo%3Achaimeleon-eu%2Fdataset-service%20eucaim&type=code
[3] https://github.com/chaimeleon-eu/jobman/tree/521f1205338a9636e9500ebc0bbd6dac3370036d
[4] https://lifescience-ri.eu/ls-login/

AARC blueprint[5] and supports the eduGAINFederation (which serves most academic and research organisations in Europe), as well as other public Identity Providers. Authentication is performed through the membership to the EUCAIM VO Group[6]. The process of creating an account and requesting the membership to the EUCAIM VO group is described in the dashboard site[7] and involves the manual verification of valid credentials.

The following services are registered in the LS-AAI:

- Dashboard, deployed in dashboard.eucaim.cancerimage.eu.
- Federated Search, deployed in explorer.eucaim.cancerimage.eu.
- Negotiator UI, deployed in negotiator.eucaim.cancerimage.eu.
- Negotiator backend, deployed in negotiator.eucaim.cancerimage.eu/api.
- Helpdesk, deployed in helpdesk.eucaim.cancerimage.eu.
- Reference Node at UPV, deployed in eucaim-node.i3m.upv.es.

Services in the development platform are authenticated through the same services to facilitate the roll-out of services into production. Figure 3 shows the management dashboard for the EUCAIM VO Group and the EUCAIM core services.

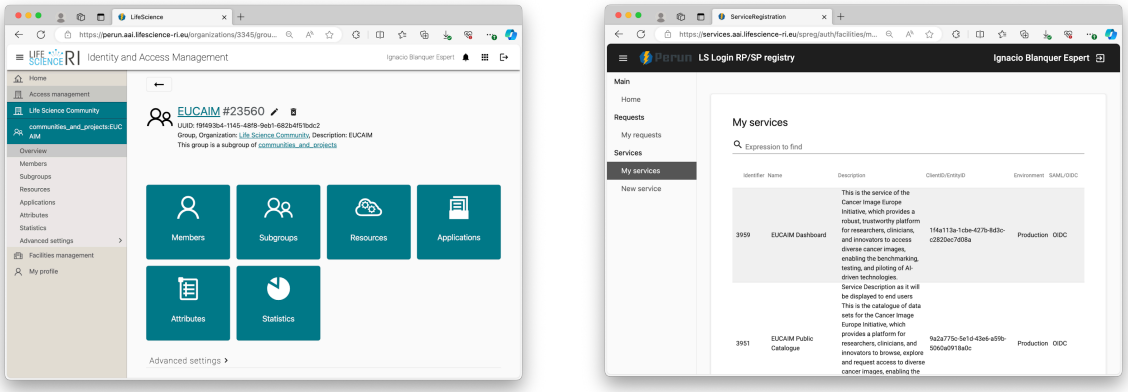


Figure 3: Snapshot of the management console of the LS AAI: VO-management panel[8] (left) and service management section[9] (right).

3.2. Public Catalogue

Data discovery is the essential first step for data re-use. The public catalogue stores the metadata of data sets, and offers the researchers descriptive information about the available datasets, while displaying data characteristics as well as access conditions. In parallel, the catalogue offers a platform for data owners to display their datasets.

[5] https://aarc-20240213064542.local/wp-content/uploads/2019/05/AARC2-DJRA1.4_v2-FINAL.pdf
[6] Enrollment URL for the EUCAIM VO Group https://signup.aai.lifescience-ri.eu/fed/registrar/?vo=lifescience&group=communities_and_projects:EUCAIM
[7] User's registration process in EUCAIM https://drive.google.com/file/d/1EsFYxbzqpyYKggyeKrKKw3FkVecDby8P/view
[8] EUCAIM VO management: https://perun.aai.lifescience-ri.eu/organizations/3345/groups/23560
[9] EUCAIM Services management page in LS-AAI: https://services.aai.lifescience-ri.eu/spreg/auth/facilities/myServices

The metadata catalogue consists of the Molgenis 10.1 platform[10] as a back-end service with a custom Javascript front-end which is based on prior catalogues[11]. The catalogue lists the datasets registered in the platform grouped into dataset series. The current implementation of the catalogue offers various options to filter the datasets based on the supplied metadata. By grouping associated datasets in dataset series, and dataset series by provider, the catalogue allows an easy overview of related datasets. The whole architecture of the Molgenis platform has been minimised to the components relevant for EUCAIM. Figure 4 shows the four components involved: the Molgenis front-end and backend, an Elasticsearch component for indexing data and a Postgres database for the persistence of all the information.

The metadata catalogue offers an API[12] through the Molgenis platform that facilitates the querying of the metadata in the catalogue. Just like the metadata which are made publicly accessible in the GUI, this information is also made accessible through the API.

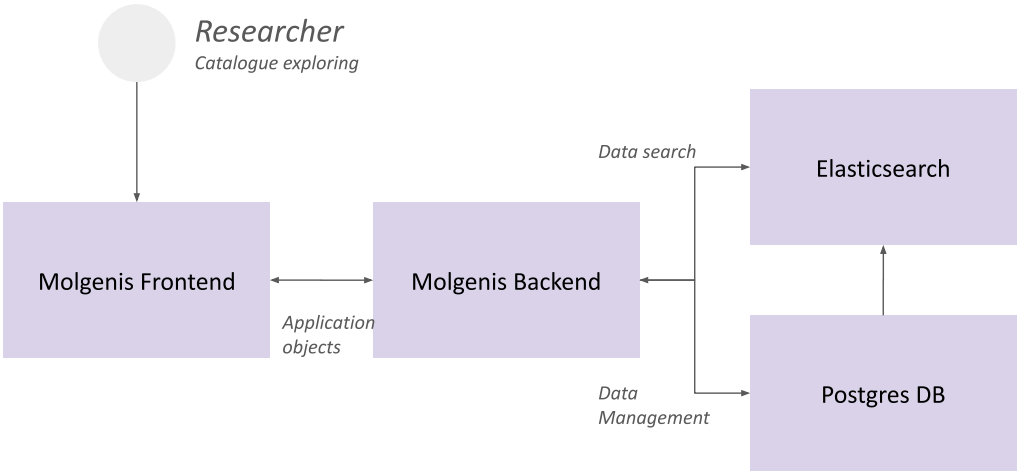


Figure 4: Snapshot of the management console of the LS-AAI: VO-management panel (left) and service management section (right).

Authenticated users belonging to the EUCAIM VO Group have access to additional functionalities like registering datasets and updating ontologies and controlled vocabularies. Apart from these, there is also a group of selected users who have administrative privileges.

In addition to storing public metadata, this public catalogue is being designed to provide main functionalities of exposing its metadata contents as well as continuous ingestion of other publicly available metadata.

To allow the dissemination of dataset metadata into multiple catalogues, without going through the trouble of repeatedly registering the datasets, the FAIR Data Point (FDP) protocol[13] is used to connect the catalogues. The FDP protocol uses the DCAT vocabulary and the DCAT-AP application profile. Through an FDP endpoint that is exposed from the metadata catalogue, other catalogues can harvest the metadata in a standardised format. The initial implementation of FDP architecture on Molgenis fairly mimics the standard FDP

[10]van der Velde, K. Joeri, et al. "MOLGENIS research: advanced bioinformatics data software for non-bioinformaticians." Bioinformatics 35.6 (2019): 1076-1078. (https://doi.org/10.1093/bioinformatics/bty742)
[11]EIBIR catalogue: https://molgenis.eibir-edc.org/#/, BBMRI catalogue: https://directory.bbmri-eric.eu/#/catalogue
[12] API: https://catalogue.eucaim.cancerimage.eu/api/v2
[13] Fair Data Point specification: https://specs.fairdatapoint.org/fdp-specs-v1.2.html

specifications. However, this is designed to be extendible to other DCAT-AP profiles since many standard profiles are in use in the field (DCAT-AP health extension, etc.).

To allow metadata from other FDP endpoints to be added to the EUCAIM metadata catalogue, a proof-of-concept tool[14] to harvest this information and use it to register the datasets in other catalogues was developed. Linking together catalogues through exposing metadata through and harvesting metadata from FDP endpoints will create a federation of catalogues, allowing datasets to be findable far beyond the catalogue where they were initially registered.

The catalogue is connected to the Federated Search application and to the access request service (negotiator). A user can request access to a dataset by accessing the proper option in the catalogue and sending the request to the negotiator.

Recently the catalogue metadatamodel was updated to incorporate the tier 1-2-3 information. The update required for this in the catalogue interface is pending and will be implemented for M24.

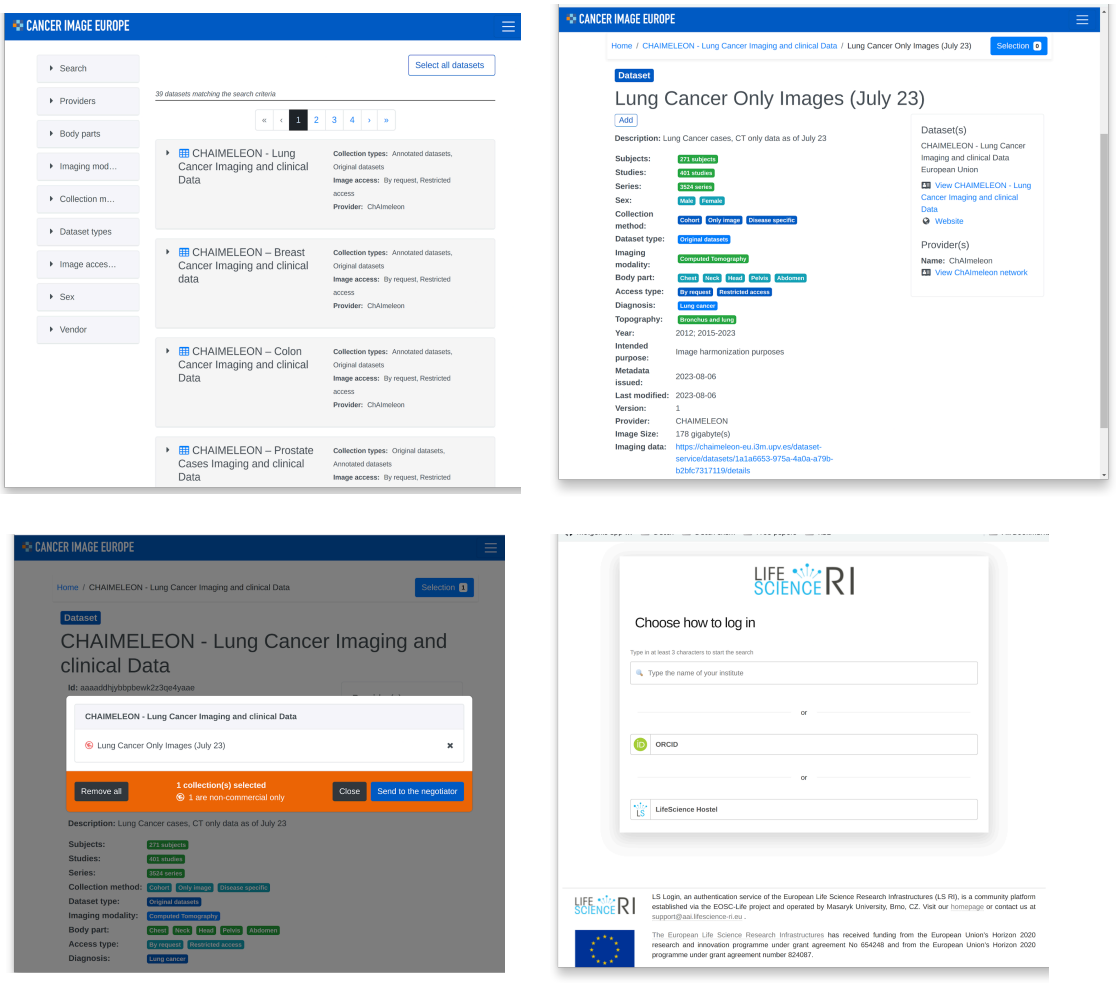


Figure 5: Four snapshots of the public catalogue showing a) the general level[15], b) specific collection information[16], c) data access request panel, and d) access by a privileged user.

[14] https://github.com/Health-RI/molgenis-fdp-harvester
[15] https://catalogue.eucaim.cancerimage.eu/
[16] https://catalogue.eucaim.cancerimage.eu/#/collection/1a1a6653-975a-4a0a-a79b-b2bfc7317119

3.3. Federated Search

Federated search enables users to retrieve the number of subjects that fulfil a specific criteria. The federated search considers two realms: The **central core services**, which consist of the front-end, its back-end, the federated query brokering system, and the certificate storage; and, on the providers' side, the **query dispatcher**, the store, and the data holders customised components to translate the query into the local format.

The central core services are the following:

- Lens - the front-end application and a set of components that propagate the search queries to different providers in the federation, and display the search results.
- Spot - Lens backend creating a task containing the query from Lens and sending it to Beam using Beam Proxy.
- Beam Proxy - handling communication with Beam Broker, taking care of authentication, encryption, and signatures.
- Beam Broker - distributed task broker.
- Vault - used to store the credentials for accessing the searching endpoints at each data holder registered in the system.

The data-holder side integrates the following components:

- Beam proxy - handling communication with Beam Broker, taking care of authentication, encryption, and signatures.
- Focus query dispatcher - receiving Beam tasks using Beam Proxy, translating queries depending on the types of endpoints, running them, and returning the results to Beam.
- Different stores (DBMS) and custom components translating the queries.

An implementation of the mediator component for connecting CHAIMELEON data holder has been integrated in CHAIMELEON Dataset service. The implementation can be found in the Github repository[17]. ProCancer-I has also implemented a mediator for the datasets available from this project.

Figure 5 shows the architecture diagram of the above components and the interactions with the data holders.

[17] https://github.com/chaimeleon-eu/dataset-service?tab=readme-ov-file#integration-with-eucaim-federated-search

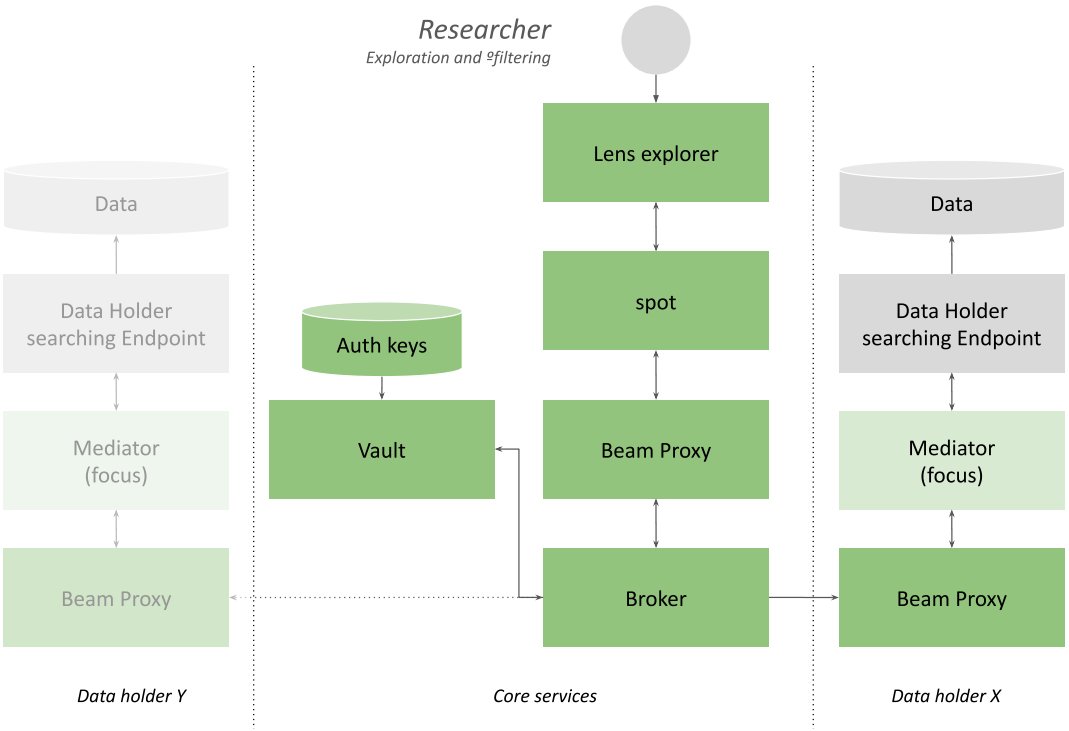


Figure 5 *Architecture of the Federated Search system.*

Details on the end-user functionality of the Federated Search are provided in deliverable *D4.13 End-user guide to the system*.

3.4. Negotiator

Access requests to sets of data are collected through the negotiator. This is a service that collects the information about a data access request, which should be evaluated by the Access Committee.

Access requests are always triggered through the catalogue. The user selects the set of datasets of interest and sends the request to the negotiator. Then, the negotiator presents a dynamic form that could depend on the type of the dataset and creates the full request. Further accesses can be performed directly on the negotiator to follow-on (or evaluate) the requests. In the case of observational studies, a “customisable” dataset should be selected, leading to a specific access form.

The negotiator comprises three services: the **front-end**, which builds interfaces according to the requests, the **backend service**, which exposes the API of the negotiator’s functionality. Finally, a **postgres database** persists the specifications of the access form and the information regarding special privileged roles for each dataset. The negotiator periodically collects this information from the catalogue. Figure 6 shows the interaction among the different components. All negotiations start from the user of the catalogue which interacts through the “export” method with the Negotiator under the user “directory”.

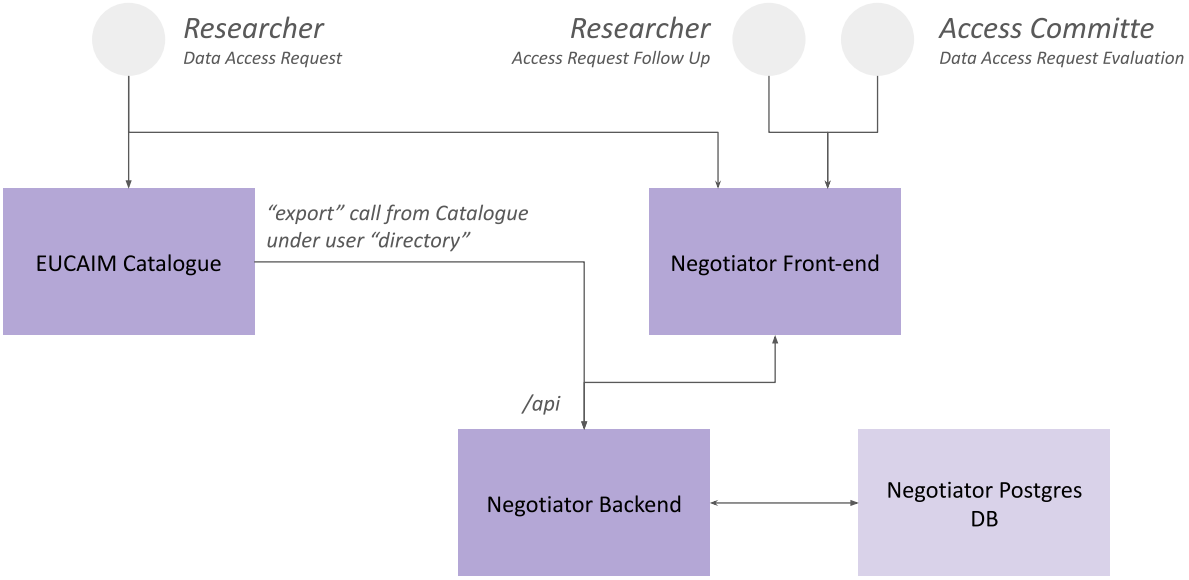


Figure 6*:* Architecture of the Negotiator Service.

3.5. Monitoring

The monitoring task enables knowing the status of the different EUCAIM components by making requests to the associated web services at certain time intervals. In this way, it is possible to know the status of the latest checks carried out for each service, as well as other important aspects such as the remaining time until the expiration of the website's TLS certificate. All this information can be consulted in a dashboard provided by Kibana in the **Uptime Monitors** section (figure 7).

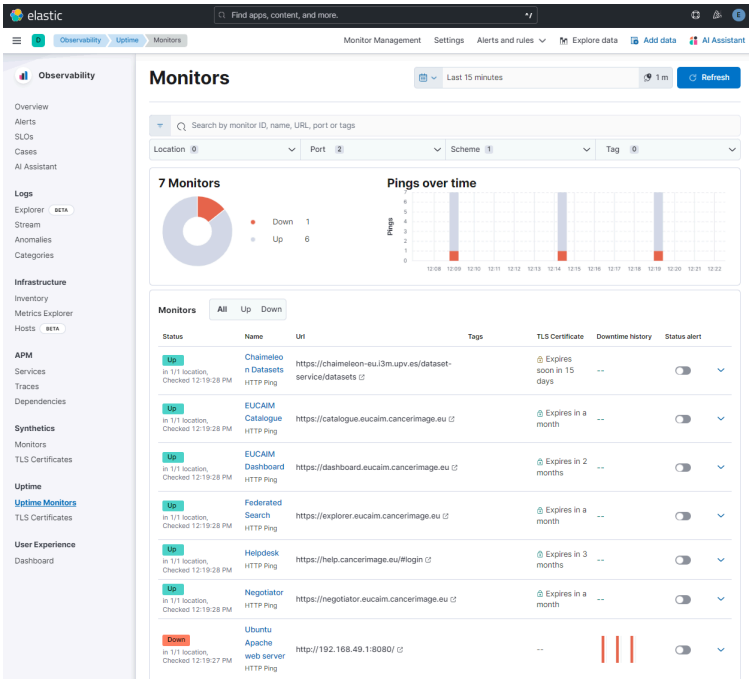
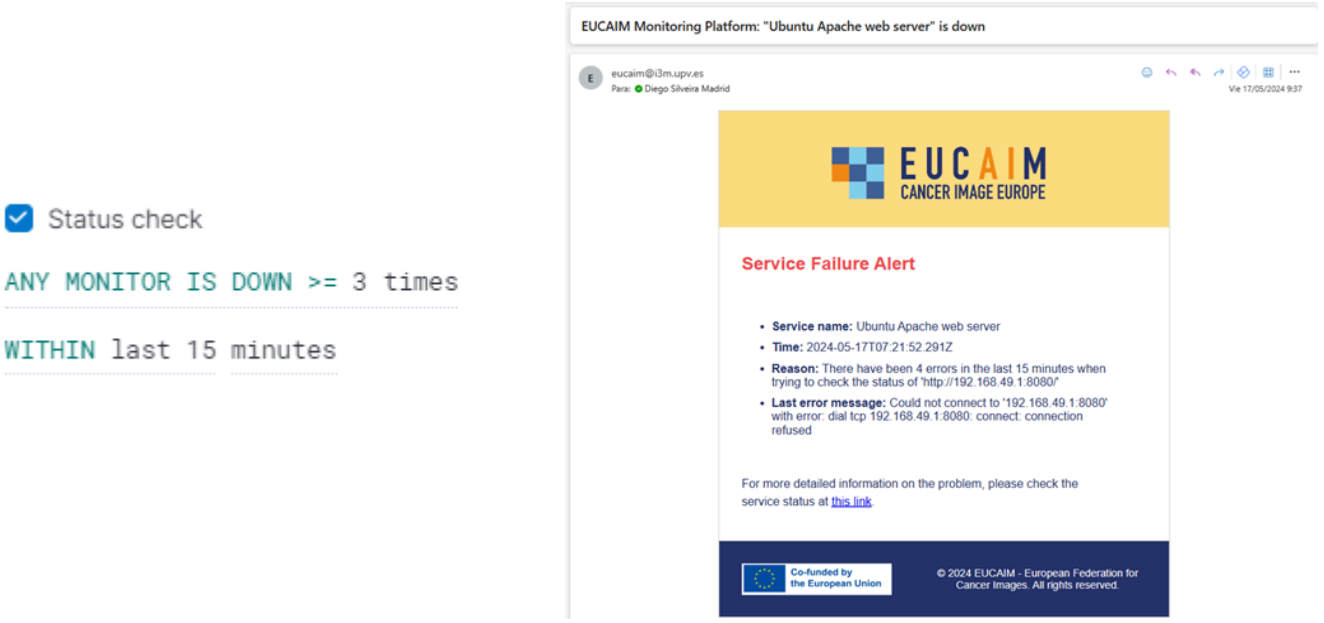


Figure 7*:* Uptime monitors dashboard

In addition to this, a set of rules has been defined that, when fulfilled, trigger the execution of certain actions. These can be useful to generate alerts and send notifications to the person in charge indicating that the service is down or if the TLS certificate is going to expire soon

or has already expired, among many other possible rules. figure 9, shows a rule that checks if any web service is working properly. Specifically, the defined rule checks if any of the web services has reported 3 or more down statuses in the last 15 minutes. If this condition is met, the rule will be executed, generating an email notification like the one shown in figure 8.



a) b)

Figure 9: a) Condition for the execution of a rule that checks the status of the different monitored services; Figure 8: b) Email sent to notify that one of the services is down and the possible cause of this problem.

For the implementation of this functionality, different services of the Elasticsearch-Logstash-Kibana technology stack (ELK stack) have been used, which are defined below. For a better understanding of how this functionality works, a flow chart representing how each of the components interact with each other is shown in Figure 10.

- **Heartbeat**. This service allows the definition of different monitors, each one associated with an EUCAIM service. In this way, the status of a web service can be verified by making requests on a repeating schedule.
- **Elasticsearch**. This service defines an index where documents will be generated each time the rule defined in Figure 9 is executed.
- **Logstash**. Each time a new document is inserted in the Elasticsearch index, this service will take the information from this document and will use it to notify via email the person responsible for the service that it is down.
- **Kibana**. This service is responsible for graphically representing the data and metrics produced or collected by other services (in this case, those produced by Heartbeat and Elasticsearch). This representation is expressed in the form of charts, which are grouped into dashboards (some with a pre-established format and others in which users can define the format you want to represent this data).

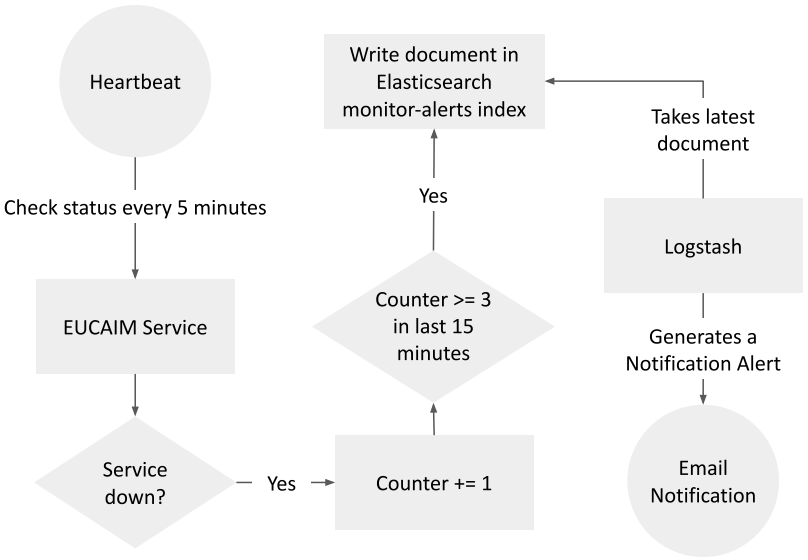


Figure 10: Interaction between the different components for the alert and notification system.

3.6. Helpdesk

The helpdesk is a single point of contact to collect and reply to questions, incidents, requests, etc., from members of the EUCAIM infrastructure The software responsible for the EUCAIM helpdesk is an independent Zammad instance which has been provided for the project.

Only authenticated/authorised users may currently access the helpdesk through the LS-AAI (Authentication and Authorization system) at https://help.cancerimage.eu/#login.

Figure 11: Helpdesk screenshot and ticket workflow.

Helpdesk users may create a new ticket, reply to it or assign a ticket to a proper group or Support Unit (SU).

Several SUs have been configured within the EUCAIM helpdesk, containing users who are specialists in distinct areas corresponding to the SUs. This way, each SU member can reply to a ticket, ask for more information and close/solve the ticket. By default, every new ticket is assigned to the First Level SU. Furthermore, a rota has been established ensuring that new tickets are processed during working hours. Figure 11 also describes the workflow that explains EUCAIM Helpdesk ticket assignment. For more information about EUCAIM Helpdesk, please refer to the EUCAIM Helpdesk page at https://confluence.egi.eu/display/EUCAIM/EUCAIM+-+Helpdesk.

4. Outline of the Video

The demonstration video shows the whole process of creating an account, browsing the catalogue, searching for data, selecting a dataset, requesting access, negotiating the access and accessing the data. It has been divided into five parts:

- Enrolment, which deals with the creation of an account and the registration of a user.
- Dataset exploration, which deals with the process of browsing the collections and filtering the data.
- Access Request, triggered from the catalogue and using the negotiation component.
- Access Negotiation, through the negotiator component, which may involve multiple steps.
- Access granted, considering two cases from two providers.

The last part of the video illustrates the process of exposing a catalogue entry through the Fair Data Point (FDP) API endpoint, and the process of harvesting this endpoint to add it to another catalogue. The figures of the video have been moved to the end of the section to show them in landscape mode to facilitate the reading.

4.1. Enrolment

Users can access the platform through the https://cancerimage.eu website. The platform is made available through the URL https://dashboard.eucaim.cancerimage.eu. In this web application, a user can request the creation of an account through the link https://aai.lifescience-ri.eu/ or by following the instructions described in the “Become a user” section of the Dashboard application.

Figure 12 shows the different steps in the enrolment process. The detailed instructions are provided online[18].

4.2. Browsing and searching for Data

The data catalogue is available through the link “Public Catalogue” in the Dashboard or directly in the URL https://catalogue.eucaim.cancerimage.eu. The catalogue shows the metadata that corresponds to the aggregated data of the collections of one of the data holders.

Figure 13 shows a few snapshots of the data filtering and searching through the catalogue, using the public catalogue and the federated search.

[18] https://drive.google.com/file/d/1MTLMzwCCfN_41cRIGfgT9tfJLyZmrMjK/view?usp=drive_link

4.3. Access request

Once a user has identified the dataset of interest, the user can request access from the catalogue by using the Negotiator. Depending on the dataset, the access request form varies. Figure 14 shows the steps of requesting access and filling in a request.

4.4. Access negotiation

The access negotiation process focuses on analysing the request and verifying that it fulfils the terms and conditions. The negotiation request can go through several statuses, which are outlined below:

- SUBMITTED, when the request has been submitted for the evaluation and it is Under Review.
- APPROVED, when the administrator of the Negotiator platform approved the request.
- DECLINED, when the administrator refuses the application.
- IN_PROGRESS, which means that the request has been approved for processing and it is being processed. During this status it can go through several substatuses. We describe in the video the following types of positive sub-status:
    - CONTACT, informing the representative about the application.
    - MARK_AS_CHECKING_AVAILABILITY, which informs that the representative of the dataset has been contacted and the application is being evaluated in terms of data and resources.
    - MARK_AS_AVAILABLE, data and resources are available.
    - INDICATE_ACCESS_CONDITIONS, the access conditions have been notified to the requester.
    - ACCEPT_ACCESS_CONDITIONS, when the requester has accepted the access conditions.
    - GRANT_ACCESS_TO_RESOURCE, when the representative of the dataset has given access to the applicant.
- CONCLUDED, when the process has been completed and the application project has been completed, and access has concluded and has been revoked..

All these status changes are notified by email. More details on the negotiation process are given in deliverable *D4.13 End-user guide to the system*. Figure 15 and Figure 16 show the different steps. This workflow is managed by three roles:

- The **researcher**, who submits the application request. The actions in the figure are executed from an “incognito” mode (black) Microsoft Edge browser.
- The **negotiator admin**, who takes the requests and sends them to the access committee, and approves or denies the processing of the request. The actions in the figure are executed from a regular (grey) Microsoft Edge browser.
- The **dataset representative**, who checks that the conditions are met and interacts with the requester for granting access. The dataset representative could be the representative of an EUCAIM reference node or the representative of each data holder. The actions are performed in a Google Chrome (blue) browser.

The actions are depicted in 2 main phases:

- Figure 15: Dataset Access application submission and approval, with the participation of the researcher submitting the application and the negotiator admin accepting the request.
- Figure 16: Dataset Access Negotiation, with the participation of the researcher and the Dataset representative.

4.5. Data Access

Once a data access request is granted, access can be performed through an endpoint provided in the catalogue through the negotiator. Two examples of this are shown in this section: one from the CHAIMELEON data holder and another from the EuCanImage data holder.

4.5.1. Access to CHAIMELEON data

After being granted access to CHAIMELEON data, the first step that a researcher should take is to register in the CHAIMELEON platform using the same credentials as the ones used in EUCAIM. There they must select the External Researcher profile and complete the registration, confirming the acceptance of the access conditions.

Before the access is granted, the responsible person of the dataset has to update the profile in the CHAIMELEON platform, granting them access to the specific dataset. Figure 17 and Figure 18 show the different steps, verifying that a user cannot access a dataset before the account is created (Figure 17) and after the account is validated (Figure 18). In this case, only the data from the Virtual Research Environment is shown.

4.5.2. Access to EuCanImage data

The data of EuCanImage is hosted on the Euro-BioImaging Medical Imaging Repository. The Euro-BioImaging Medical Imaging Repository service is an XNAT instance operated by Health-RI. The first step that a researcher user should take is to request an account for XNAT[19] at the EUCAIM Helpdesk, who will forward the request to the Health-RI helpdesk[20]. The researcher will receive an email to agree with the Health-RI data access policy and will receive their credentials. The Euro-BioImaging XNAT is currently in the process of implementing SRAM which will allow using the same credentials as the ones used in EUCAIM in the future.

Before the access is granted, the responsible person of the dataset has to update the profile in the XNAT, granting them access to the specific dataset. Figure 19 shows the different steps from logging in to visualising the imaging data on XNAT.

4.6. Exporting to and harvesting from a Fair Data Point

As explained in section 3.2, the metadata of a catalogue entry can be shared in a standardised manner through the use of a Fair Data Point (FDP). To enable this, two sets of database tables are maintained in the database of the Molgenis system, one to store the metadata according to the internal catalogue metadata model and the other one to store the metadata according to the Fair Data Point metadata model.

The first step of exposing a catalogue entry through an FDP is to retrieve the metadata as is from the catalogue, convert it to the FDP metadata model, and then re-upload it to the FDP database tables. This is done by running a custom Python script that is provided by EUCAIM.

After the information is added to the FDP database tables in Molgenis, the metadata will become available through the FDP endpoint. From there it can be harvested using the proof-of-concept tool mentioned in section 3.2. In the demonstration video the FDP endpoint from the EUCAIM catalogue is harvested and a catalogue entry is added to the EIBIR

[19] Euro-BioImaging Medical Imaging Repository: https://xnat.bmia.nl
[20] Health-RI helpdesk: https://trait.topdesk.net/

catalogue with that metadata. Currently, the FDP does not expose enough information to satisfy all the required attributes of the EIBIR catalogue metadata model. In order to perform this demonstration additional default values are supplied. Finally, by using the FDP to add an entry to the EIBIR catalogue, the entry on the EUCAIM catalogue is edited and after propagating the changes to the metadata exposed by the FDP, these changes are harvested and the entry on the EIBIR catalogue is updated.

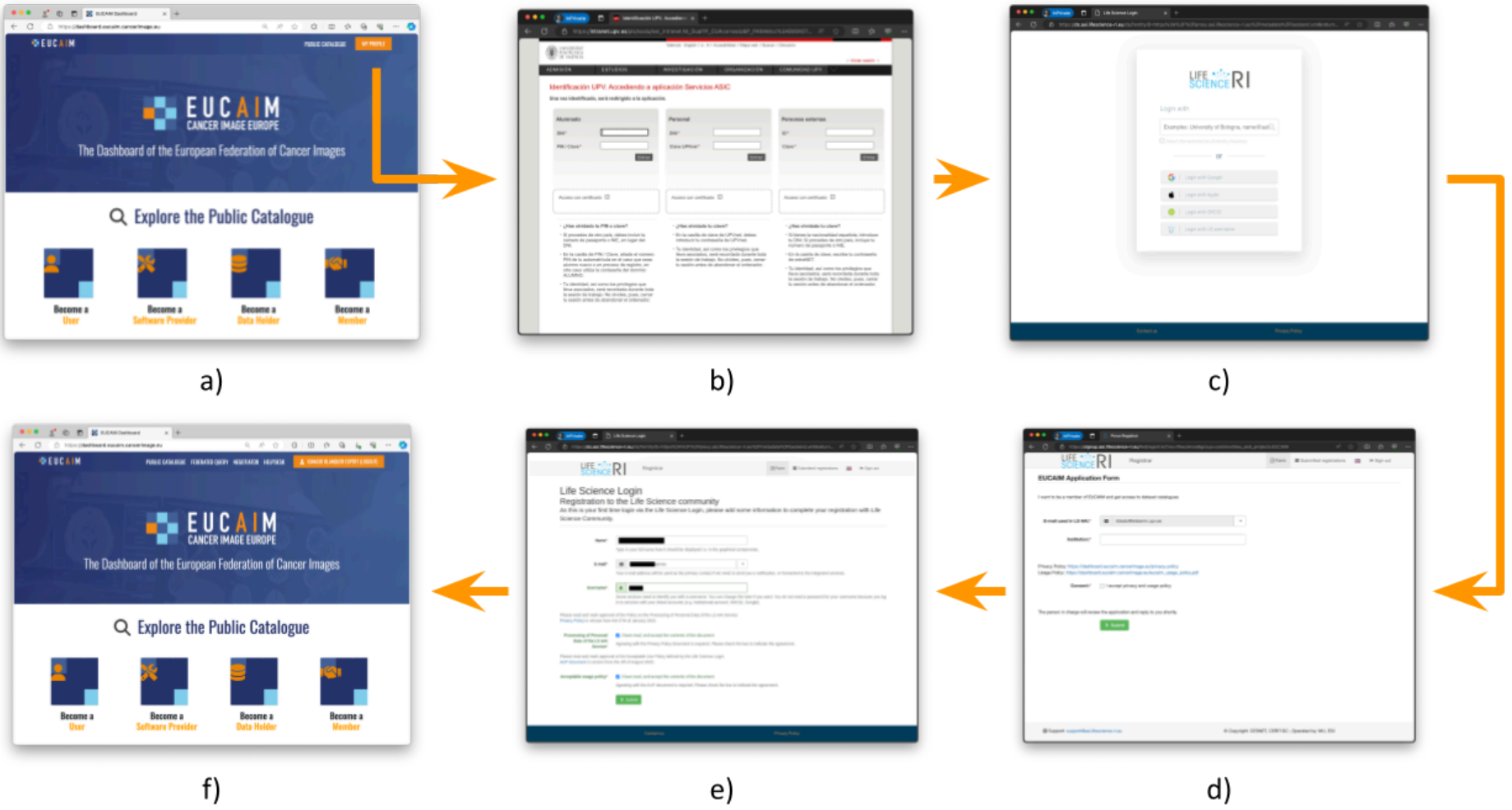


Figure 12: User registration and EUCAIM VO enrolment process: a) Unauthenticated access to the Dashboard; b) selection of the institutional IdP; c) authentication with the local IdP; d) creation of the LS-AAI account; e) request of membership to the EUCAIM VO; and f) authenticated access to the Dashboard.

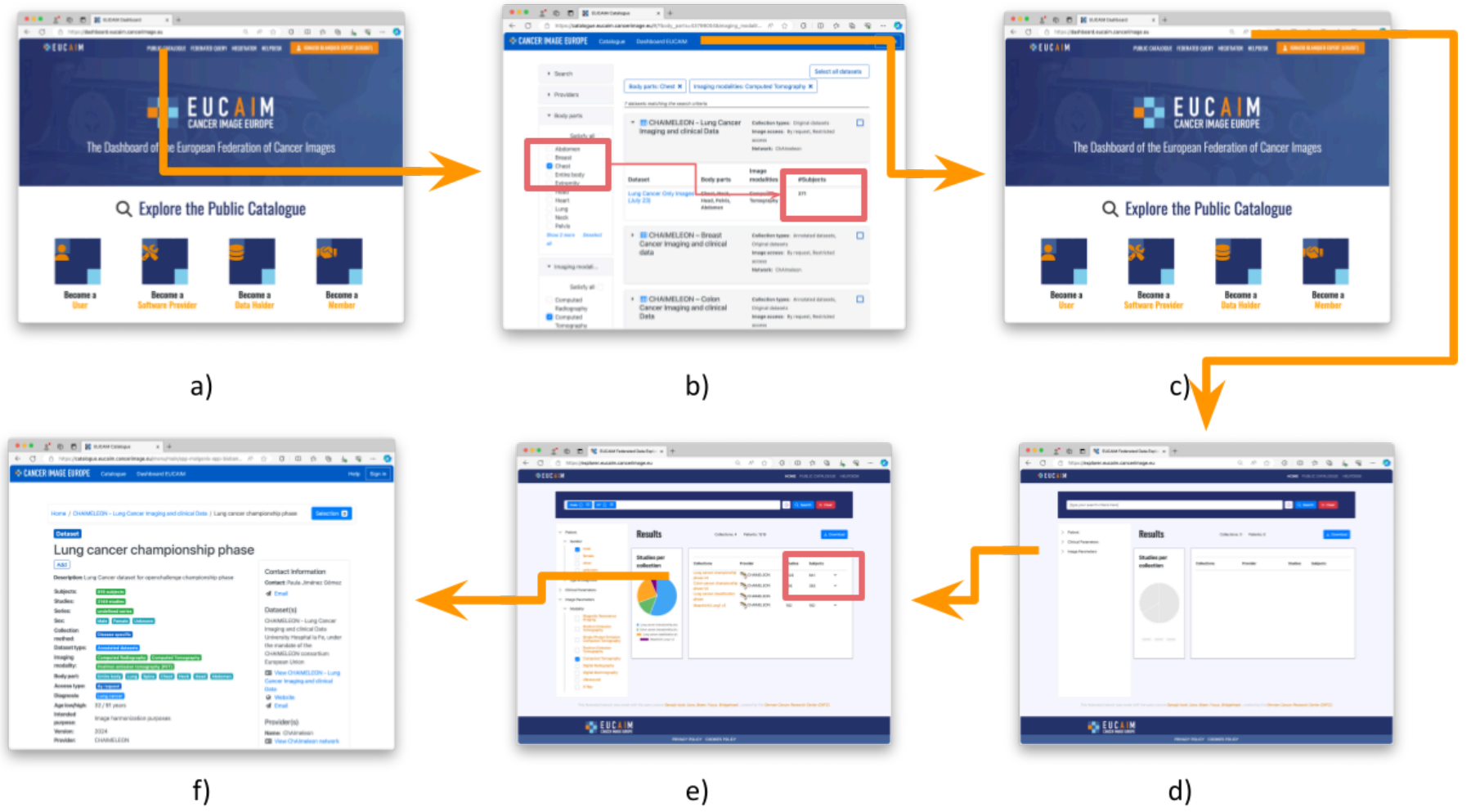


Figure 13: Data filtering and searching: a) Access to the Public catalogue through the Dashboard; b) filtering datasets by body part; c) access to the federated search through the dashboard; d) selection of the searching criteria; e) visualisation of the number of subjects that fulfil the searching criteria; and f) access to the dataset metadata through the link in the federated search to the catalogue.

Figure 14: Access request: a) Adding the dataset to the request; b) send the request to the negotiator; c) opening the dataset access request form; d) filling out the access form; e) submission of the access request; and f) browning the access request status.

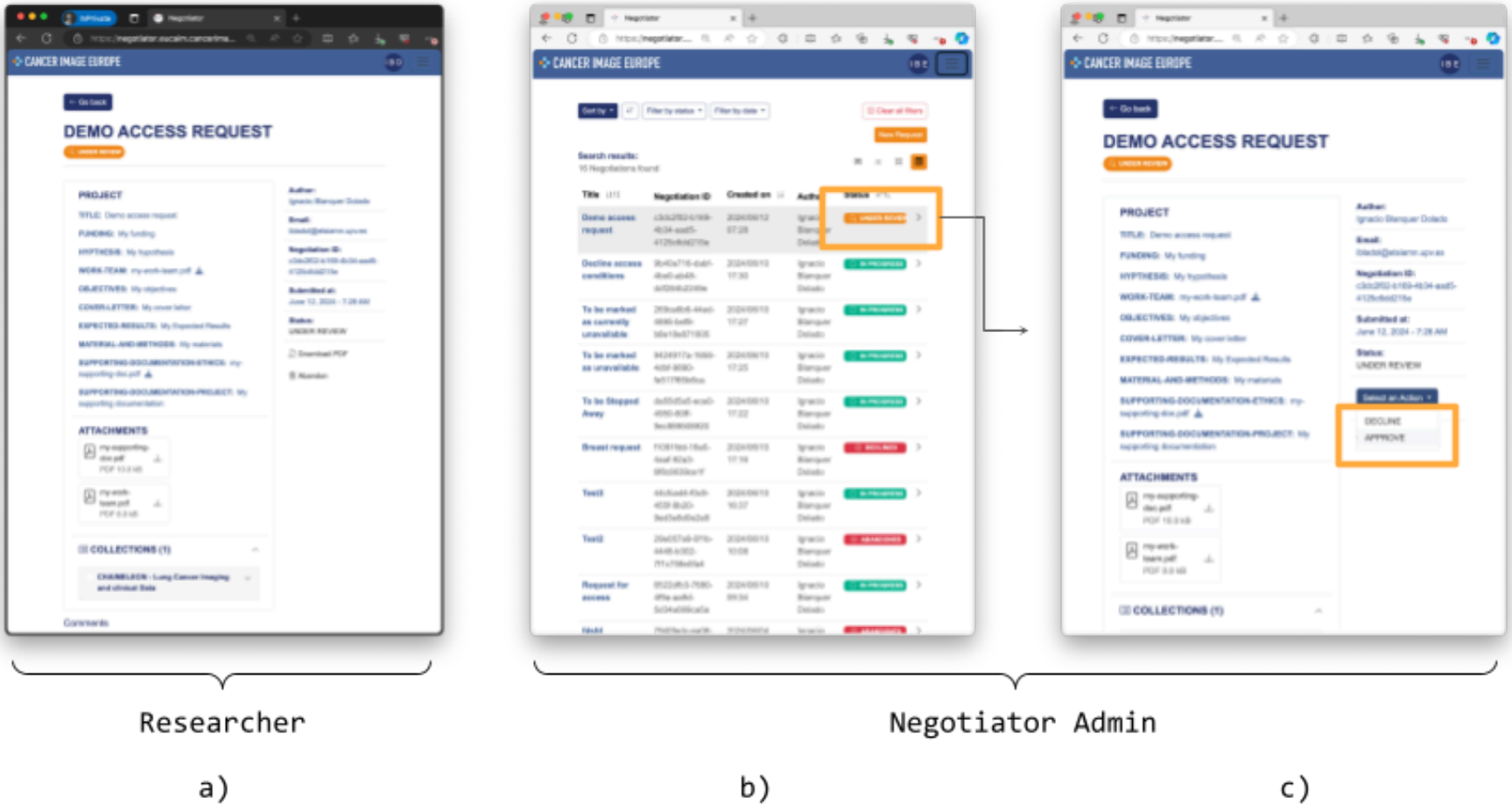
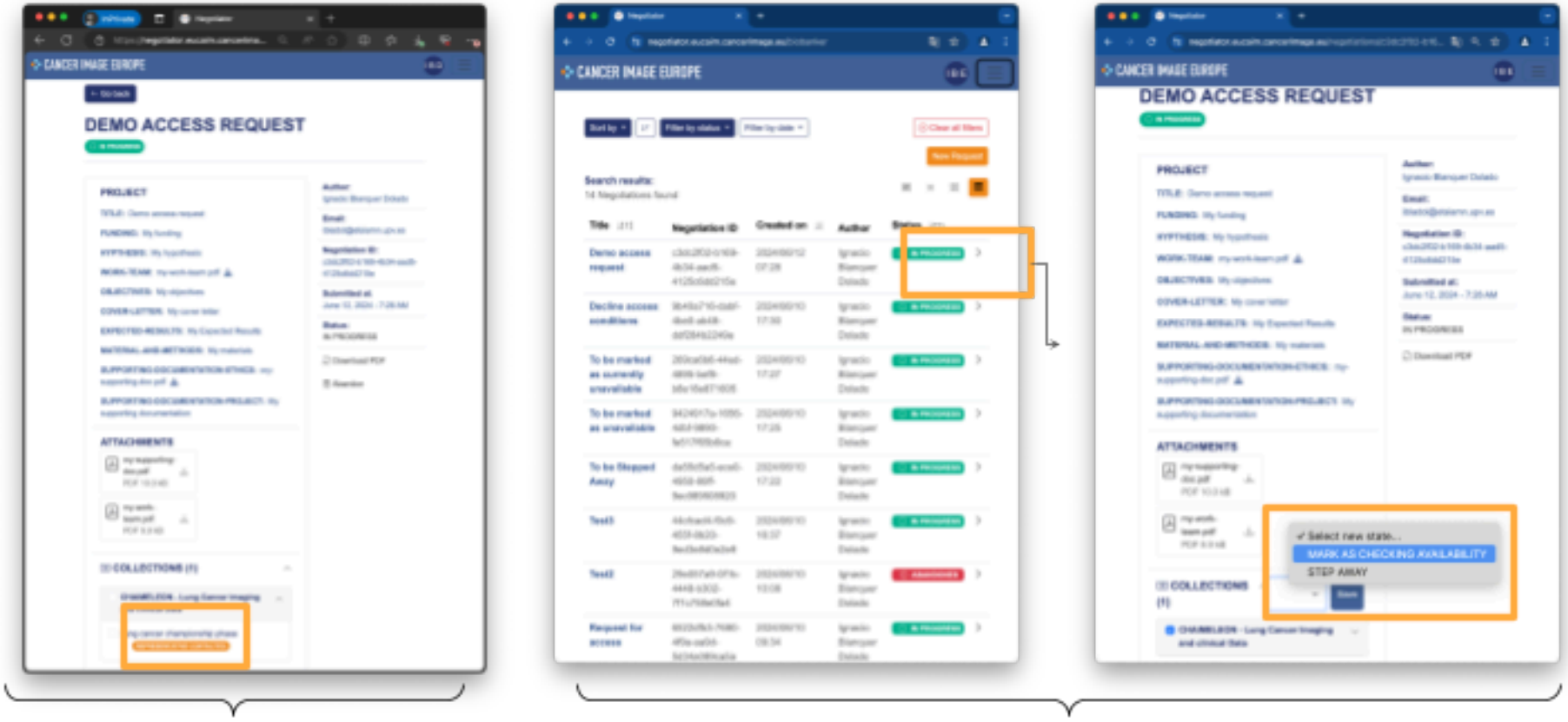


Figure 15: Dataset Access application submission and approval: a) Researcher submits an application; b) Negotiator admin checks the application; and c) Negotiator admin approves the application.



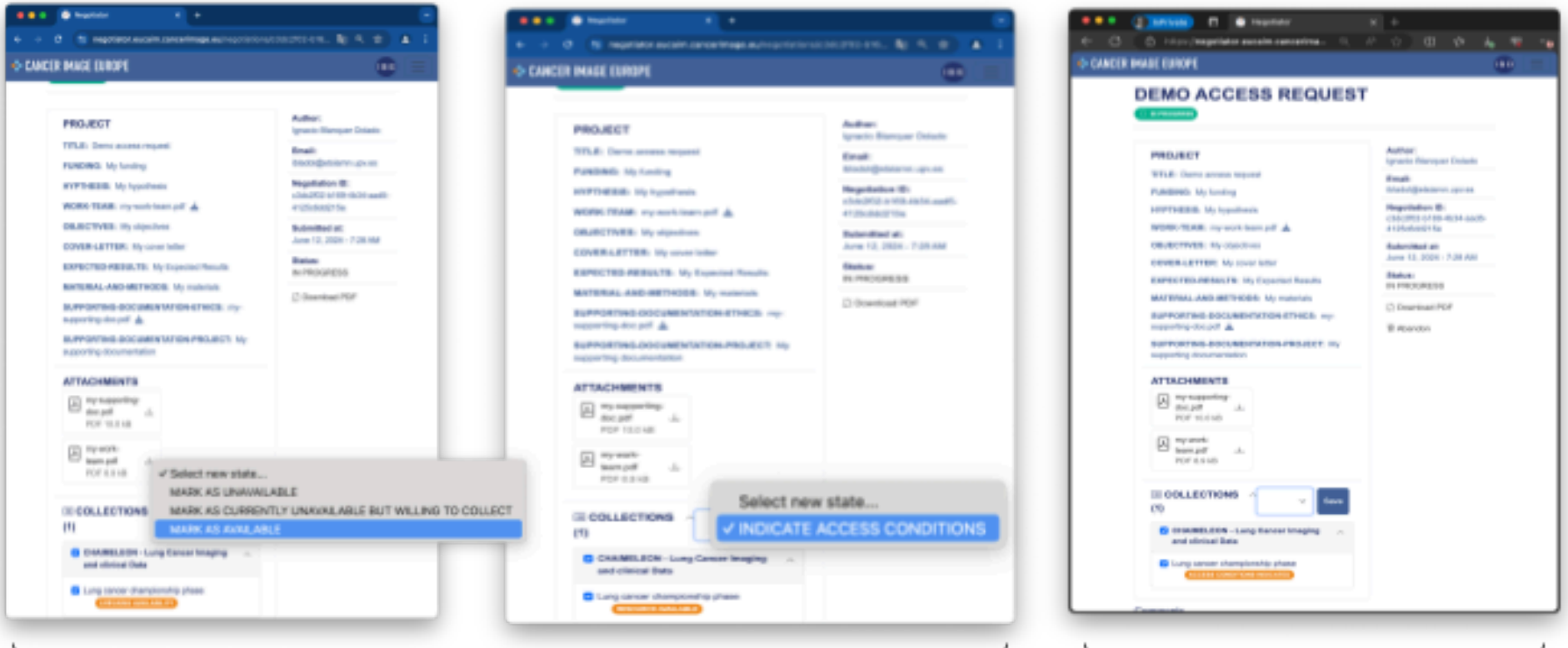
Researcher

Dataset Representative

a)

b)

c)



Dataset Representative

Researcher

d)

e)

f)

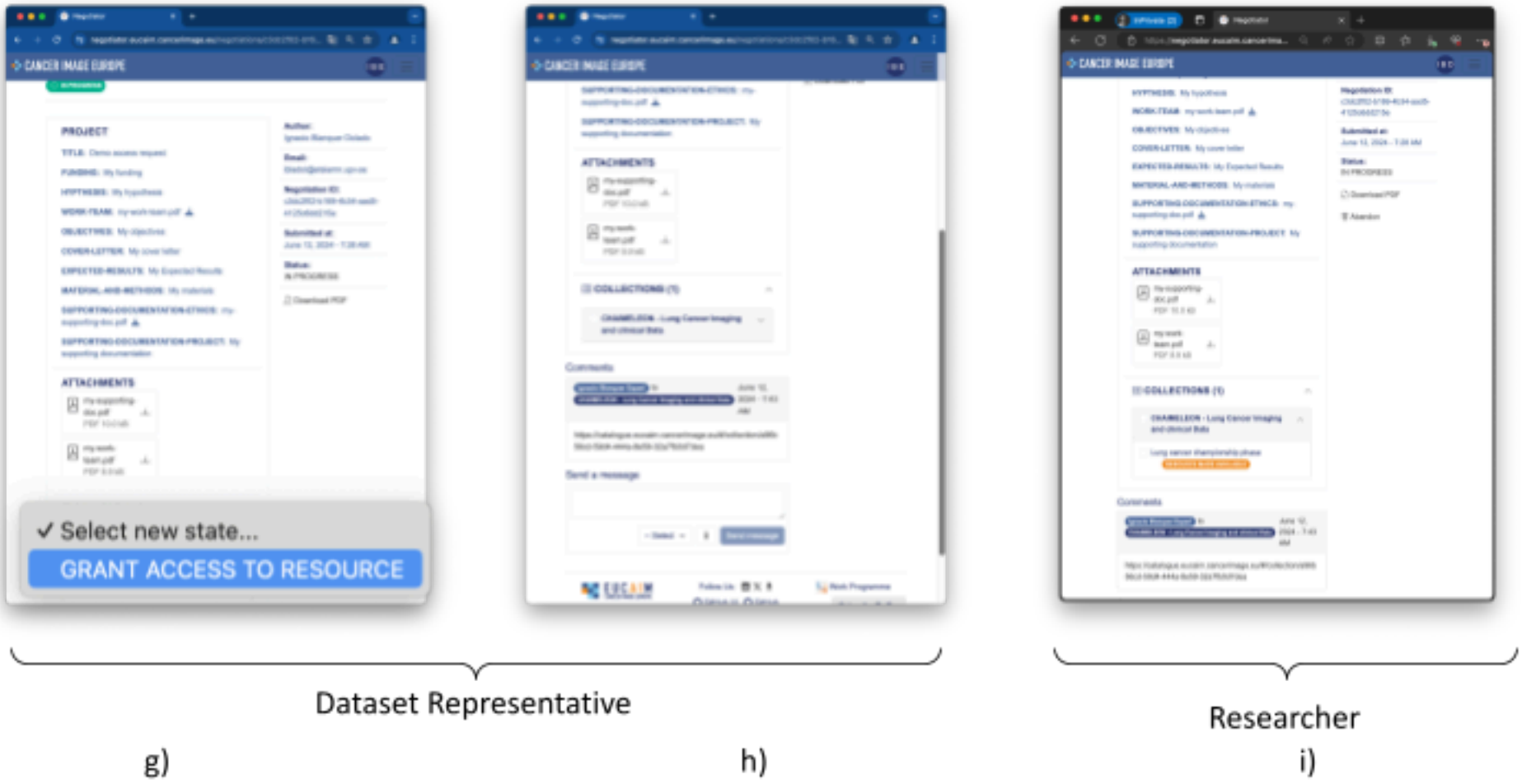


Figure 16: Dataset Access Negotiation: a) Contacting the dataset representative; b-c) accepting the request and checking the availability of the data requested; d) confirmation of the availability of data; e) send access conditions; f) researcher accepts access conditions; g-i) access granted and communication of the URL to the researcher.

Figure 17: Access to the dataset before registering in the platform: a) Successful access to the dataset metadata; b) Creating a Virtual Research Environment (VRE); c) Failure due to the lack of permissions.

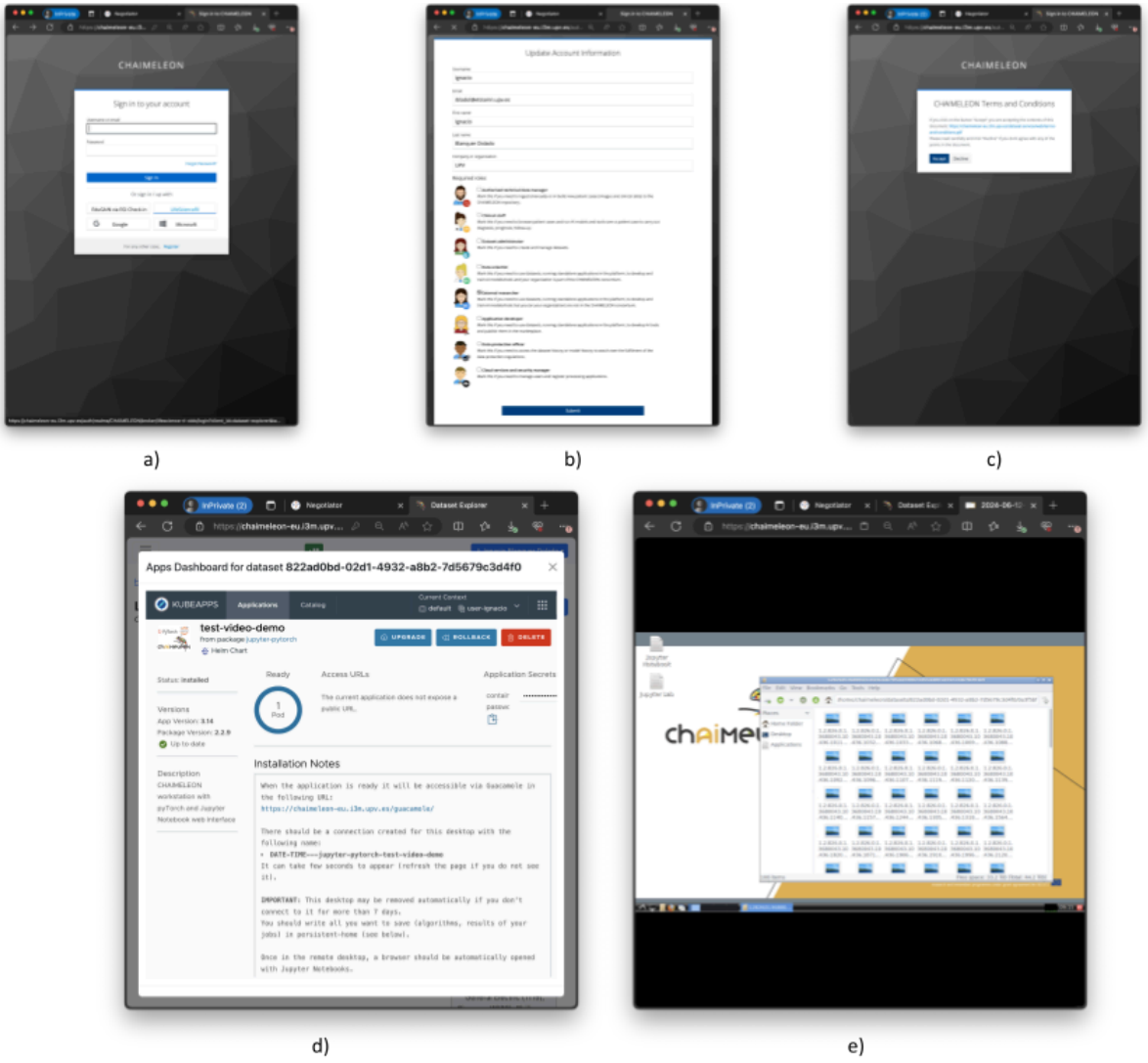
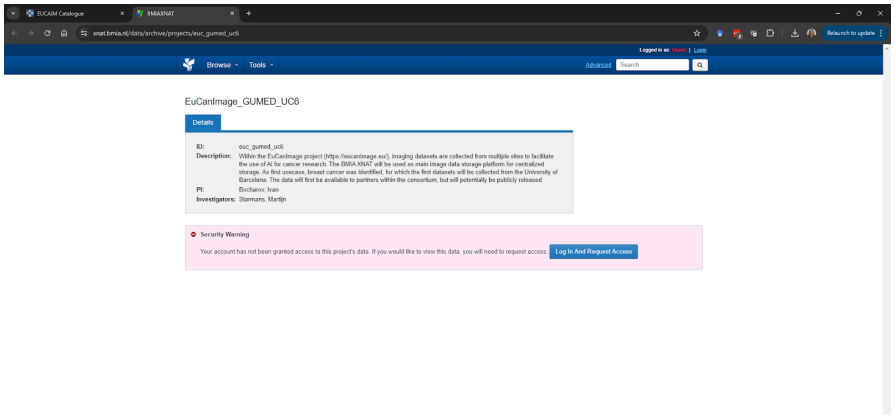
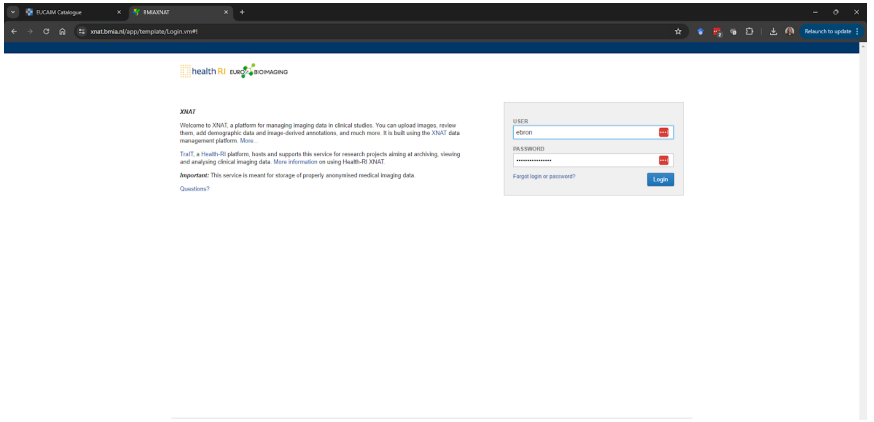


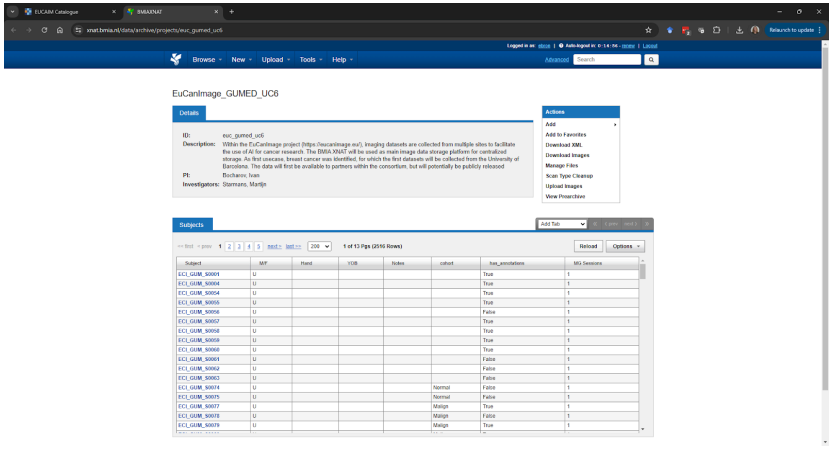
Figure 18: Dataset Access: a) logging in CHAIMELEON through the LS-AAI; b) requesting the “External Researcher” role; c) accepting the terms and conditions; d) successfully creating a Virtual Research Environment (VRE); and e) accessing the data through the VRE.



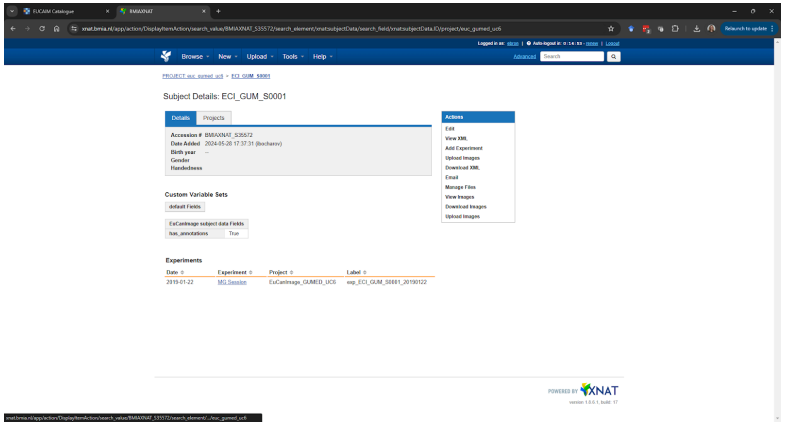
a)



b)

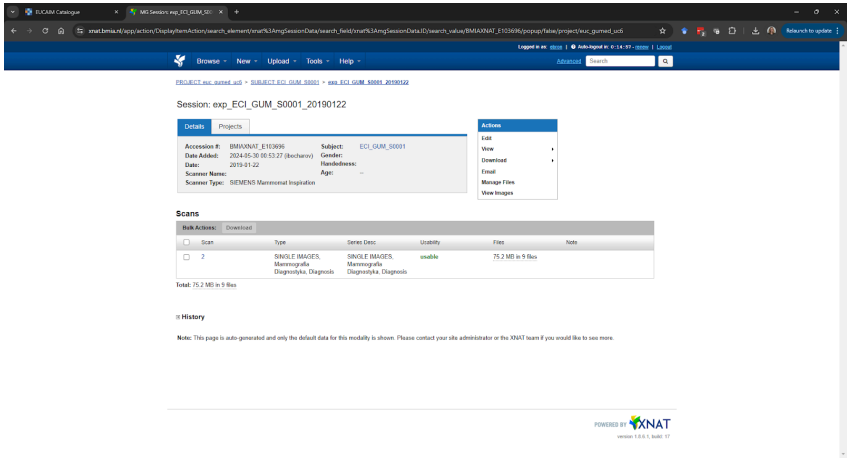


c)

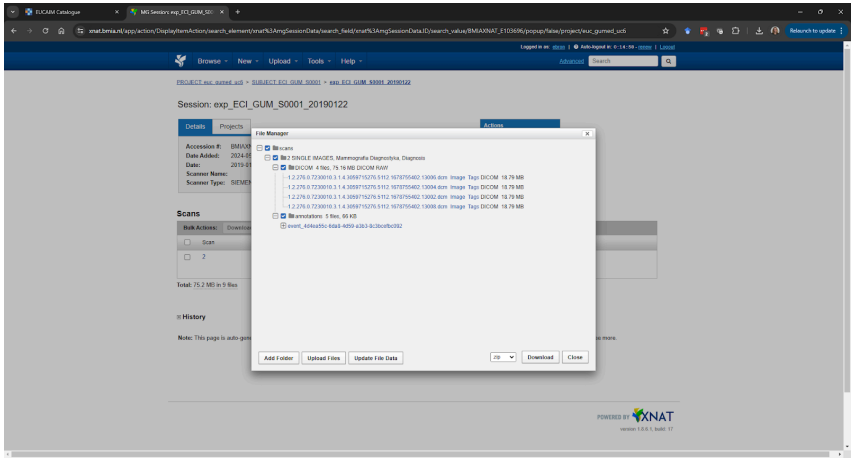


d)

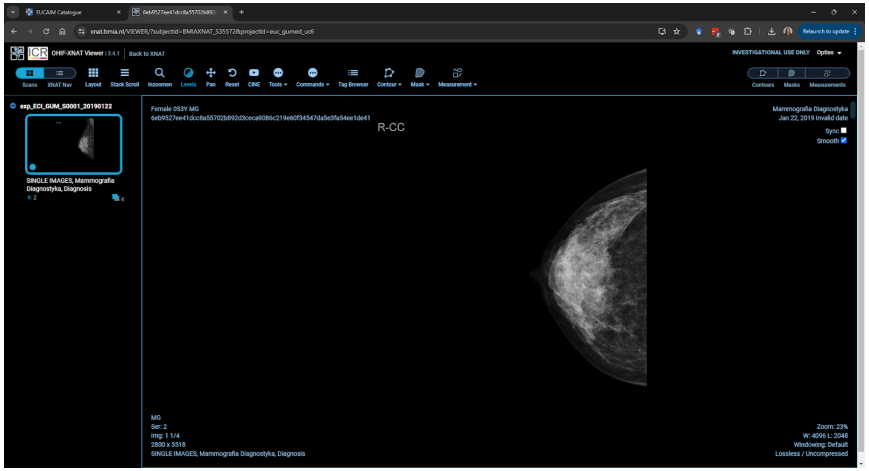
Figure 19: Access to EuCanImage data: a) the dataset link on Euro-BioImaging XNAT for unauthorised users; b) logging in the Euro-BioImaging XNAT through a valid user account; c) accessing the data through on the Euro-BioImaging XNAT: subject list; d) accessing the data for a specific subject;



e)

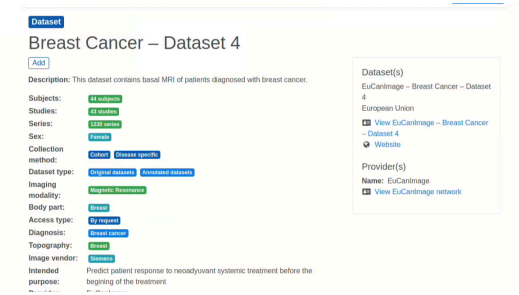


f)

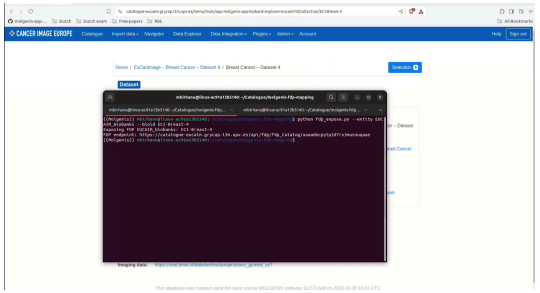


g)

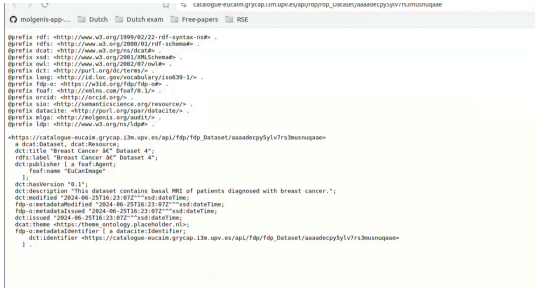
Figure 20: Access to EuCanImage data: e) access the data for a specific experiment; f) downloading all available files via “Manage Files”; g) visualising a specific scan using “View Imaging” with the OHIF Viewer.



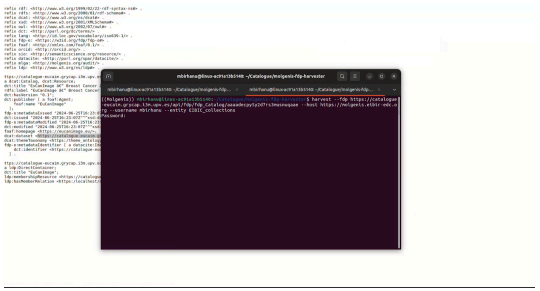
a)



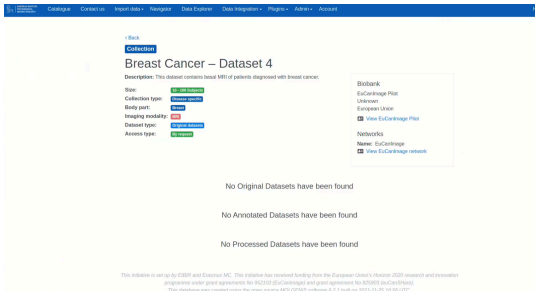
b)



c)



d)



e)

Figure 21: Exposing to and harvesting from a Fair Data Point (FDP): a) the dataset as shown on the EUCAIM catalogue; b) running the Python script to expose the catalogue entry through the FDP; c) the dataset metadata is made available through the FDP in Turtle format; d) running the Python program to harvest from the FDP; e) the resulting catalogue entry on the EIBIR catalogue.

5. Conclusions

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 *D4.5 First Federated Core Services*. EUCAIM Core services are those that support the main functionalities of the federation, and comprise: the Authentication and Authorisation, the public catalogue, the federated search service, the data access request service, the service monitoring and the helpdesk.

The architecture of the platform considers this federated model where nodes can store and process data. The authentication and authorisation is managed through the Life Sciences AAI service, the proposed de-facto standard for biomedical Research Infrastructures. The services differentiate between anonymous and authenticated access and comprise the capability of browsing and searching for data, according to the dataset metadata and the aggregated information of the federated search. The access to data requires the approval of an access committee, and data can be accessed according to different models: downloading for public anonymous datasets, on-site processing for restricted datasets that can be visualised and federated processing for those datasets that have the highest access restriction.

The information of this deliverable is complemented in the rest of WP4 M18 deliverables:

- *D4.7 First EUCAIM Dashboard*, describing the interoperability of the Dashboard.
- *D4.9 Central Core Infrastructure set-up*, which describes the resources and recipes used to deploy the core services.
- *D4.10 Technical evaluation of the Platform*, describing the validation of the services at the technical and user-level.
- *D4.13 End-user guide to the system*, a user manual of the core services.