Project title: European Federation for Cancer Images

Project acronym: EUCAIM

Grant Agreement: 101100633

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

D4.2: Final EUCAIM Operational Platform

Partner(s):   HULAFE

Author(s):   Irene Marín, Carina Soler, Pedro Miguel Martínez, Patricia Serrano, Luis Martí Bonmatí (HULAFE), Ignacio Blanquer (UPV), Carles Hernández (BSC)

Contributors: Celia Martín (QUIBIM), Diego Silveira (ITI), Laure Saint-Aubert (MEDEX), Alma Vitro (UV), Ana Miguel (MAT), Kurt Majcen (BBMRI-ERIC)

Reviewers: Peter Gordebeke (EIBIR), Ricard Martínez (UV), Salvador Capella-Gutierrez (BSC)

Date of delivery: 20/12/2024

Version: 1

Table of contents

List of abbreviations

AAI = Authentication and Authorisation Infrastructure

AB = Advisory Boards

AC = Access Committee

ADC = Apparent Diffusion Coefficient

AdmCo = Administrative Project Coordinator

AI = Artificial Intelligence

AI4HI = Artificial Intelligence for Health Imaging

API = Application programming interfaces

CA = Consortium Agreement

CDM = Common Data Model

CHO = Central Hub Office

CPU = Central Processing Unit

CT = Computed Tomography

D = Deliverable

DH = Data Holder

DPA = Data Processing Agreement

DPO = Data Protection Officer

DSA = Data Sharing Agreement

DTA = Data Transfer Agreement

DU = Data User

DW = Data Warehouse

EC = European Commission

EHDS = European Health Data Space

EHDS-R = European Health Data Space Regulation

ELB = Ethical and Legal Board

ETL = Extract Transform Load

EU = European Union

FAIR = Findable, Accessible, Interoperable, Reusable

GA = Grant Agreement

GB = Governing Body

GDPR = General Data Protection Regulation

GPU = Graphics Processing Unit

HDAB: Health Data Access Body

IPR = Intellectual Property Rights

LS = Life Science

MB = Management Board

MR = Magnetic Resonance

MS = Milestone

RC = Research Community

R&I = Research and Innovation

RI = Research infrastructures ç

RWDH = Real World Data

RWDH = Real World Data Holder

SC = Steering Committee

SCo = Scientific Coordinator

SW = Software

SLA = Service Level Agreement

SP = Software Provider

TB = Technical Board

T&C = Terms and Conditions

ua = User Action

UML = Unified Modelling Language

us = User Story

WP = Work Package

WPLs = Work Package leaders

1. Introduction

Aim and scope of the deliverable

This deliverable is an update to Deliverable D4.1[1], *First EUCAIM Operational Platform*, of which the initial version was submitted in June 2023. In this version, concepts have been updated and improved as the project has advanced, incorporating the additional information and developments gathered throughout the process. It is worth noting that in this document, the term 'EUCAIM' refers to the platform developed under WP4 within the framework of the project. In the future, and within the WP8 and the EDIC framework, it will be replaced by 'Cancer Image Europe'.

This deliverable provides a detailed overview of the various user roles that will interact with the Cancer Image Europe infrastructure deployed by the EUCAIM project, referred to in this deliverable as EUCAIM platform. It outlines the specific actions expected from each of them within the infrastructure and through the Dashboard, which is the website intended for users who want to use data provided through the platform for analysis in the context of research and innovation. In addition, this document highlights the various levels of authorization and authentication required for each specific action. For further information complementing this document, readers may refer to the EUCAIM Glossary[2] as well as to the 'Architecture of the Federated Repository' document[3], both of which are continuously being updated.

A range of user stories and potential user actions, which serve as descriptive accounts of how users from different roles interact with the platform are presented. These user stories and user actions play a significant role in the definition of the operational procedures, considering the associated needs, restrictions, capabilities, and other relevant factors.

By detailing the operational procedures, D4.2 provides clear guidance on how to navigate and utilise the EUCAIM platform effectively, addressing the requirements and constraints associated with each user role and providing insights into the capabilities available to them. These procedures ensure a streamlined and secure user experience within the EUCAIM ecosystem. As a result of this analysis, the functional requirements of the platform have been defined, which will be directly reflected in the design and implementation of the platform and, more specifically, in the functionalities and workflows made available through the Dashboard.

In addition to the aforementioned functional and technical aspects, D4.2 encompasses a comprehensive set of definitions for specific terminology used within the EUCAIM project, which are harmonised with the terms of the future European Health Data Space Regulation[4]. These terms span various domains, including technical, AI, medical, legal, and ethical fields. These definitions hold significant importance as they will be extensively utilised throughout the

---

[1] First EUCAIM Operational Platform: https://cancerimage.eu/wp-content/uploads/2024/02/D4.1-First-EUCAIM-Operational-Platform-HULAFE-DECEMBER-REVIEW.pdf

[2] EUCAIM Glossary https://cancerimage.eu/glossary/

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

[4] European Parliament (27/11/2024). CORRIGENDUM to the position of the European Parliament adopted at first reading on 24 April 2024 with a view to the adoption of Regulation (EU) 2024/… of the European Parliament and of the Council on the European Health Data Space. P9_TA(2024)0331 (COM(2022)0197 – C9-0167/2022 – 2022/0140(COD)). https://www.europarl.europa.eu/doceo/document/TA-9-2024-0331-FNL-COR01_EN.pdf

project's life span. These terms have served as the foundation for the project's glossary, which is continuously updated as the project progresses.

Along with the operational procedures defined, D4.2 presents an overview of the governance bodies responsible for the operation of the Central Hub (comprising the Reference Nodes and the services and software provided by the EUCAIM platform). These bodies are integral to the project's management structure, including the Central Hub's governance structure, which is envisioned to be an evolution of the existing EUCAIM Project management structure. The overview included herein offers insights into the roles, responsibilities, and decision-making processes of the governance bodies, ensuring effective coordination and collaboration within the project.

Furthermore, an initial approach to defining the rules for participation for Data Holders was established within deliverable D4.1 towards MS23. However, these rules have been excluded from the current version, as they are now covered in detail in Deliverable D4.3[5], *First Rules for Participation Report*, and its final version, D4.4 *Final Rules for Participation Report (expected by January 2025).*

This document also presents in Section 7 an updated version of the Service Level Agreement (SLA) template, specifically designed to regulate the integration of a federated node into the EUCAIM Network. The guidelines and agreements in the SLA ensure standardised practices and facilitate smooth collaboration among federated participants.

Please note that the content of this current version has undergone both minor and major revisions in response to the project's evolution and the recent approval of two key legislations: the "EU Health Data Space: more efficient treatments and life-saving research"[6] and the Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC)."[7] Our objective is to ensure maximum alignment with these regulations.

Relation with other deliverables and Work Packages (WPs)

WP4 (Governance and implementation of the Central Hub) is central to all other WPs in the project which feed into and out of it. As such, D4.2 is paving the path towards the design and definition of many important project implementations, especially those related to WP3 (Ethical and legal aspects), WP5 (Data federation and interoperability framework), and WP6 (Federated data processing and analysis).

The rules for participation drafted in the previous version of this document (D4.1) have been further documented in D4.3 (First rules of participation report) and finalised in D4.4 (Final rules of participation report), and excluded from this current version (D4.3). Likewise, operational procedures herein defined will be further refined and documented in deliverable D4.12, all of

---

[5] D4.3 First Rules for Participation Report: https://cancerimage.eu/wp-content/uploads/2024/02/EUCAIM_D4.3_DECEMBER-REVIEW.pdf

[6] EU Health Data Space: more efficient treatments and life-saving research https://www.europarl.europa.eu/news/en/press-room/20240419IPR20573/eu-health-data-space-more-efficient-treatments-and-life-saving-research

[7] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). https://eur-lex.europa.eu/eli/reg/2024/1689/oj

which pertain to the design and implementation of the core services, dashboard, and set up of the core infrastructure, eventually producing the related operational report.

It is also worth noting that the operational framework described in this deliverable is very closely linked to the legal framework specified in the project's Data Management Plan (D3.6).

2. User roles

This section focuses on the User Roles, which refer to those users interacting with the EUCAIM platform, from a functional and technical point of view. Depending on the roles assigned to a user, they can perform specific actions corresponding to functional requirements (e.g., ingest data, request access to data, provide processing and AI software, etc.), which starts an interaction flow supported by the components of the final architecture. The User Roles identified so far are as follows:
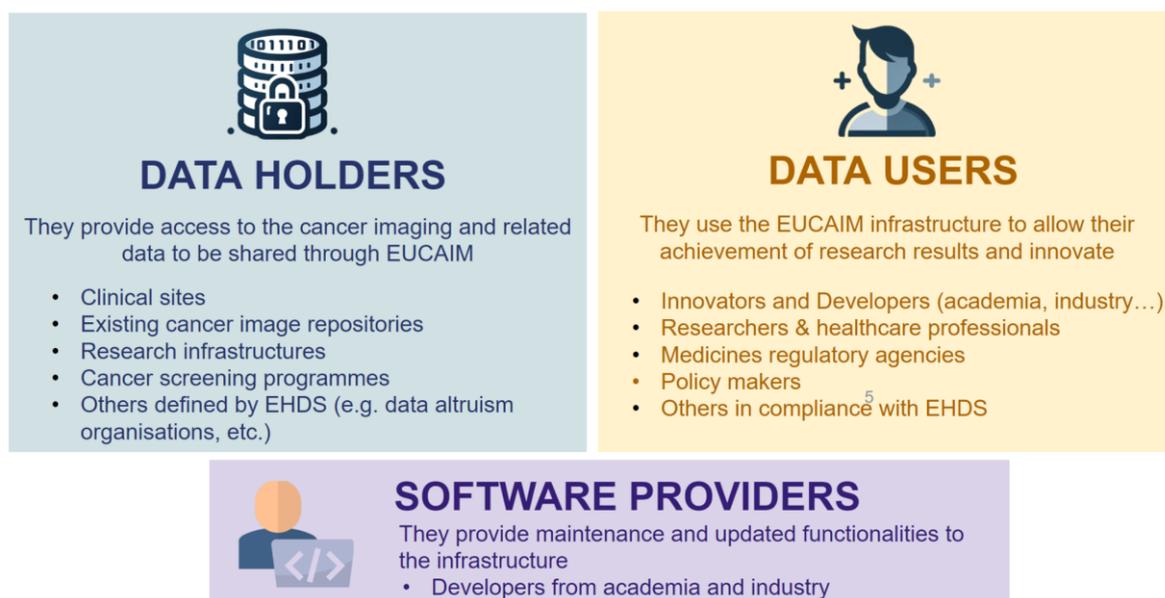
Figure 1. Summary of the EUCAIM user roles

The roles defined here depend heavily on the final text of the European Health Data Space Regulation (EHDS-R). This future regulation gives EU Member States the power to expand the categories of data holders and to create new ones, such as intermediary services. In some cases, it should be understood that the definition of the roles in the platform may correspond to a *de facto* action or condition. For example, research consortia do not usually have legal personality, but nothing prevents them from generating datasets and integrating them into the EUCAIM environment within the framework of the consortium agreement or a specific agreement.

1. **Data Holder (DH):**

    A Data Holder[8] refers to any natural or legal person, including entities, bodies, and research organisations in the health or care sectors, as well as European Union

---

[8] According to future EHDSR a 'health data holder' means any natural or legal person, public authority, agency or other body in the healthcare or the care sectors, including reimbursement services where necessary, as well as any natural or legal person developing products or services intended for the health, healthcare or care sectors, developing or manufacturing wellness applications, performing research in relation to the healthcare or care sectors or acting as a mortality registry, as well as any Union institution, body, office or agency, that has either:

institutions, bodies, offices, and agencies, who has the right, obligation, or capability to make certain data available for secondary uses under EHDS-R, including registering, providing, restricting access, or exchanging the data. Examples of Data Holders include data repositories, infrastructures, regional biobanks, clinical centres, cancer screening programs, public entities, pharmaceutical companies, data altruism initiatives, and publication repositories.

Since EUCAIM is a hybrid platform, DHs have two options for joining the federation: they can technically either become a federated node or transfer their anonymized data to the Reference Nodes. It is important to note that the term "federated node" in this document refers to a technical component for data processing and is distinct from the concept of "node" in institutional or governance contexts, such as those envisioned within the EDIC framework.

Additionally, depending on their origin, two distinct environments can be identified (*Figure 2*):
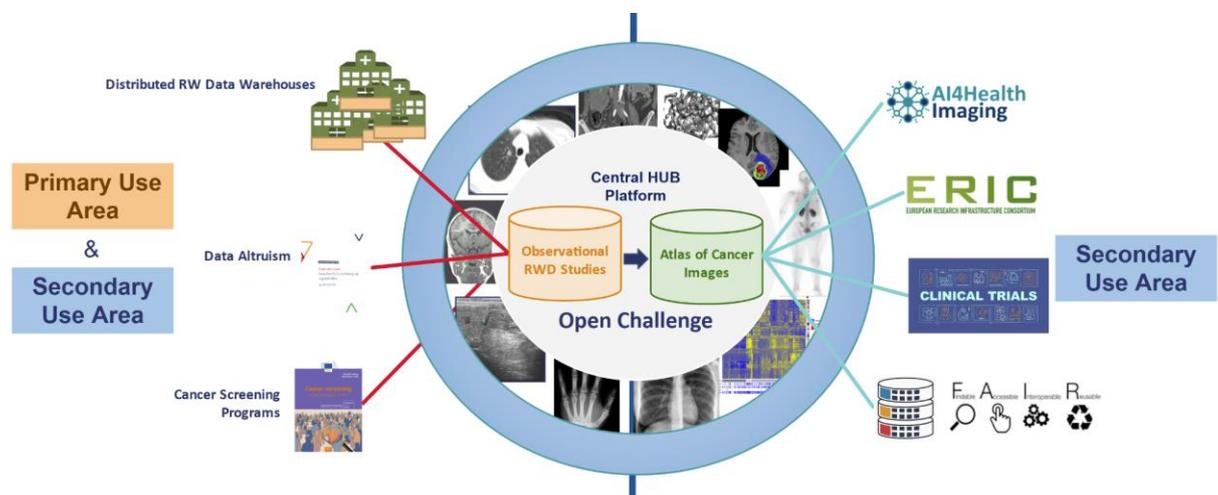
Figure 2. EUCAIM hybrid platform

- **Research and Innovation Environment** (right column): Comprising existing research repositories originating from European framework projects, beginning with the AI4Health Imaging initiatives, but also including other European Research Infrastructures, clinical trials, and more. Together, these secondary-use data repositories form the European Atlas of Cancer Images.
- **Real World Data Environment** (left column): Composed of distributed Real World Data Holders (RWDHs), such as hospitals and research institutes with access to primary health data, cancer screening programs, and data altruism initiatives. These RWDHs will make primary health data readily and efficiently available for Research and Innovation (R&I) through observational studies.

(i) the right or obligation, in accordance with applicable Union or national law and in its capacity as a controller or joint controller, to process personal electronic health data for the provision of healthcare or care or for the purposes of public health, reimbursement, research, innovation, policy making, official statistics or patient safety or for regulatory purposes; or

(ii) the ability to make available non-personal electronic health data through the control of the technical design of a product and related services, including by registering, providing, restricting access to or exchanging such data;

Depending on these two scenarios, two different data provisioning models are considered in EUCAIM (*Figure 3 and Figure 4*):
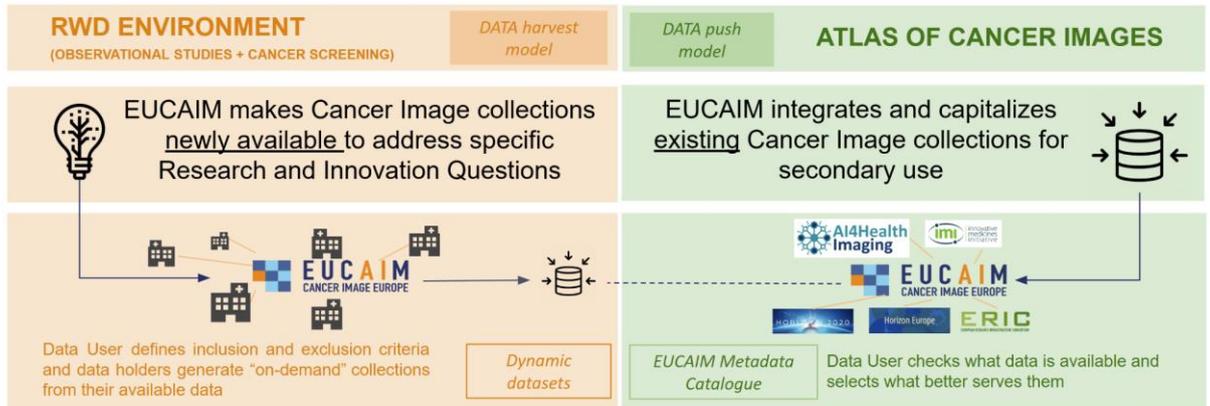
Figure 3. EUCAIM data provisioning models
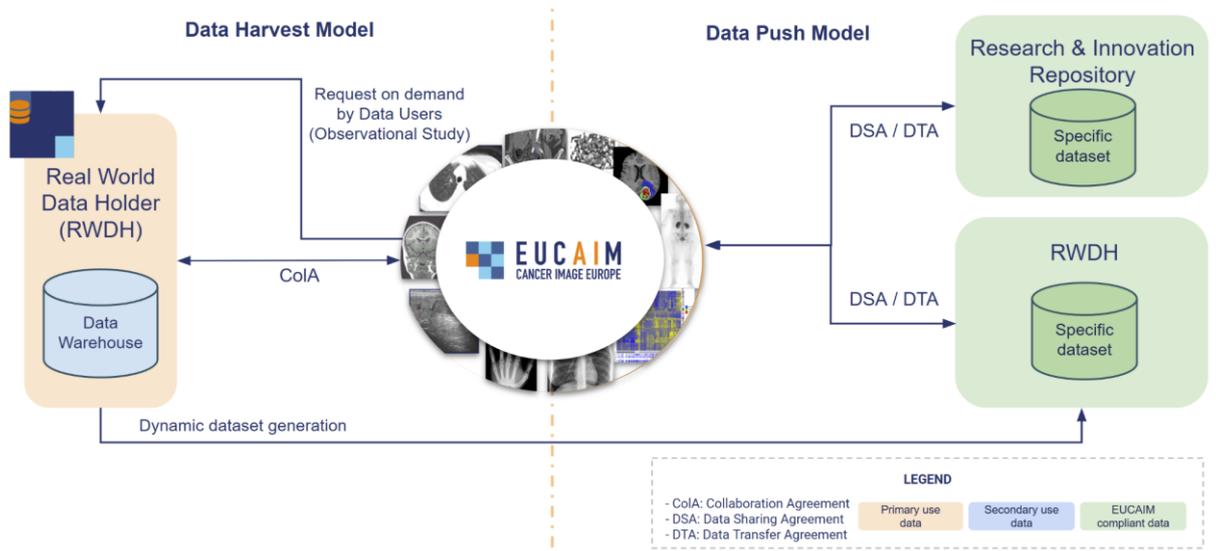
Figure 4. EUCAIM data provisioning models workflow

- **Research environment data push model:** to populate the Atlas of Cancer Images, EUCAIM will integrate existing Cancer Images datasets coming from secondary use. In this model, DHs will register their datasets in the EUCAIM Public Metadata Catalogue, making them searchable and available upon request by Data Users *(this user role is defined in Section 2.2)*.

- **Real World Data harvest model**: in this case, EUCAIM will make Cancer Image datasets newly available to address specific Research Questions through new observational studies. Data Users will define inclusion and exclusion criteria and RWDHs will generate "on-demand" datasets from their available data. Unlike the data push model, RWDHs will not register their entire Data Warehouse (DW) to the EUCAIM Public Metadata Catalogue. Instead, they will be approached individually whenever there is a new observational

study. If they opt to participate, they will prepare the necessary anonymized datasets within their DW and share them with the federation either through a federated node or by uploading them to the Reference Nodes.

Additionally, sharing or transferring some pre-existing repositories to EUCAIM and registering them in the public catalogue will have the opportunity to do so. This fosters broader collaboration and enhances the secondary use of health data.

As mentioned before, DHs will become a federated node (by signing a Data Sharing Agreement (DSA)) or will transfer their data to the Reference Nodes (by signing a Data Transfer Agreement (DTA)). Additionally, to ensure transparency, facilitate data reuse, and respect governance in EUCAIM, DHs will provide data (in both cases) under specific licences detailing usage conditions and/or national law restrictions, such as permissions for commercial use, data visibility, or downloadability. Based on these licences, the Access Committee (*defined in Section 6.9*) will evaluate Data Users' requests. If the requests clearly comply with the terms of use described, the Access Committee will have the authority to grant access to the data. The DH has the right to be notified but does not need to be consulted beforehand, unless it considers it necessary to be consulted on any relevant aspect before granting the data permit.

DHs are responsible for exposing clinical data and imaging metadata in a standardised manner, using the EUCAIM Common Data Model (CDM) schema and standard vocabularies (EUCAIM Hyper-ontology), developed by WP5, so that the data can be queried and used in federated learning processes to train AI models requested by the researchers and innovators. However, to facilitate as much as possible their participation in EUCAIM, three technical tiers of compliance with EUCAIM have been defined, each with distinct levels of requirements:

1. Tier 1: Low compliance: datasets are searchable in the Public Metadata Catalogue.
2. Tier 2: Medium compliance: datasets are accessible through Federated Query.
3. Tier 3: Full compliance: datasets are available for Federated Processing Services.

Detailed information on the tiers and requirements for DHs can be found in D4.4 Final Rules for Participation Report.

To ensure proper legal compliance with data privacy and security regulations, DHs must verify that their clinical and imaging data meet all applicable standards and legal requirements defined by *D4.4 Final rules for participation report.*

As part of the DH role, the Local Data Manager, an authorised technical expert or a team of experts on the provider's site, is responsible for installing, configuring, operating, and maintaining the local services that support the federation. This also includes managing the ingestion of data into the Reference Nodes when applicable. The Local Data Manager will receive support from the EUCAIM Helpdesk and all relevant teams.

Example of a **Data Holder**: One research institute that will provide imaging and clinical data to the platform, collected under several research projects it has performed. The

entity is willing to share data altruistically for research and innovation purposes, aligning with its datasets sustainability models. With extensive data accumulated over the last years, the site can share and/or transfer datasets that meet EHDS-R standards.

2. **Software Provider (SP):**

The Software Provider refers to entities (such as startups, enterprises, research institutions, government agencies, and non-profit organisations) that would like to contribute processing software (SW), services, data preprocessing software, or applications they have developed to the Marketplace for use in federated processing and/or data preprocessing within the platform. The SP must ensure that their developed SW comply with EUCAIM technical requirements and standards, defined by WP5 and WP6 in the working document entitled "*EUCAIM software on-boarding guideline*". When required, the SP must prove that the SW have been developed in accordance with the risk analysis and data protection methodologies by design and by default imposed by GDPR, both from the point of view of risks to the rights of data subjects and security, and that, when mandatory, it has developed a Data Protection Impact Assessment. Compliance with these obligations must be proven with appropriate documentary evidence. In the case of AI SW, it shall be evidenced that their development takes into account the principles of human centric approach, guarantee of fundamental rights, human oversight and risk focused development. Moreover, in the future, software that is provided and may qualify as an artificial intelligence information system or that incorporates embedded AI should specify its risk level rating under the EU Artificial Intelligence Act (AIA) and ensure compliance.The SW development shall be aligned with the recommendations of the High-Level Expert Group as specified in the Ethics Guidelines for Trustworthy AI, and with the standards applicable in the EU. Additionally, Software Providers must offer their SW with clear instructions and documentation on how to use them, any prerequisites required, and any limitations or restrictions that may apply. They must also ensure that their SW are compatible with the data stored in the federated nodes and that they do not violate any data privacy or security regulations or ethical and legal requirements of the platform. The Software Provider must comply with the terms and conditions of the network, including any requirements for intellectual property rights, licensing, or revenue sharing.

Example of a **Software Provider**: A startup that provides, after agreement, an AI explainability software which helps data scientists, business analysts, and developers to explain, analyse, and monitor the behaviour of AI models in real-time. Their platform offers model explainability, bias detection, and mitigation, and data validation SW.

3. **Data User (DU):**

The Data User refers to any natural or legal person who wants to make use of the data that is made accessible through the EUCAIM infrastructure for research, development and innovation purposes[9]. To this end, DUs explore the public catalogue of available (meta)data and eventually request access to data and process them using the software

---

[9] According to EHDS-R health data user' means a natural or legal person, including Union institutions, bodies, offices or agencies, which has been granted lawful access to electronic health data for secondary use pursuant to a data permit, a health data request approval or an access approval by an authorised participant in HealthData@EU

available in the platform and/or their own software. This data access request by the DU must be made under a R&I project approved by a competent Ethics Committee of the country where the project is going to be conducted.

As explained before, DUs will be able to request data from two different environments:

- Research environment: requests will be made to access data sourced from already existing research repositories exposed in the public catalogue.
- Real World Data environment: requests will be made under a proposal for an observational study using real-world data. DUs will define inclusion and exclusion criteria and RWDHs will generate "on-demand" datasets from their available data.

Depending on the type of request, different documents will be needed, as further explained in the Data Access Requests section. Both types of requests will undergo evaluation by the Access Committee (*defined in* [Section 6.9](#)).

Within the different purposes for requesting data access and software, we can group them in two profiles, depending on the intended purpose:

- **Research & Academia**: DUs, such as clinicians, researchers, research teams, or education professionals in the health and care sectors. These users typically request data to conduct studies, research, or analysis with the intention of generating new knowledge in the field of medicine and publishing the findings. Their focus is on discovering new insights, patterns, or trends in the data, and their work often contributes to academic or scientific research. Data Users may require specialised access to certain datasets or software that enable complex data analysis.
- **Innovation**: This profile includes DUs such as data scientists and innovators focused on the development and enhancement of AI algorithms. These users typically request data to contribute to public health or health technology assessment, or to ensure high levels of quality and safety of health care, of medicinal products or of medical devices, with the aim of benefiting the end-users, such as patients, health professionals and health administrators. Their activities encompass the development and innovation of products and services, as well as the training, testing, and evaluation of algorithms, including those used in medical devices, AI systems, and digital health applications.
- **Policy making and regulatory activities:** This profile includes DUs such as policymakers and regulatory authorities who support public sector bodies, Union institutions, agencies and bodies, including regulatory authorities, in the health or care sector to carry out their tasks defined in their mandates. Policymakers typically gather information through consultations and research to develop policies or strategic frameworks addressing specific issues.

Examples of Data Users:

- **Data User-Researcher**: A DU is leading a research project related to prostate cancer, with one of the objectives being the treatment allocation based on the analysis of baseline Magnetic Resonance images at the time of diagnosis of the disease. The research team will incorporate Artificial Intelligence software and experience in interpreting the results obtained, validating and then applying them in a clinical setting for routine clinical practice.

- **Data User-Innovator**: A group of Data Scientists is developing AI software to analyse health images and related clinical data on the most prevalent cancers in Europe in a federated manner. They have an initial model that they aim to improve using new federated data. They seek high-quality, labelled data and do not accept unstructured data or data lacking a logical folder structure for their federated processing.

- **Data User-Policy making:** A policymaker in the health sector is leading an initiative to analyze data on a given cancer profile in several EU countries within EUCAIM. The objective is to evaluate key aspects of health system performance, such as early detection rates, treatment durations, types of treatments, and survival outcomes. By comparing these factors across different healthcare systems, the policymaker aims to identify best practices, inform policy decisions, and improve cancer care standards across the EU.

Finally, it is important to highlight that any of the User Roles explained in this section will have the possibility to join or create, in case it does not already exist, a virtual organisation within the platform: a **Research Community (RC)**. Research Communities are formal multidisciplinary teams composed of researchers, innovators, clinicians, data scientists, engineers, AI specialists, etc, dedicated to exploring, innovating and advancing a specific field or topic, to improve the role of imaging in healthcare, fostering collaborations and improving outcomes. This virtual organisation encompasses the previously described roles and will function accordingly, based on the functions the team members wish to perform. The Main Tasks of the RC (such as hepatocellular carcinoma, neuroblastoma or breast screening) are:

- Gathering large annotated Data Collection.
- AI Model designing, training, and optimization to integrate data and recognize patterns link to outcomes.
- Validation of models and benchmarking against established methods. Integration into clinical workflows for clinical decision support.
- Collaboration with regulatory bodies to comply with regulations and standards.
- Collaborative work and publications, including the general population.

Example of **Research Communities**: the Pediatric Cancer RC focuses on enhancing diagnosis and treatment through innovative AI applications in medical imaging. This dedicated community within EUCAIM originated from the consortium of the finalised project PRIMAGE[10]. The outcomes and data generated from PRIMAGE have been successfully transferred to the Central Hub, ensuring their continued usability and alignment with the overarching goals of the EUCAIM initiative. This community actively fosters collaboration and innovation in AI-driven healthcare solutions by engaging in new research projects and forming strategic partnerships.

Management Profiles

In addition to user roles, there are some management profiles, within the project consortium, who will also interact with the platform, described as follows:

1. **Platform Manager**: The Platform Manager is a technical team of experts who will coordinate and operate the core services of the EUCAIM platform. The Platform

---

[10] PRIMAGE project: https://www.primageproject.eu/

Manager is responsible for managing and maintaining the underlying EUCAIM technical infrastructure, including the Reference Nodes, servers, databases, and other resources. The Platform Manager manages user accounts and access permissions, deploys applications, and services, uploads new SW to the marketplace (provided by Software Providers), and ensures their proper integration into the platform. They support the orchestration of federated processing, working with DHs and SPs to integrate (meta)data, SW, and services, and ensuring that DUs queries are properly executed. As a team of experts, it is possible to have multiple platform managers assigned to different roles such as security and data privacy, administration, development, system management, etc. Additionally, the Platform Manager offers specialised support teams, who provide user support, respond to inquiries, provide documentation, and troubleshoot issues that arise with the platform, to the different user roles.

> Example of a **Platform Manager**: A system administrator is a member of the team employed by the institution that operates the Reference Nodes who has technical and project management skills, as well as specific knowledge of data management principles, including data storage, data integration, and data governance, and understanding data privacy and security regulations. The platform manager might have a basic understanding of machine learning and artificial intelligence, including the principles of federated learning and how it works, and be able to support the orchestration of this process.

2. **Governing Body**: The Governing Body is the decision-making board of EUCAIM that plays a crucial role regarding the coordination, governance, and operation of the federated infrastructure. The Governing Bodies and their decision making are subject to modifications during the project's life span. The bodies are described below, in Section 6. Operative boards managing the Central Hub.

> Examples of Governing Body:
>
> - **The Access Committee (AC)** (responsible for assessing data access requests and recommending acceptance or rejection to the Steering Committee) receives access requests to specific datasets of the Atlas of Cancer Images from Data Users. This specific Governing Body receives an email notification that a new request has been submitted through the Negotiator service, and initiates the established procedure to perform an evaluation (depending on the type of agreement signed with the Data Holder organisation). Any rejection is fully justified by the AC in writing form, according to objective criteria defined in the internal procedures of the AC.
>
> - **The EUCAIM SW Validator** is responsible for ensuring that software and services meet privacy, security, integrity, and ethical standards. When a SP applies to make software available on the marketplace, the validator checks for risks, ensures data protection, fairness, and transparency, and verifies compliance with technical and security protocols on the EUCAIM platform.

3. User Stories

User Stories are descriptions of full interactions of a User Role within the EUCAIM platform, described in natural language. User Stories define in general terms the needs, restrictions,

performance limitations, desired features, innovation capabilities and business models for the repository.

The project has identified 26 User Stories that identify actions that could be performed by the actors from the above user profiles. In the following tables, these User Stories are listed and graphically represented for each of the roles and they are further detailed with specific examples and accompanied by Unified Modeling Language (UML) diagrams. Only the User Stories related to the Governing Body, which do not involve any technical interaction process with the platform, are not accompanied by a UML diagram.

Lists and diagrams

*Table 1. List of User Stories related to **Data Holders.***

#	User Story	User Roles	Mapping
usDH1	Application to join the federation	Data Holder	Local Data Manager with duly representation
usDH2	Setup of the data node to connect to EUCAIM for acting as federated nodes	Data Holder	Local Data Manager
usDH3	Dataset creation by a RWDH for an specific observational study request	Only for Real World Data Holder	Local Data Manager
usDH4	Data preparation on premises: anonymization, quality control and curation (legal/technical)	Data Holder	Local Data Manager
usDH5	Register metadata to the Public Catalogue (upload/harvest)	Data Holder	Local Data Manager
usDH6	Ingest data on the Reference Nodes	Data Holder	Local Data Manager
usDH7	Manual or semi-automatic data annotation using the reference node DICOM viewer	Data Holder	Local Data Manager
usDH8	Monitoring the access of the dataset	Data Holder	Local Data Manager

*Table 2. List of User Stories related to **Software Providers.***

#	User Story	User Roles	Mapping
usSP1	Application of a software to the EUCAIM	Software Provider	AI developer

#	User Story	User Roles	Mapping
	Marketplace		
usSP2	Monitoring the access of software	Software Provider	AI developer

*Table 3. List of User Stories related to **Data Users**.*

#	User Story	User Roles	Mapping
usDU1	Exploration of datasets from the Public Catalogue	Data User	Researcher/ Data Scientist/ Innovator
usDU2	Federated query of aggregated data in the datasets	Data User	Researcher/ Data Scientist/ Innovator
usDU3	Request access to existing datasets or propose building a new observational study using RWD through the negotiator	Data User	Researcher/ Data Scientist/ Innovator
usDU4	Get an overview of the datasets to which they have been granted access	Data User	Researcher/ Data Scientist/ Innovator
usDU5	Exploring the software available in the federation and their provenance	Data User	Researcher/ Data Scientist/ Processing Requester
usDU6	Run a federated (distributed) experiment on multiple data-sets from the User's Datasets Library using software from EUCAIM Marketplace	Data User	Researcher/ Data Scientist/ Innovator
usDU7	Training an AI model on multiple data-sets from the User's Datasets Library leveraging EUCAIM federated architecture	Data User	Researcher/ Data Scientist/ Innovator
usDU8	Monitoring job status in the Federated Processing architecture	Data User	Researcher/ Data Scientist/ Innovator
usDU9	Retrieve results from a federated experiment	Data User	Researcher/ Data Scientist/ Innovator
usDU10	Retrieve an AI model trained in EUCAIM datasets	Data User	Researcher/ Data Scientist/ Innovator

*Table 4. List of User Stories related to the **Governing Body**.*

#	User Story	User Roles	Mapping
usGB1	Evaluate and accept/reject datasets from Data Holders	Governing Body	Access Committee
usGB2	Evaluate and accept/reject software from Software Providers	Governing Body	SW Validator
usGB3	Evaluate and accept/reject a data access request application	Governing Body	Access Committee
usGB4	Evaluate and accept/reject a RWD observational study application	Governing Body	Access Committee

*Table 5. List of User Stories related to the **Platform Manager**.*

#	User Story	User Roles	Mapping
usPM1	Uploading of software to the EUCAIM Marketplace	Platform Manager	Platform Manager, SW Validator
usPM2	Platform services administration	Platform Manager	Platform Manager

Examples of User Stories related to Data Holders

Example of **Application to join the federation** User Story**:**

Table 6. usDH1.

#: usDH1	Story name: Application to join the federation	
Related WP: WP2	**Triggered by User Role (Fictitious Name):** Data Holder (Raul)	**Related User Roles (Fictitious Name):**
Overview Description: Raul is a representative of a research organisation that seeks to join the EUCAIM federation and contribute imaging and clinical data to the platform. The provided datasets will be used in federated learning processes to train AI models by DUs, within the federated network. Raul's organisation has accumulated a large amount of imaging and clinical datasets over the past 10 years, as they have conducted multiple research projects in this field, within the European Framework. Compliance with data privacy and security regulations, including GDPR, is a top priority for them.		

To apply for data sharing with the federation, Raul accesses the Dashboard and reviews the process and conditions. He follows the documentation available in the Data Holders area to register his organisation. Raul submits the Expression of Interest, where he highlights the research organisation's capacity to provide valuable datasets while adhering to stringent data privacy, anonymisation, curation and data quality standards. The application demonstrates their commitment to maintaining high-quality and up-to-date datasets and their plan to align with EUCAIM's Tier model, maximising interoperability and usefulness in federated learning processes. Moreover, his entity either possesses or plans to acquire the necessary computational resources and processing capabilities to become a federated node.

By joining the federation, Raul's organisation will become a federated node, allowing DUs to access and utilise their available clinical and imaging datasets upon request. This collaboration contributes to the advancement of research and development of AI models that can ultimately benefit the medical community and patient care.

Preconditions:

- Datasets availability.
- Identification of a legal representative (signatory).
- Knowledge of participation rules.
- Processing of supervisory procedures by the Access Committee and approval by the Steering Committee (see usGB stories).

Postconditions:
- Signing of CoIA and DSA/DTA (Data Sharing Agreement/Data Transfer Agreement).
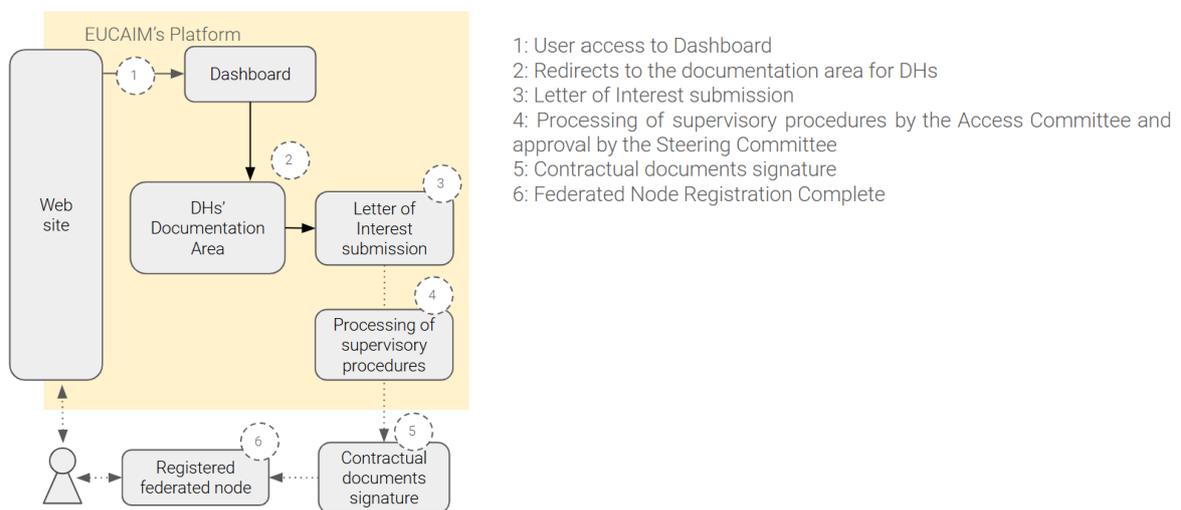
1: User access to Dashboard
2: Redirects to the documentation area for DHs
3: Letter of Interest submission
4: Processing of supervisory procedures by the Access Committee and approval by the Steering Committee
5: Contractual documents signature
6: Federated Node Registration Complete

Figure 5. UML diagram of usDH1

Example of Setup of the data node to connect to the federation User Story:

Table 7. usDH2.

#: usDH2	**Story name:** Setup of the data node to connect to EUCAIM for acting as federated nodes	
Related WP: WP2, WP4, WP5	**Triggered by User Role (Fictitious Name):** Local Data Manager (Carol)	**Related User Roles (Fictitious Name):** Data Holder (Raul) / Platform Manager (Ignacio)

Overview Description:

Carol, as the Local Data Manager in Raul's Data Holder centre, which has recently joined the EUCAIM federation (meaning that the centre has received the approval from the Governing Body), is responsible for the configuration of the data node that will be connected to the network, which will be able to execute federated processing.

To do this, she accesses the Data Holder workspace corresponding to her organisation through the Dashboard, where she can find the specific documentation on how to set up the federated node in terms of technical requirements, both regarding software and hardware, and instructions on how to install and configure the necessary software components and dependencies to connect the node to the network. She accesses the Marketplace and searches for the software needed to set up the connection, downloads it and installs it in the node, as for example the *Federated Processing Daemon*. In addition, she can contact the Technical Support team through the Helpdesk, to receive support for any of the processes that she needs to perform.

Once she finishes the software and hardware installations and configurations, she contacts Ignacio, the Platform Manager, to test the integration of the node with the rest of the platform, validating the connectivity, and launching multiple test jobs with validation/test datasets (datasets containing imaging and associated clinical data). After due verifications, Ignacio issues a final positive report, which is incorporated into the documentation.

Preconditions:
- Access to necessary hardware and software for configuring the node.
- Access to the Dashboard documentation.

Postconditions:
- Successful configuration of the node and connection to the EUCAIM network.
- Validation of node connectivity with the Core Services.
- Conducting integration tests with the rest of the Platform.
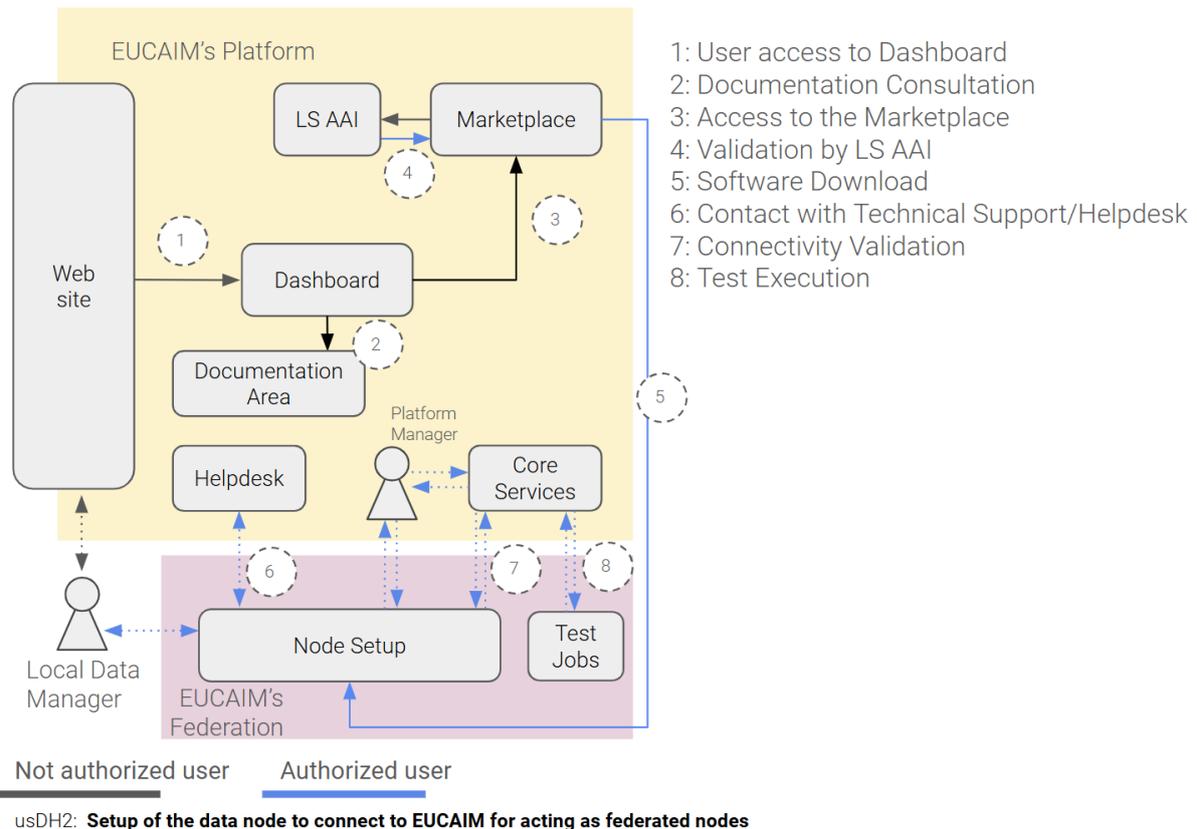
Figure 6. UML diagram of usDH2

Example of **Dataset creation by a RWDH for an specific observational study request** User Story:

Table 8. usDH3.

#: usDH3	Story name: Dataset creation by a RWDH for an specific observational study request	
Related WP: WP2, WP3, WP4, WP5	**Triggered by User Role (Fictitious Name):** Local Data Manager in a Real World Data Holder (Irene)	**Related User Roles (Fictitious Name):** Platform Manager (Ignacio)
Overview Description: Irene is the Local Data Manager at a hospital with RWD, who has recently signed a Collaboration Agreement with EUCAIM. At the hospital, Irene is building a DW where the hospital's primary health information systems undergo an ETL process daily and are stored in a structured database, following the EUCAIM CDM and standards.		

19

This setup allows Irene to dynamically create datasets with newly available data for secondary use in accordance with the characteristics of the published and available dataset catalogue.

When a DU proposes a new observational study through the Negotiator, Irene is contacted by the AC. She is asked if she would like to participate as a partner in the project and which data she can provide. Since Irene's DW is already mature and well-structured, she can quickly respond to the request, specifying the number of cases she can provide.

Once selected as a partner for the new observational study, Irene creates the dataset following the specific inclusion criteria of the project.

With high-quality RWD readily available for secondary use, Irene helps accelerate cancer research and contributes to populating the Atlas of Cancer Images with new data from her hospital.

Preconditions:

- A Collaboration Agreement and DSA/DTA with EUCAIM has been signed with due guarantees of legal compliance, particularly with regard to GDPR and EHDSR
- The hospital's Data Warehouse is already in place

Postconditions:

- The hospital is selected as a partner for the new observational study
- The Hospital specifies the data that can be provided
- New datasets are created with RWD

1: RWDH accesses to the Dashboard
2: RWDH signs ColA with EUCAIM
3: RWDH has a DW and a federated node in place
4: DU proposes a new observational study
5: AC receives the proposal of new observational study via negotiator and contacts with the RWDH
6: RWDH replies and is accepted in the observational study
7: RWDH prepares a new dataset coming from RWD

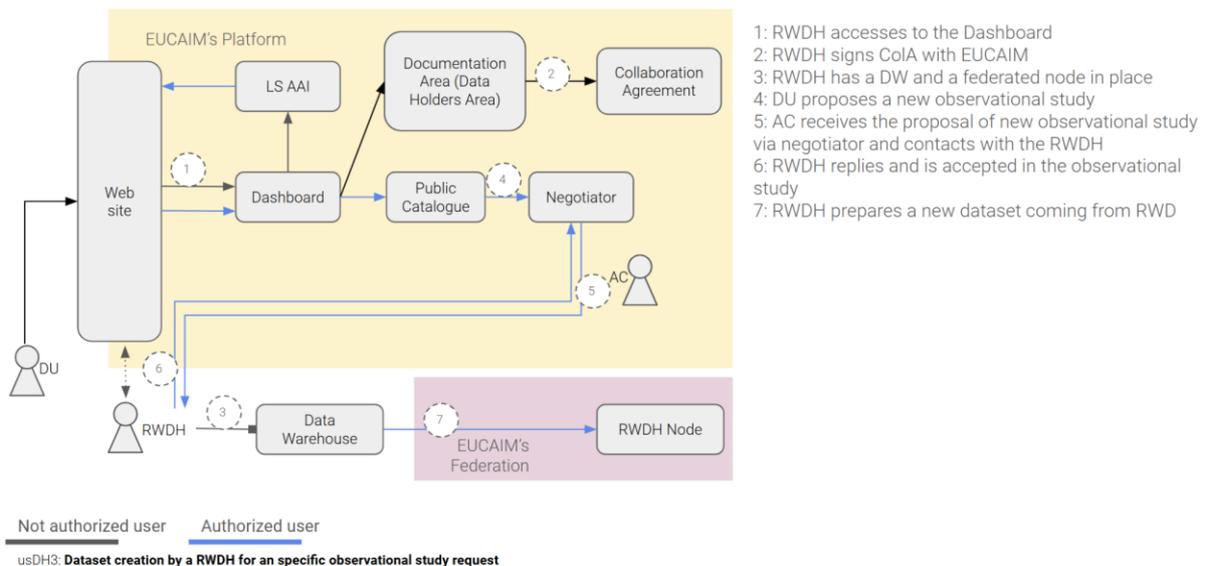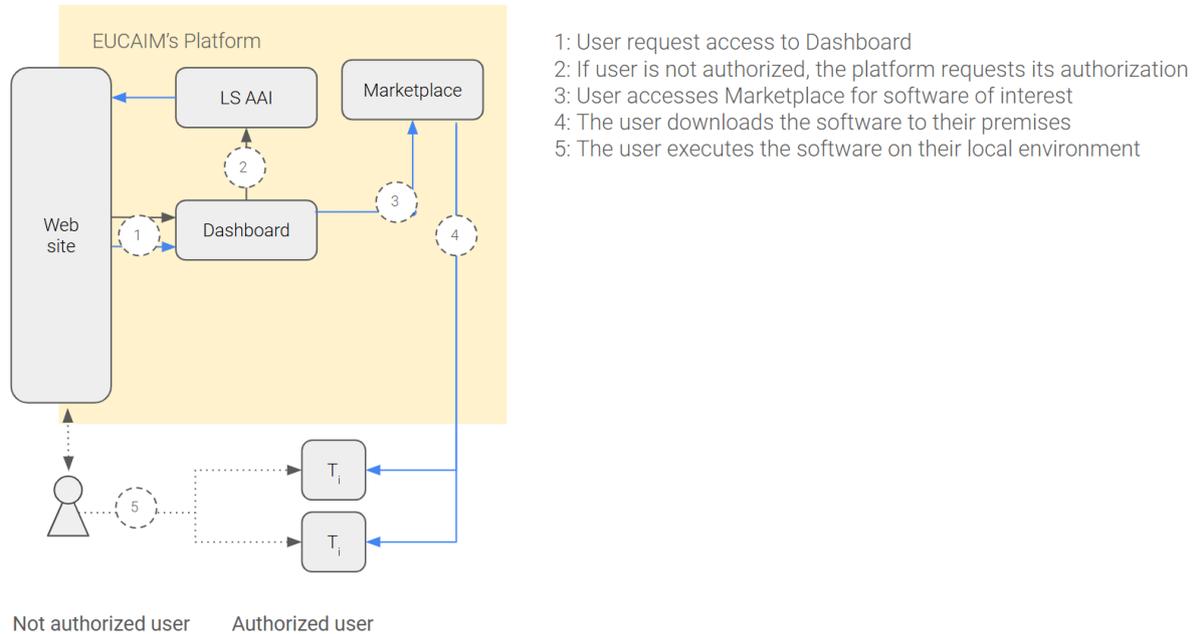usDH3: **Dataset creation by a RWDH for an specific observational study request**

Figure 7. UML diagram of usDH3

Example of **Data preparation on premises: anonymization, quality control and curation (legal/technical)** User Story:

Table 9. usDH4.

#: usDH4	Story name: Data preparation on premises: anonymization, quality control and curation (legal/technical)	
Related WP: WP2, WP3, WP5	Triggered by User Role (Fictitious Name): Local Data Manager (Carol)	Related User Roles (Fictitious Name):

Overview Description:

Carol is the Local Data Manager in a Data Holder centre that has recently joined the EUCAIM federation (meaning that the centre has received the approval from the Governing Body). She has already completed the process of "Setting up the data node", so the node is able to interact with the EUCAIM infrastructure.

In order to prepare her dataset(s), she has to apply some pre-processing software for it to be compliant with the EUCAIM data requirements. The aspects to be processed/evaluated are:

- Anonymization of the datasets
- Re-identification risk assessment
- Quality assurance (data integrity and data coherence)
- FAIR compliance evaluation
- Data curation (e.g. conversion to EUCAIM data standards and structuring, and data completeness)
- Transform data using an ETL tool to align with the EUCAIM Common Data Model (CDM), because she wants her datasets to comply with Tier 2 (partial transformation) or 3 (complete transformation) (*optional*)

Carol's datasets must meet the requirements of the specific Tier before their ingestion into the Reference Node or being made available in the federated node. She can accomplish this either by utilizing software she is already familiar with or by leveraging those provided by EUCAIM to address specific needs. To do it, after reviewing the guidelines available in the training material, she accesses the Data Holder workspace corresponding to her organisation through the Dashboard, where she accesses the tool marketplace and selects the list of data preparation software that she wants to download and run on-premise over the datasets that she selected. The processing pipeline and necessary software depend on the characteristics of the dataset and the intended Tier. She receives the necessary information to be able to deploy the software locally, avoiding federated processing or any kind of orchestration coming from the Central Hub at this point.If the Platform Manager becomes aware of any risk, such as inadequate anonymisation, he/she will notify Carol so that appropriate safeguards can be put in place to address the identified risk.

Preconditions:

- Completion of the usDH2 to enable interaction with the EUCAIM infrastructure, in case of sharing data as a federated node.
- Availability of datasets extracted from the Local DW or from an existing repository coming from a R&I project.
- Availability of the software and hardware resources for deployment of the required preprocessing software.
- Access to the training material in which the data preparation process is described.

Postconditions:

- Deployment of some pre-processing software locally to check the quality and risks of the datasets, and to curate them to comply with EUCAIM requirements.
- Generation of a report allowing evaluation of the quality, risk analysis and degree of preparation of the datasets.

1: User request access to Dashboard
2: If user is not authorized, the platform requests its authorization
3: User accesses Marketplace for software of interest
4: The user downloads the software to their premises
5: The user executes the software on their local environment

usDH4: **Data preparation on premises: anonymization, quality control and curation (legal/technical)**

Figure 8. UML diagram of usDH4

Example of **Register metadata to the Public Catalogue** User Story:

Table 10. usDH5.

#: usDH5	Story name: Register metadata to the Public Catalogue (upload/harvest)	
Related WP: WP2, WP7	**Triggered by User Role (Fictitious Name):** Local Data Manager (Carol)	**Related User Roles (Fictitious Name):**
Overview Description: Carol, as the Local Data Manager, is responsible for uploading metadata to the Public Catalogue. Recently, Carol's center agreed to participate in sharing data with the Federation. As part of this collaboration, she has prepared her curated and anonymised datasets following the EUCAIM standards, so they are ready to be shared with the platform through her federated node. Carol now focuses on preparing the metadata at the datasets level. The goal is to register it into the Public catalogue, which showcases metadata from diverse datasets following the DCAT-AP specifications. It also aims to make the available data interoperable with the rest of the federated data.		

To ensure data security, the EUCAIM platform implements an access system with different levels of privileges. The Public Catalogue is available to anonymous users, offering the view of the metadata and basic search options at the dataset level.

In order to register the metadata, Carol fulfills a template with the required metadata, provided by EUCAIM. She takes care of giving the information ensuring that the descriptions of the datasets are comprehensive and accurate, following the EUCAIM specifications. This includes details about the dataset's type, provenance and access conditions, among others. Carol's registration will be validated and confirmed by the corresponding person in charge of the Platform Manager team.

Preconditions:

- Valid user account, registered and validated through the authentication and authorization service (AAI).
- Access to the Public Catalogue Service with permission to upload the metadata from the datasets.
- Knowledge of the metadata requirements and standards set by EUCAIM.
- Validation of the registered metadata.

Postconditions:
- Metadata registration of the dataset in the Public Catalogue.
- Dataset metadata is discoverable by anonymous users.

1: User access to Dashboard
2: Validation by LS AAI
3: Catalogue Service Access
4: Metadata Preparation
5: Metadata Uploading & Organization
6: Public & Authenticated Access

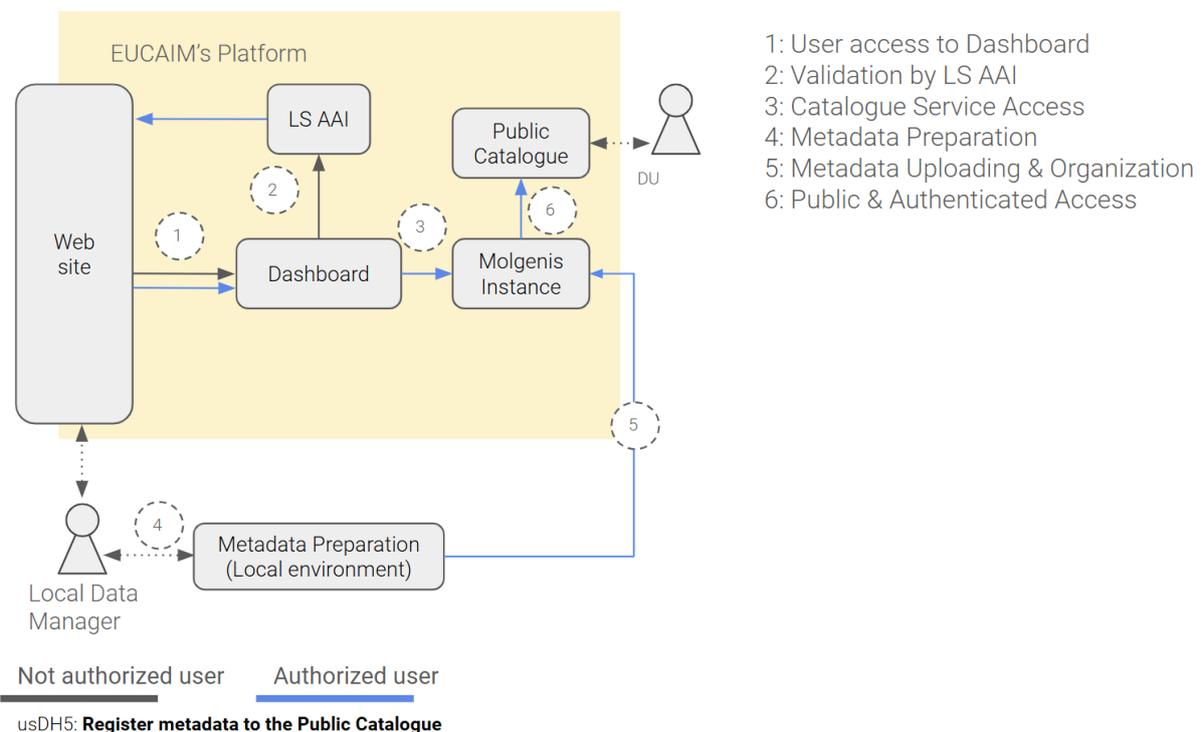usDH5: **Register metadata to the Public Catalogue**

Figure 9. UML diagram of usDH5

23

Example of **Ingest data on the Reference Nodes** User Story:

Table 11. usDH6.

#: usDH6	Story name: Ingest data on the Reference Nodes	
Related WP: WP2, WP7	**Triggered by User Role (Fictitious Name):** Local Data Manager (Carol)	**Related User Roles (Fictitious Name):** Data Holder (Raul)

Overview Description:

Carol also aims to populate the Reference Nodes of EUCAIM with some of their dataset, as part of their research project's sustainability plan. The Reference Nodes host data (both medical imaging and associated clinical data) coming from the DH. To do so, data must be curated and fully anonymised before being uploaded into the repository.

Before Carol starts uploading the data into the Reference Nodes, her colleague Raul, as a Data Holder, has extracted all information from the different clinical databases involved in the research project and Carol has performed the actions of curation and anonymisation. Once everything is prepared and they have the approval of the Governing Body, Carol starts uploading the studies with the associated clinical data (if available).

For uploading cases, Carol has two options. The first one is an in-bulk ingestion, she can use a specific tool (In-Bulk Ingestion tool) installed at her clinical centre, that connects directly the local databases with the Reference Nodes through the Data Access/Ingestion Service. The second option is a Case-by-case ingestion tool, that allows uploading data using an intuitive user interface. This last option is best suited for new datasets containing RWD for observational studies. Her datasets will be placed in a quarantine state so they will be validated by the technical and legal corresponding boards and, only after validation, the datasets can be released on the platform when all the requirements are complied. A report from this evaluation will be generated by the corresponding boards in charge. Additionally, the DTA/DSA will specify that the DHs are responsible for any issues, such as the presence of personal data.

Preconditions:

- The Reference Nodes are accessible and ready to receive data.
- There is connectivity between the local node and the Reference Nodes.
- Datasets have been prepared and curated locally.
- Governing Body approval.
- Valid user account, registered and validated through the AAI.
- Data Transfer Agreement (DTA) in place.

Postconditions:
- The Reference Nodes of EUCAIM is populated with curated medical images and clinical data if it is anonymized. If not, it will be placed in a quarantine state and the dataset must be improved and processed before its usage, with the corresponding Data Processing Agreement (DPA).
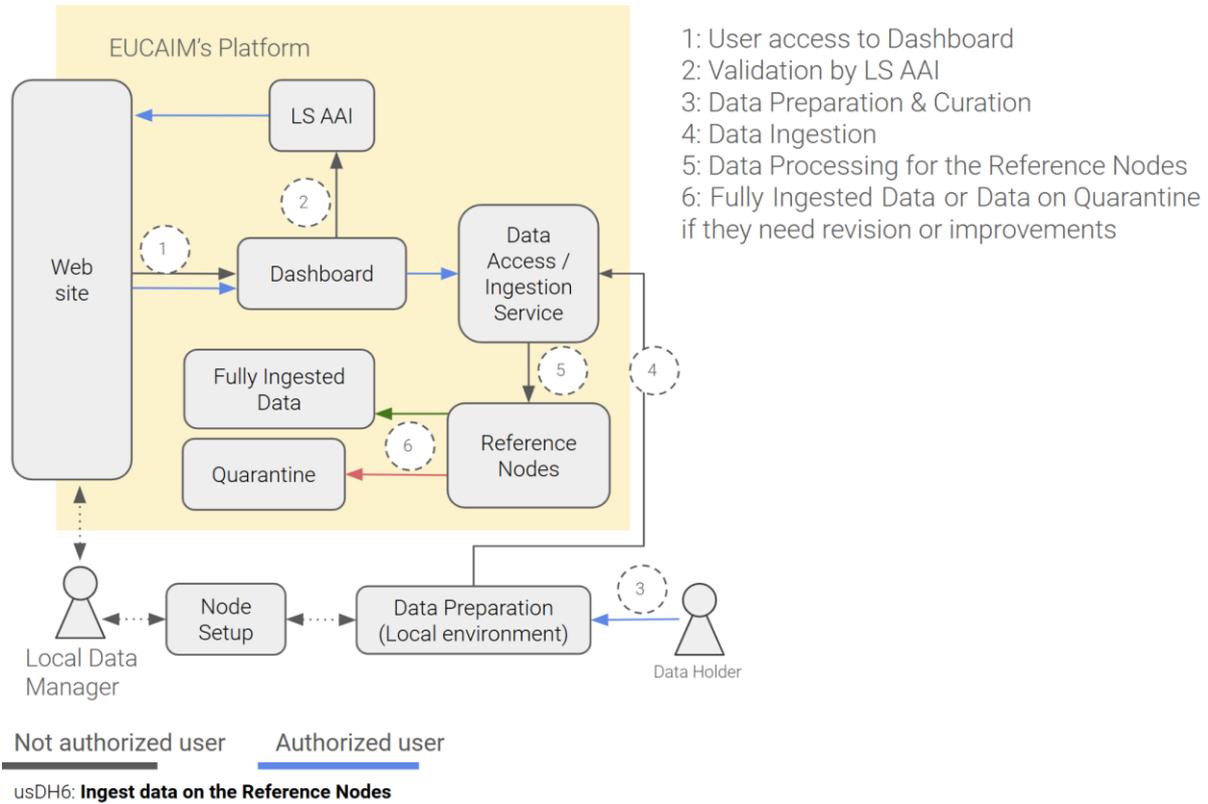
Figure 10. UML diagram of usDH6

Example of Manual or semi-automatic data annotation using the reference node DICOM viewer User Story:

Table 12. usDH7.

#: usDH7	**Story name:** Manual or semi-automatic data annotation using the reference node DICOM viewer	
Related WP: WP2, WP4, WP5	**Triggered by User Role (Fictitious Name):** Data Holder (Raúl)	**Related User Roles (Fictitious Name):** Local Data Manager (Carol), Platform Manager (Ignacio)
Overview Description: Raúl, is a radiologist in a Data Holder centre that has recently ingested a dataset in the Reference Node. He now wants to use the automatic segmentation software provided in the Reference Node DICOM Viewer to annotate the images that were previously uploaded by her colleague Carol. Raúl begins by logging into the system through the AAI to securely access the Data Holders' workspace. From here, he accesses the DICOM Web Viewer and visualises all DICOM		

images to which he has access permission. The web viewer includes a set of annotation software, allowing him to manually segment the images as needed. Additionally, built-in segmentation algorithms can be applied to pre-segment the tumours or other anatomical regions, with the option for Raúl to review and manually adjust the results for accuracy. This interactive process ensures that the annotations are precise and optimised for deep learning. The annotations performed by expert clinicians as Raúl will be permanently stored with the original dataset.

Preconditions:

- Valid user account, registered and validated through the AAI.
- The Data Holder has already ingested a dataset into the Reference Node.

Postconditions:

- The annotations generated by expert clinicians are stored together with the image data in the Reference node repository.

1: The user requests access to the Reference Node.
2: If the user is not authorized, the platform prompts them to obtain the necessary authorization.
3: The user opens studies with granted access using the DICOM Viewer.
4: An automatic segmentation tool can be executed directly from the DICOM Viewer.
5: The user can review and refine the preliminary segmentation generated by the algorithm.
6: The manual or semi-automatic annotations are saved in the central node's storage.

usDH7: **Manual or semi-automatic data annotation (using the Reference Nodes DICOM viewer)**
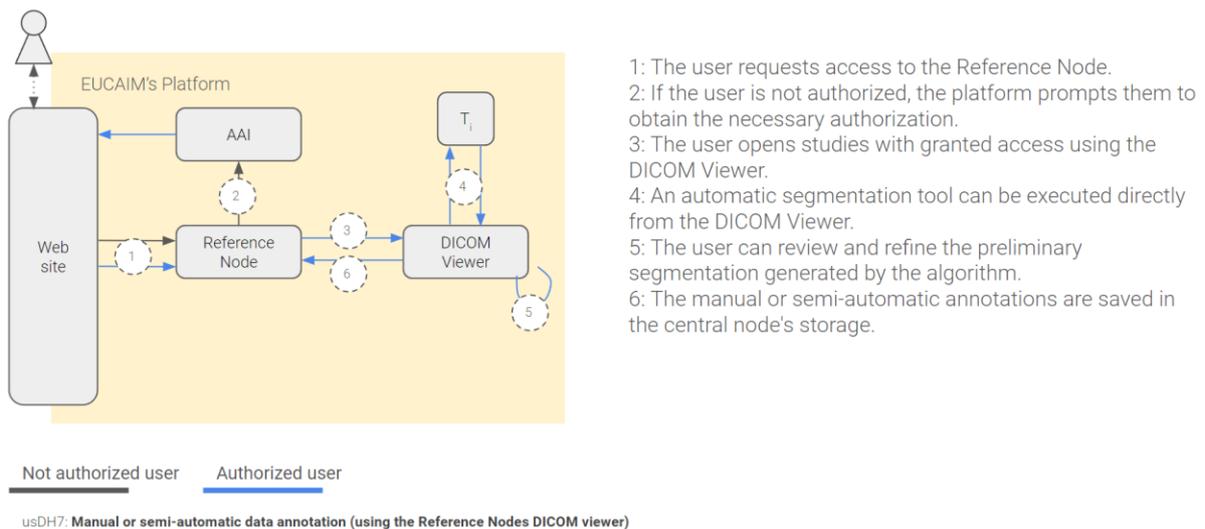
Figure 11. UML diagram of usDH7

Example of **Monitoring the access of the dataset** User Story**:**

Table 13. usDH8

#: usDH8	Story name: Monitoring the access of the dataset	
Related WP: WP4	Triggered by User Role (Fictitious Name): Local Data Manager (Carol)	Related User Roles (Fictitious Name):

Overview Description:

Carol has successfully added data to the Reference Nodes. Now, she is interested in understanding the access pattern to her uploaded data, only accessing pseudonymised information at the level of the user. This information would be very important for understanding the relevance of their data (e.g. she could compute the number of different users accessing the data, the number of accesses, the mean number of accesses for a dataset, etc.).

To achieve this, Carol accesses the Dashboard by logging through AAI.

Once logged in, Carol navigates to the Dataset Explorer of the Reference Node within the Dashboard. Using this feature, she can browse through her uploaded data and select a specific dataset to gain insights into its usage within the platform.

Preconditions:

- Valid user account, registered and validated through the AAI.
- Data is successfully available in the EUCAIM Federation.

Postconditions:
- Consult logs on accesses and processes of uploaded datasets, through one of the Central Services.
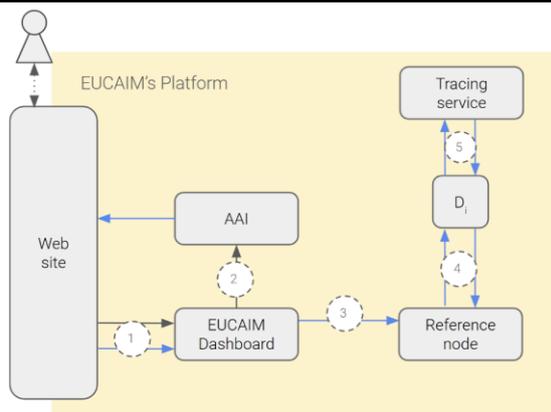
1: The user requests access to the EUCAIM Dashboard.
2: If the user is not authorised, the platform prompts the user to obtain the necessary authorisation.
3: Once it is logged in, the user navigates to the Dataset Explorer.
4: In the Dataset Explorer of the reference node, the user consults the uploaded data and selects a specific dataset (D_i).
5: For this specific dataset, the user can query usage-related information, which is obtained from the Tracing Service.
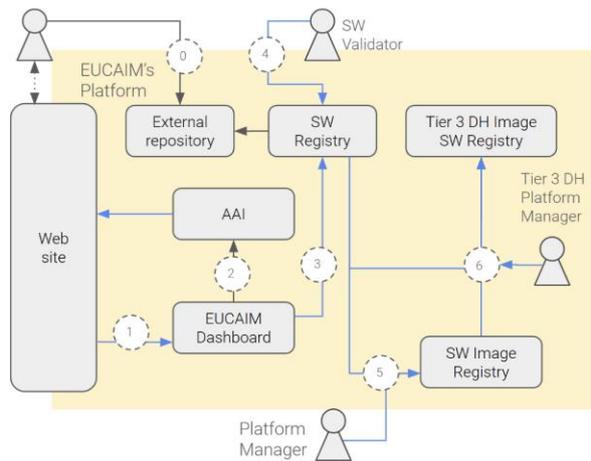
Figure 12. UML diagram of usDH8

Examples of User stories related to Software Providers

Example of Application of a software to the EUCAIM Marketplace User Story:

Table 14. usSP1.

#:	Story name:
usSP1	Application of a software to the EUCAIM Marketplace

Related WP: WP4, WP6	Triggered by User Role (Fictitious Name): Software Provider (Paolo)	Related User Roles (Fictitious Name): Platform Manager (Ignacio)

Overview Description:

Paolo, an application developer, aims to contribute to the Marketplace by applying a specialised software. His goal is to share a software that enhances image processing capabilities within the federated ecosystem.

Once Paolo has diligently developed the software, he submits a registration request to the Governing Body. To do so, Paolo follows all the steps described in the "*EUCAIM software on-boarding guideline*". During the steps defined in the guideline, multiple aspects of Paolo's software are evaluated (from conceptual to technical, from ethical and legal to integration into EUCAIM's architecture) by the EUCAIM SW Validator.

Once Paolo has followed all the steps in the guideline, his software will be registered in the Marketplace, and a containerized image of his software located at EUCAIM's SW Image Repository. Then, the nodes of the Federation will need to test the image locally before it is finally approved and accepted in EUCAIM.

By registering the software in the Marketplace, registering its image and providing all the requested documentation, Paolo aims to contribute to the collaborative and innovative environment of federated learning, fostering further advancements in the field.

Paolo's software becomes a valuable asset within the federation marketplace, enabling researchers and data scientists to leverage its capabilities in their own federated learning projects.

Preconditions:
- Access to the Marketplace registration system is available.
- The software is fully developed, tested, and compliant with EUCAIM's standards and software requirements.

Postconditions:
- The software registration request is submitted to the Governing Body for its validation.
- If it is accepted, the software becomes available on the Marketplace for researchers and innovators.

0: Prerequisite: The user has the SW and the container image description in an external repository.
1: The user requests access to the EUCAIM Dashboard.
2: If the user is not authorised, the platform prompts the user to obtain the necessary authorisation.
3: Once it is logged in, the user navigates to the SW Registry, where (s)he adds the description of the SW in the SW Registry, referencing the information from the external repository.
4: The SW validator, verifies if the SW corresponds to the application approved by the access committee and that it is compliant to the SW development guidelines.
5: The EUCAIM Platform Manager creates the container image and pulls it to the SW Image Registry.
6: Each Tier 3 Data Holder Platform Manager that could support this new SW checks its description and eventually customises the image to run it on the Tier 3 platform (e.g. by adding new environment variables).

NOTE: For readability, only the authentication of the user is shown in the diagram. Every actor performing an action that requires authorisation (blue arrows) should have acquired valid credentials before performing the actions.

usSP1: **Application of a software to the EUCAIM's marketplace**
usPM1: **Uploading of software to the EUCAIM marketplace**
usGB2: **Evaluate and accept/reject software from Software Providers**

Figure 13. UML diagram of usSP1, usPM1 and usGB2

Example of **Monitoring the access of software** User Story**:**

Table 15. usSP2.

#: usSP2	Story name: Monitoring the access of software	
Related WP: WP4	Triggered by User Role (Fictitious Name): Software Provider (Paolo)	Related User Roles (Fictitious Name):
Overview Description: Paolo, software developer, has previously uploaded a software application onto the Marketplace and is now seeking to ensure that the users are using it without encountering any issues. To accomplish this, Paolo gains access to the Dashboard by logging in using the AAI. Once successfully logged in, Paolo proceeds to navigate through the EUCAIM SW Image Repository and locate his uploaded SW in order to access detailed information about it. Within the expanded menu of the SW, he can assess how it is being used by the DUs, which is internally collected in the Monitoring Platform. Paolo has previously signed a Data Processor Contract that includes an additional non-disclosure commitment. Paolo can only deploy this function when his contribution is necessary for the operational maintenance of the software and the guarantee of its security. Paolo may under no circumstances exploit the information for its own benefit or infringe competition laws, trade secrets or intellectual property.		

Preconditions:
- Login is performed using the AAI.
- The SW has been successfully uploaded to EUCAIM's SW Image Repository.
- Signature of a processor's agreement with a non disclosure commitment.

Postconditions:
- Detailed usage metrics and statistics of the tool are visible, including user interactions and performance.
- Any identified issues or concerns regarding the SW's usage can be addressed to ensure a smooth user experience.
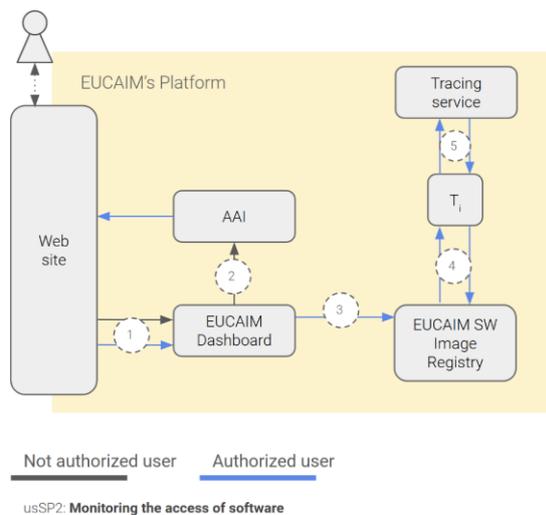
1: The user requests access to the EUCAIM Dashboard.
2: If the user is not authorised, the platform prompts the user to obtain the necessary authorisation.
3: Once it is logged in, the user navigates to the EUCAIM Marketplace.
4: In the EUCAIM SW Image Repository, the user selects one of the tools that he/she has uploaded (T_i).
5: For this specific tool, the user can consult different metrics collected by the Tracing service. These metrics will allow him/her to know how the tool is being used by the marketplace users.

usSP2: **Monitoring the access of software**

Figure 14. UML diagram of usSP2

Examples of User stories related to Data Users

Example of **Exploration of datasets from the Public Catalogue** User Story**:**

Table 16. usDU1.

#: usDU1	**Story name:** Exploration of datasets from the Public Catalogue	
Related WP: WP4, WP5	**Triggered by User Role (Fictitious Name):** Data User (Alice)	**Related User Roles (Fictitious Name):**
Overview Description:		
Alice, a researcher in a medical imaging research group, is preparing a research project focused on prostate cancer. As an anonymous user of the EUCAIM platform, she seeks to explore the diverse metadata of the datasets available in the EUCAIM Atlas of Cancer Images.		

Her primary objective is to perform an initial evaluation of the potentially available datasets, assessing their relevance and suitability for her research project.

Alice explores the Public Catalogue within the EUCAIM platform as an anonymous user and navigates through the available metadata, which are a combination of generic metadata and metadata following the DCAT-AP standards. She uses basic search and filtering options to identify datasets that could align with her research interests and objectives.

For example, Alice wants to know which datasets have breast cancer cases and the cohort's age is between 40 and 50. As a result, she finds that INCISIVE and CHAIMELEON datasets satisfy her criteria. Additionally, Alice can view the access conditions for each data dataset offered by the Data Holders.

In order to perform advanced filters, Alice will need to register into the platform via the Life Science AAI and accept the Terms of Usage and Privacy.

Preconditions:
- Stable internet connection is required for the data user to access the public catalogue within the EUCAIM platform.
- Metadata from the datasets must have been registered by the Data Holders onto the Public Catalogue.

Postconditions:
- Datasets meeting the specified criteria are successfully identified by using basic search and filtering options.
- The available metadata from the exposed datasets can be visualised.

Example of Federated query of aggregated data in the datasets User Story:

Table 17. usDU2.

#: usDU2	Story name: Federated query of aggregated data in the datasets	
Related WP: WP4, WP5	**Triggered by User Role (Fictitious Name):** Data User (Alice)	**Related User Roles (Fictitious Name):**
Overview Description: After signing up in the EUCAIM platform, Alice wants to make an advanced search on the multiple sources executing a federated query. For this purpose, Alice accesses into the EUCAIM Federated Query and uses the available options (cancer type, imaging modality, annotation availability, etc.) to perform queries over the different federated nodes, including the Reference Nodes which act as another node. This federated query is based on the hyper-ontology concepts and is more detailed than the previous search she performed. Now, as a result, she can not only determine the datasets but also the number of cases (as aggregated numeric results) that meet her criteria. For example, Alice searches for datasets		

that have breast cancer with a specific TNM stage value and with the MR series from a specific manufacturer. The query reveals that the INCISIVE dataset includes 123 cases, while the CHAIMELEON dataset contains 400 cases matching her criteria.[11]

Additionally, a report is generated containing the results of the query, which can be downloaded by Alice.

Preconditions:
- Valid user account, registered and validated through the AAI.
- Federated nodes, including the Reference Nodes, are interconnected and functional through the mediator component.
- Datasets are minimum in Tier 2.

Postconditions:
- Federated query is successfully performed over multiple sources.
- Aggregated results are obtained based on the hyper-ontology.
- Datasets meeting the specified criteria are successfully identified by using advanced search and filtering options, through federated queries.
- A report is generated with the results of the query.
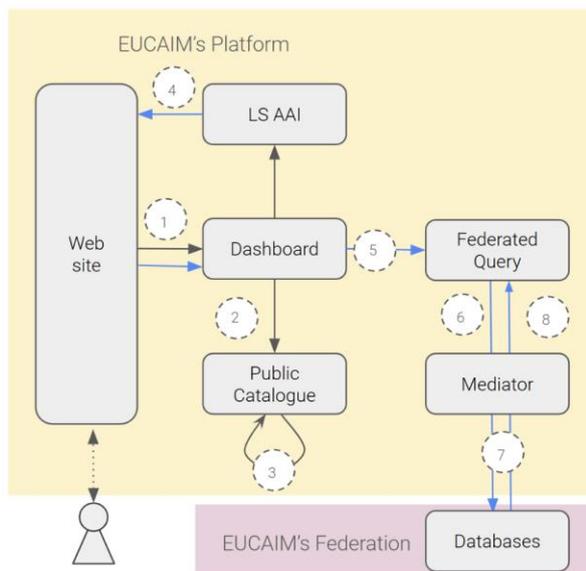
1: Anonymous user accesses the Dashboard
2: User accesses and explores the Public Catalogue
3: User performs basic search and filtering options
4: User authenticates into the platform
5: Authenticated user accesses the Federated Query
6: User performs advanced filters with aggregated data
7: The Mediator component interacts with the Federated Nodes and sends the query
8: The Mediator component returns the aggregated results

usDU1: **Exploration of datasets from the Public Catalogue**
usDU2: **Federated search of aggregated data in the datasets**

Figure 15. UML diagram of usDU1 and usDU2

Example of **Request access to existing datasets or propose building a new observational study using RWD through the negotiator** User Story**:**

Table 18. usDU3.

#: usDU3	Story name: Request access to existing datasets or propose building a new observational study using RWD through the negotiator	
Related WP: WP3	Triggered by User Role (Fictitious Name): Data User (Alice)	Related User Roles (Fictitious Name): Access Committee

Overview Description:

In the EUCAIM platform, Alice often needs access to federated data for her R&I projects. As described above, she has explored the Public Catalogue and used the Federated Query to search for potentially available datasets meeting her specific criteria.

After identifying the relevant datasets, Alice returns to the Public Catalogue, where she can select the datasets she wishes to access. By selecting these datasets, she adds them to a list that can be submitted to the Negotiator. Additionally, if she intends to initiate a new observational study with newly available RWD, she can choose this option from the Public Catalogue too.

In both cases, Alice accesses the Negotiator tool to initiate the process. She begins by submitting her access request to the Access Committee, providing all the required documentation. [12] [13]

The Access Committee reviews her request, assessing its alignment with the project's objectives, compliance with data privacy regulations, and adherence to other applicable policies. Based on this evaluation, the AC decides whether to approve or reject the request and inform the Steering Committee for the final decision. This process may involve iterations, as it is a negotiation process between the DU and the AC. Alice may have obtained a data permit from another actor in EHDS such as the National Health Data Access Body or HealthData@EU. In this case her authorised configuration will depend on the data permit issued by these bodies.

In cases where the intended purpose of the data request is not covered under the existing usage licenses agreed upon between the Data Holder and EUCAIM, the request is forwarded to the Data Holder's Access Committee. This committee can evaluate the request and issue the final decision to either approve or reject it.

Once the access request is approved, Alice is granted access to the federated data. With access to this rich and diverse dataset, Alice can conduct her analysis, enhancing the scope and accuracy of her findings.

Preconditions:
- The user has successfully logged in through the AAI.
- Datasets matching the specified criteria have been identified and shortlisted.
- An R&I project approved by an Ethics Committee is in place, under which the requested data will be utilized.

Postconditions:
- The access request has been successfully submitted via the Negotiator.
- All required documentation supporting the access request has been provided.

---

[12] https://drive.google.com/file/d/1F-DFw7-PXlBh29sQ2C0qjM_BYnRdO9Yx/view

[13] https://drive.google.com/file/d/1pmWyMfoVCVfS5ZHBJ7aOdKwXqlsog3cp/view

-	The Access Committee has reviewed the request and the Steering Committee issued a decision.
-	The user of the data has a valid data permit issued by a third party authority.
-	Upon approval, the DU is granted access to the federated data to proceed with her project.

Example of Get an overview of the datasets to which they have been granted access User Story:

Table 19. usDU4.

#: usDU4	**Story name:** Get an overview of the datasets which they have access to	
Related WP: WP4	**Triggered by User Role (Fictitious Name):** Data User (Alice)	**Related User Roles (Fictitious Name):**

Overview Description:

Following the successful evaluation and acceptance of their access request applications in the Negotiator, Alice now seeks to obtain an overview of the datasets to which she has been granted access. This allows her to explore her User's Dataset Library with the datasets to which she has access for her R&I projects.

Having received positive evaluation from the Access Committee regarding their approved access, Alice logs into the Dashboard and navigates to the section where the datasets list is shown.

In this section, Alice can review the metadata and details of each dataset she has been granted access to. This overview enables her to fully assess the suitability of each dataset for her specific R&I objectives.

Alice can also access information on any restrictions or conditions applicable to the datasets, ensuring compliance with privacy regulations, EHDS, AI ACT, data usage agreements, data permit conditions or licensing requirements.

Preconditions:
- Login is performed using the AAI (Authentication and Authorization Infrastructure).
- Access request applications within the Negotiator have already been evaluated and successfully granted.

Postconditions:
- The User'sUser's Datasets Library is available in the Dashboard.
- Metadata and details of datasets granted access can be successfully reviewed, as well as the information on any restrictions or conditions applicable to the datasets.
- Informed decisions regarding the selection and use of the available data resources can be performed by the data user.
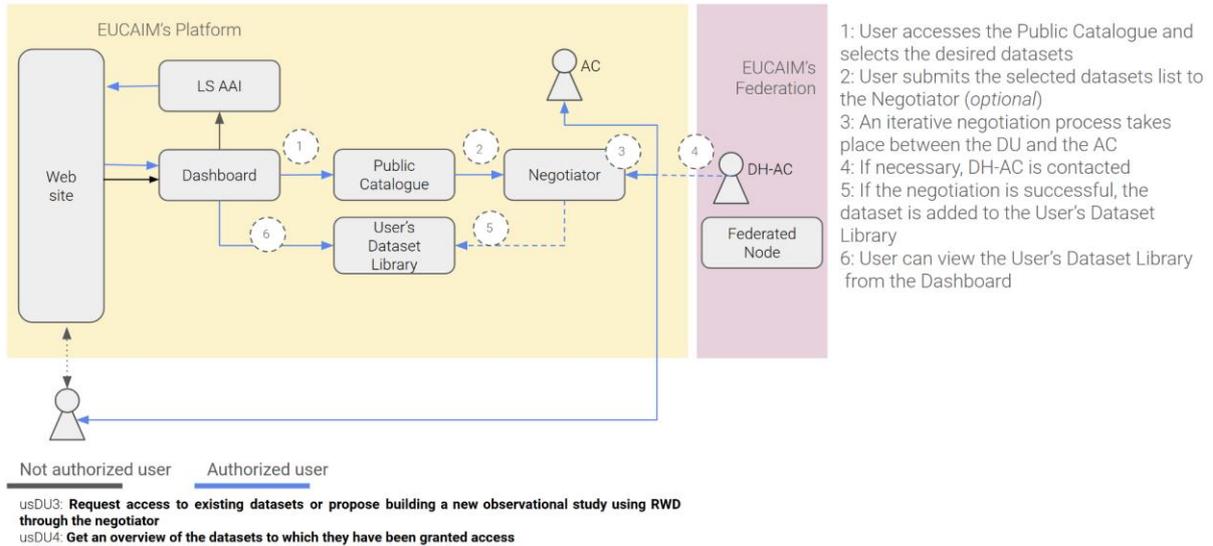
Figure 16. UML diagram of usDU3 and usDU4

Example of **Exploring the software available in the federation and their provenance** User Story:

Table 20. usDU5.

#: usDU5	Story name: Exploring the software available in the federation and their provenance	
Related WP: WP6	**Triggered by User Role (Fictitious Name):** Data User (Alice)	**Related User Roles (Fictitious Name):**
Overview Description: As a Data User of the EUCAIM platform, Alice has the opportunity to explore the diverse range of federated processing software available within the federated environment. Her goal is to identify suitable software for her specific research needs while understanding their provenance. As an anonymous user, Alice navigates the Marketplace, a centralised service that displays all the software contributed. These software encompass a wide array of functionalities, including image processing, data analysis, machine learning, and more. By navigating through the Marketplace, Alice gains insights into the capabilities and features of each tool, enabling her to make informed decisions about tool selection based on her specific research requirements. Additionally, the platform provides traceability and provenance information of the software. This valuable insight allows Alice to understand the datasets used for training the software, the performance metrics, and the overall popularity and reputation of each tool within the EUCAIM community. Throughout her exploration, Alice can leverage the Marketplace interface to browse, search, and filter the available software based on various criteria, such as tool type or functionality. The		

platform ensures a seamless and efficient experience, facilitating Alice's exploration of the software and enabling her to discover innovative solutions for her research projects.

Preconditions:
- The Marketplace interface for browsing, searching, and filtering software must be functional.

Postconditions:
- Exploration of the diverse range of software available within the EUCAIM platform.
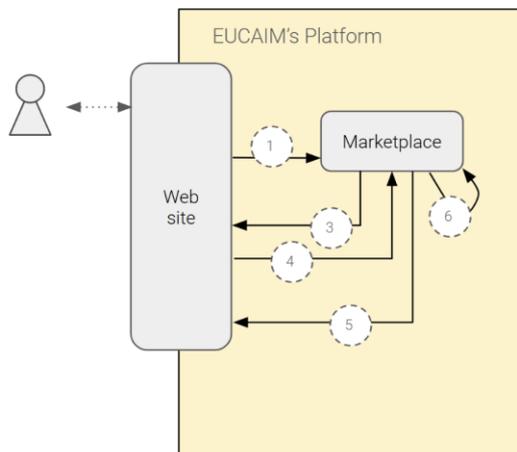
1: User accesses to Marketplace
2: Marketplace returns the list of software available in EUCAIM
4: User requests the provenance of a specific software
5: Marketplace returns requested information (provenance, performance metrics, usage of the software in EUCAIM)
6: User can filter the Marketplace according to categories, tags, functions and node restrictions

Not authorized user   Authorized user

usDU5: **Exploring the tools available in the federation and their provenance**

Figure 17. UML diagram of usDU5

Example of Run a federated (distributed) experiment on multiple datasets from the User's Datasets Library using software from the Marketplace User Story:

Table 21. usDU6.

#: usDU6	Story name: Run a federated (distributed) experiment on multiple datasets from the User's Datasets Library using software from the Marketplace	
Related WP: WP6	**Triggered by User Role (Fictitious Name):** Data User (Alice)	**Related User Roles (Fictitious Name):**

Alice, a Data User, has already explored the Public Catalogue and has been granted access to a series of datasets of her interest, that can help on her research project on prostate cancer. She later identifies the right software available in the Marketplace. Her main objective is to analyse baseline MR at the time of diagnosis to allocate treatment accurately.

Once Alice has selected the target datasets (the ones that compose her negotiation for the specific prostate project) and chosen the right software for her needs from the marketplace using the Federated Processing Dashboard, she configures the "federated analysis".

To do so, Alice accesses the Federated Processing Dashboard where she will find the datasets already selected as well as the software she previously picked. If she would have accessed this Dashboard without prior access to the Public Catalogue and the Marketplace, she would be able to select the target datasets (by selecting one of the negotiations she completed) and the software in this Dashboard.

The configuration interface allows her to identify which nodes from the federation are the holders of the datasets she selected. Moreover, the configuration might allow her - according to the nodes in the federation included in the experiment - to specify some computational resources (such as the number of CPU cores, memory allocation, and GPU usage) as well as storage settings (the destination folder for processed data, file formats, compression options, and retention policies).

During this process, Alice checks the configuration interface for clear and concise documentation or tooltips explaining the purpose and impact of each setting, aiding her in making informed decisions. Additionally, Alice can rely on the support services offered by EUCAIM, such as technical assistance, for any help she may require during datasets selection, software selection, analysis configuration and analysis processing. The user-friendly interface and documentation resources of the platform enable Alice to effectively utilise the selected software, even without extensive experience in AI development.

Alice would also like the interface to provide options to select pre and post steps, in order to create an analytical workflow, in case the datasets need some sort of pre-processing before running the selected software. Same idea applies as for post-processing.

By integrating the chosen software into her research project, Alice enhances the analysis of MRI data for treatment allocation in prostate cancer. Her ability to interpret the results obtained from these software, contributes to advancing the field of precision medicine and improving outcomes for patients in the context of prostate cancer management.

Preconditions:
- Login is performed using the AAI (Authentication and Authorization Infrastructure).
- Specific datasets suitable for the Data User's research project have been identified and appropriate access has been requested and been granted (Negotiator).
- Specific software suitable for the Data User's research project have been identified by accessing to the Marketplace of available software within the EUCAIM federated environment
- Basic understanding of software selection and data processing capabilities is possessed by the Data User or made available through support services.

Postconditions:
- The selected software is effectively utilised for processing EUCAIM's datasets.

- Results obtained from these software can be interpreted and used for research.

Example of **Training an AI model on multiple datasets from the User's Datasets Library leveraging EUCAIM's federated architecture** User Story**:**

Table 22. usDU7.

#: usDU7	**Story name:** Training an AI model on multiple datasets from the User's Datasets Library leveraging EUCAIM's federated architecture	
Related WP: WP6	**Triggered by User Role (Fictitious Name):** Data User (Alice)	**Related User Roles (Fictitious Name):**

Overview Description:

A team of data scientists led by Alice aims to publish a research project focused on developing an innovative federated AI model using the EUCAIM platform, a powerful infrastructure supporting federated learning.

For this purpose, Alice accesses the Dashboard and uses the catalogue to identify suitable datasets for their research. Alice explores the metadata datasets through a federated query. Once she has identified the datasets to use, she requests access to the datasets through the Access Negotiator, which manages the access requests. After finishing the process, Alice has access to the desired datasets (the negotiation was completed successfully).

With the datasets and processing software ready, they begin training the model using EUCAIM's distributed federated learning approach: each node in the federation holding one of the selected datasets securely retains its local data, ensuring privacy while contributing to the model's intelligence. At the end of the process, Alice stores the aggregated model, intermediate results are erased and evaluation metrics are stored in EUCAIM's storage, promoting collaboration within the team.

To do so, Alice accesses the Federated Processing Dashboard, where she will be able to select the datasets of interest (given a negotiation) and one of the "federated learning frameworks" available in the Marketplace as the software to use. During the configuration of the software, she will be able to select the AI libraries her model requires and she will provide a URL to a git repository with the AI model - compatible with the selected libraries - to train.

After successfully training and evaluating the model, Alice proceeds to publish their research project. Their work not only adds to the field of privacy-preserving machine learning but also gains recognition within the EUCAIM platform. As a result, Alice is accepted to provide their developed AI model to the Marketplace, adhering to recommended standards for interoperability and reusability, by following the "including new software for EUCAIM". To achieve this outcome Alice will need to provide evidence that her model complies with the legal and ethical obligations set out in the GDPR and AI Act and a legal representative of her entity

will grant the appropriate licences and permissions. This will follow the processes defined by EUCAIM for SPs.

Preconditions:
- Login is performed using the AAI.
- Datasets and processing software are prepared and available for training the federated AI model.

Postconditions:
- The federated AI model is successfully trained using the distributed federated learning approach.
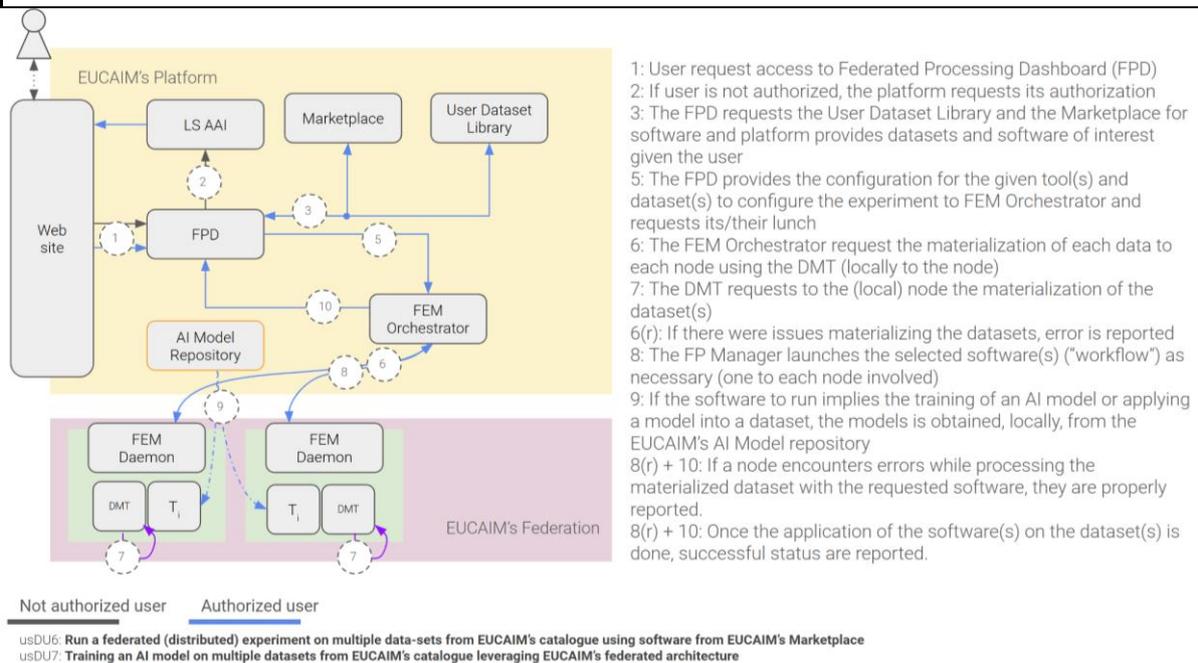

1: User request access to Federated Processing Dashboard (FPD)
2: If user is not authorized, the platform requests its authorization
3: The FPD requests the User Dataset Library and the Marketplace for software and platform provides datasets and software of interest given the user
5: The FPD provides the configuration for the given tool(s) and dataset(s) to configure the experiment to FEM Orchestrator and requests its/their lunch
6: The FEM Orchestrator request the materialization of each data to each node using the DMT (locally to the node)
7: The DMT requests to the (local) node the materialization of the dataset(s)
6(r): If there were issues materializing the datasets, error is reported
8: The FP Manager launches the selected software(s) ("workflow") as necessary (one to each node involved)
9: If the software to run implies the training of an AI model or applying a model into a dataset, the models is obtained, locally, from the EUCAIM's AI Model repository
8(r) + 10: If a node encounters errors while processing the materialized dataset with the requested software, they are properly reported.
8(r) + 10: Once the application of the software(s) on the dataset(s) is done, successful status are reported.

Not authorized user   Authorized user

usDU6: **Run a federated (distributed) experiment on multiple data-sets from EUCAIM's catalogue using software from EUCAIM's Marketplace**
usDU7: **Training an AI model on multiple datasets from EUCAIM's catalogue leveraging EUCAIM's federated architecture**

Figure 18. UML diagram of usDU6 and usDU7

Example of Monitoring job status in the Federated Processing architecture User Story:

Table 23. usDU8.

#: usDU8	Story name: Monitoring job status in the Federated Processing architecture	
Related WP: WP6	**Triggered by User Role (Fictitious Name):** Data User (Alice)	**Related User Roles (Fictitious Name):**
Overview Description: Alice, a Data User with a Data Scientist profile, wants to be able to monitor the status of her job in the Federated Processing Dashboard, so that she can track the progress and ensure timely completion of her data analysis tasks. In order to do so, Alice will access the Federated		

Processing Dashboard where she will have a user-friendly section that displays the status of her jobs in an intuitive interface. The information provided by this interface is job name, start time, estimated completion time (based on user's configuration not real time estimation), and current status (e.g., running, queued, completed). The interface allows users to filter out the jobs according to job name, start time, or status to easily locate and track the desired job. In case of any errors or issues encountered during job processing, the interface will provide a warning icon and the users, like Alice, can move into the job description to obtain the full error message and, if the software selected provides it, the full stack of the error. If the error reveals a security or operation incident, Alice will report to the Platform Manager through the Helpdesk, using the appropriate channel. The interface will maintain a menial number of metrics per job, successfully completed or not, for future reference and analysis. Alice will get email notifications, if desired, when a job completes or encounters an error, allowing her to take prompt action if required.

The system will have appropriate security measures in place to ensure that only authorised users, such as Alice, can access the job status information.

Preconditions:
- Login in the Monitoring component is performed using the AAI.
- A federated processing (analysis/inference or learning) has been initiated, and the user wishes to monitor its status.

Postconditions:
- The progress of jobs in the processing pipeline can be monitored using a clear and intuitive Dashbaord interface.
- Real-time updates on the progress of each job, including job name, start time, estimated completion time, and current status (e.g., running, queued, completed), are provided.
- Notifications or alerts are received when a job completes or encounters an error, enabling prompt action if required.

1 - User request access to Federated Processing Dashboard (FPD)
2 - If user is not authorized, the platform requests its authorization
3 - User requests the status of their jobs to the FPD
4 - The FPD requests the status of the jobs to the FP Manager
5 - From the internal database, the FP Manager queries the node's FP API having jobs with the status "running".
6 - The FP API returns the status of the jobs and the FP Manager updated the internal database
7 - The FP Manager returns the list of the jobs for the given user with all the status updated.

usDU8: **Monitoring job status in the Federated Processing architecture**
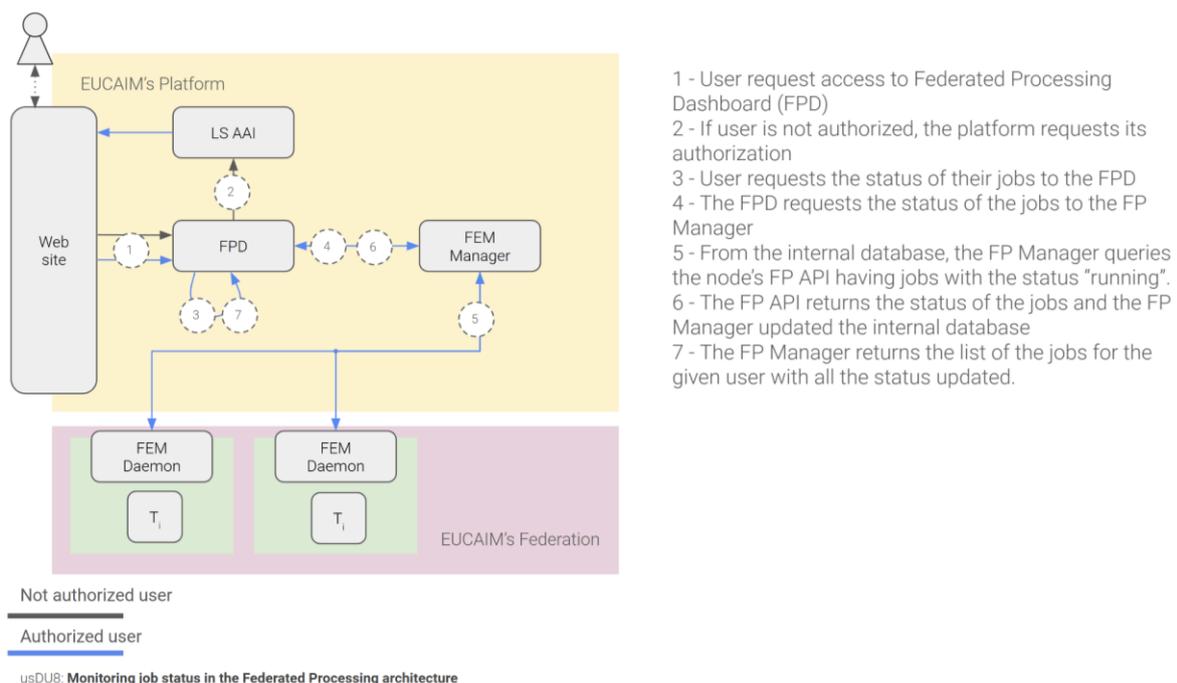
Figure 19. UML diagram of usDU8

Example of **Retrieve results from a federated experiment** User Story:

Table 24. usDU9.

#: usDU9	**Story name:** Retrieve results from a federated experiment	
Related WP: WP6	**Triggered by User Role (Fictitious Name):** Data User (Alice)	**Related User Roles (Fictitious Name):**

Overview Description:
 	
As a data scientist, Alice wants to retrieve and review the results obtained after running her federated AI model using the EUCAIM platform. After logging in the platform, she navigates the Federated Processing Dashboard and to the list of her finished jobs and accesses one of them to check the results. There, she explores the available options for retrieving the results.
 	
Alice utilises the interface to locate the file, or files, of interest, retrieved from the federated nodes; result of running a software or a workflow in the federation. The interface will allow her to obtain the desired file(s) as a ZIP-like file with some simple clicks.
 	
Preconditions:
  -   Login is performed using the AAI.
  -   A federated AI model has been executed
 	
Postconditions:
  -   The results obtained by running the federated AI model can be successfully retrieved and reviewed.
  - A comprehensive report generated by the EUCAIM platform can be exported and downloaded as a CSV/XLSX file, containing detailed information about the trained model's performance.		

Example of Retrieve an AI model trained in EUCAIM datasets User Story:

Table 25. usDU10.

#: usDU10	**Story name:** Retrieve and review obtained results	
Related WP: WP6	**Triggered by User Role (Fictitious Name):** Data User (Alice)	**Related User Roles (Fictitious Name):**
Overview Description: As a data scientist, Alice wants to retrieve and review the results obtained after running her federated AI model using the EUCAIM platform. After logging in the platform, she navigates up to EUCAIM's Federated Processing Dashboard and to the list of her finished jobs and accesses		

one of them to check the results. There, she explores the available options for retrieving the trained AI model.

Alice utilises the software provided by EUCAIM to obtain the AI model composed by the aggregation of the models trained in the federation. She can obtain basic metrics - according to the availability provided by the framework she used for training the AI model - like accuracy, precision, recall, and other relevant metrics. The EUCAIM platform integrates data visualisation software within the Federated Processing Dashboard, allowing her to effectively interpret the results. If available, Alice can download these metrics as a CSV file.

Preconditions:
- Login is performed using the AAI.
- A federated AI model has been trained.

Postconditions:
- The results obtained by training an AI model in the federation can be successfully retrieved and reviewed.
- A series of basic metrics - if provided by the framework - can be downloaded as a CSV file, containing information about the training performance.
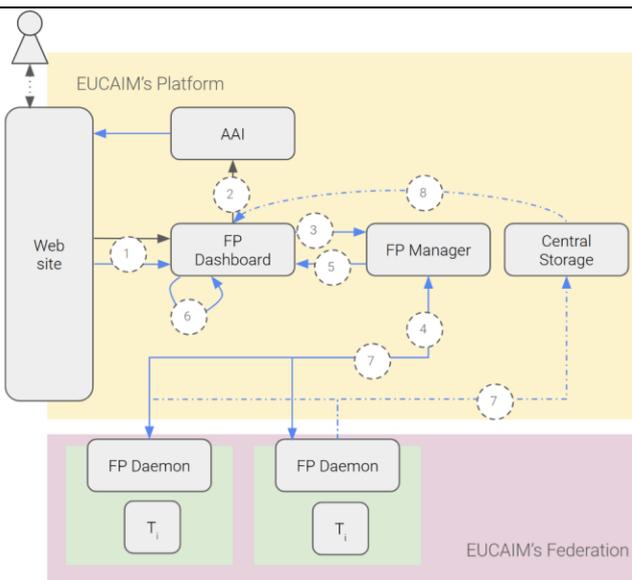
1: User request access to Federated Processing Dashboard (FPD)
2: If user is not authorize, the platform requests its authorization
3: The FPD requests to FP Manager to scan the results of a specific job on the node(s) were it was run, reporting the status of the files
4: The FP Manager runs the scan on the selected nodes
5: The FP Manager updates its internal database and the FPD with the status of the files
6: User select the files of interest to obtain
7: The FPD requests the FP Manager the data for a specific job, being copied into the user's space at the Central Storage (using a 3rd party tool or the data moving capabilities of the "FP Services")
8: FPD can provide a bridge with user space at the EUCAIM's Central Storage so the user can download the results of the job

usDU9: **Retrieve results from a federated experiment**
usDU10: **Retrieve an AI model trained in EUCAIM datasets**

Figure 20. UML diagram of usDU9 and usDU10

Examples of User Stories related to the Governing Body

As mentioned earlier, there have also been identified User Stories or situations in which profiles other than the defined users of the platform may interact with it. These situations may involve actions performed by these other profiles either directly or indirectly.

Example of **Evaluate and accept/reject datasets from Data Holders** User Story**:**

Table 26. usGB1.

#: usGB1	**Story name:** Evaluate and accept/reject datasets from Data Holders	
Related WP: WP2	**Triggered by Profile (Fictitious Name):** The Access Committee (AC)	**Related User Roles (Fictitious Name):** Data Holder (Raul)
Overview Description: Raul, as a member of a research organisation that seeks to join the EUCAIM federation and to contribute with different imaging and clinical datasets, has sent an application for providing datasets to be incorporated within EUCAIM. Depending on its scientific quality and taking into account its relevance, implementation and impact, as well as its compliance with the EUCAIM ethical and legal requirements the Access Committee reviews Raul's application and issues a positive or negative report or indicates to the applicant what additional documentation or evidence should be provided. The positive report is forwarded to the Steering Committee for decision. Raul receives a reply as soon as possible. **Preconditions:** - An application has been made by a Data Holder **Postconditions:** - The Access Committee accepts/rejects the application. - If accepted, new data is incorporated into the EUCAIM federation.		

Example of **Evaluate and accept/reject software from Software Providers** User Story**:**

Table 27. usGB2.

#: usGB2	**Story name:** Evaluate and accept/reject software from Software Providers	
Related WP: WP2, WP3	**Triggered by Profile (Fictitious Name):** The Access Committee (AC)	**Related User Roles (Fictitious Name):** Software provider (Paolo)
Overview Description: Paolo sends an application for providing software to be incorporated within EUCAIM. The AC provides to Paolo the "*EUCAIM software on-boarding guideline*". This document describes the multiple steps that Paolo needs to follow for his software to be validated and accepted in		

EUCAIM. Depending on the type of software, Paolo will need to provide evidence of the documentation related to the design that the risk-based approach and responsible design methodologies of the regulations (GDPR, AI Act, EHDSR) have been followed.

The process includes a detailed revision of the conceptual, technical and legal alignment of the software with EUCAIM's goals and architecture. Moreover, Paolo will need to provide some videos describing the software as well as showcasing how the software works in order to be used as training material for his software. The later stages correspond to the registration and description of the software in the Marketplace as well as providing a containerized image of the software so the federation of nodes can test their local execution.

Once the AC makes a decision of acceptance or rejection, supported by the internal governance bodies on technical, ethics and legal compliance. Paolo receives a reply as soon as possible.

Preconditions:
- An application has been made by a Software Provider.

Postconditions:
- The Access Committee accepts/rejects the application.
- If accepted, new software is incorporated into the EUCAIM federation.

Example of **Evaluate and accept/reject a data access request application** User Story**:**

Table 28. usGB3.

#: usGB3	**Story name:** Evaluate and accept/reject a data access request application	
Related WP: WP2, WP3	**Triggered by Profile (Fictitious Name):** The Access Committee (AC)	**Related User Roles (Fictitious Name):** Data User (Alice)
Overview Description: The AC receives an email notification of a new request from Alice, who is willing to access specific datasets of the Atlas of Cancer Images she has selected from the Public Catalogue to carry out her research Project. Through the Negotiator Component, the AC reviews all the application documents and initiates the established procedure to perform the evaluation according to the alignment of her project. The AC also checks if the different EUCAIM resources are available and if the access conditions are met, changing the states of the negotiation if needed. Any rejection is fully justified by the AC in writing form, according to objective criteria defined in the internal procedures of the AC. The members of the AC are appointed by the EUCAIM Management Board for a given period of time and periodically renewed.		

If the Alice request has been previously authorised by an HDAB or HealthData@EU and communicated to EUCAIM, the CA only verifies that it is consistent with the configuration and cataloguing of the datasets concerned.

Preconditions:
- An application has been made by a Data User.

Postconditions:
- The Access Committee reports positively or negatively on the application for decision making by the Steering Committee.
- If accepted, a research project is carried out with data available from EUCAIM.

Example of **Evaluate and accept/reject a RWD observational study application** User Story:
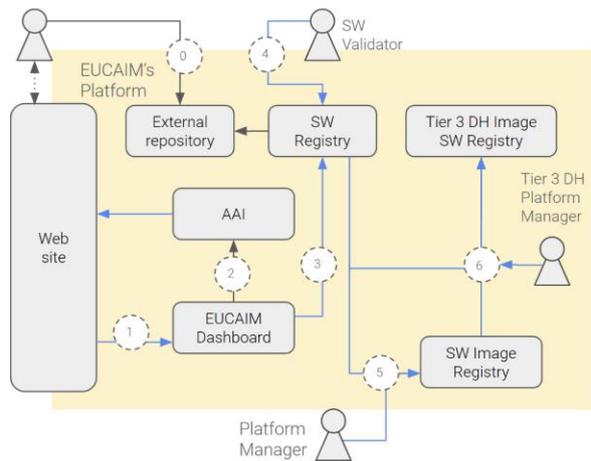
Table 29. usGB4.

#: usGB4	Story name: Evaluate and accept/reject a RWD observational study application	
Related WP: WP2, WP3	**Triggered by Profile (Fictitious Name):** The Access Committee (AC)	**Related User Roles (Fictitious Name):** Data User (Alice) Local Data Manager in a Real World Data Holder (Irene)

Overview Description:

The AC receives an email notification of a new request from Alice, who is willing to build a new observational study with Real World Data (RWD). Through the Negotiator Component, the AC reviews the documentation attached and performs a first evaluation of the application.

If it is aligned with EUCAIM's objectives, the AC contacts all the RWD Holders connected to EUCAIM, asking if they are interested in participating as partners in the project and how many cases they would provide according to the inclusion and exclusion criteria defined by Alice.

When the AC receives the answer from the RWDHs, informs Alice about the hospitals that are eligible to participate and together make the selection of those that will finally take part in the project. At that point, the Local Data Managers of the hospitals, such as Irene, will be contacted again to dynamically generate the secondary use datasets from their Data Warehouse and the AC will be in charge of giving access to Alice when they are ready through the Negotiator.

Preconditions:
- An application has been made by a Data User.

Postconditions:
- The Access Committee accepts/rejects the application.
- If accepted, a research project is carried out with data newly available in EUCAIM.

Examples of User Stories related to the Platform Manager

Example of **Uploading of software to the federation marketplace** User Story:

Table 30. usPM1.

#: usPM1	Story name: Uploading of software to the federation marketplace	
Related WP: WP4	**Triggered by Profile (Fictitious Name):** Platform Manager (Ignacio)	**Related User Roles (Fictitious Name):** Software provider (Paolo)
Overview Description: As a member of the Platform Manager team, Ignacio is responsible for uploading authorised and tested software provided by the SPs to the federation marketplace. Recently, Ignacio created a code repository for Paolo to upload the necessary files to build the container following the specifications of software development in EUCAIM. This way, Ignacio gets access to the tool from Paolo and can build the containerized tool. Ignacio stores the tool in the appropriate container registry, such as Docker Hub or Harbor. In addition, Paolo has defined all the required associated information for the tool, such as its name, version, requirements, tags, and more. Ignacio registers this information through a user interface on the platform. **Preconditions:** - Necessary specifications for software development in EUCAIM are defined and accessible. - A code repository is created for software providers to upload code and files for container building. **Postconditions:** - The containerized tool is built following the specifications of software development in EUCAIM. - The built tool is stored in the appropriate container registry (e.g. Docker Hub, Harbor). - The uploaded tool is available for access and use by users within the federation marketplace.		

0: Prerequisite: The user has the SW and the container image description in an external repository.
1: The user requests access to the EUCAIM Dashboard.
2: If the user is not authorised, the platform prompts the user to obtain the necessary authorisation.
3: Once it is logged in, the user navigates to the SW Registry, where (s)he adds the description of the SW in the SW Registry, referencing the information from the external repository.
4: The SW validator, verifies if the SW corresponds to the application approved by the access committee and that it is compliant to the SW development guidelines.
5: The EUCAIM Platform Manager creates the container image and pulls it to the SW Image Registry.
6: Each Tier 3 Data Holder Platform Manager that could support this new SW checks its description and eventually customises the image to run it on the Tier 3 platform (e.g. by adding new environment variables).

NOTE: For readability, only the authentication of the user is shown in the diagram. Every actor performing an action that requires authorisation (blue arrows) should have acquired valid credentials before performing the actions.

Figure 21. UML diagram of usPM1, usSP1 and usGB2

Example of **Platform services administration** User Story:

Table 31. usPM2.

| #: usPM2	Story name: Platform services administration	
Related WP: WP4	**Triggered by Profile (Fictitious Name):** Platform Manager (Ignacio)	**Related User Roles (Fictitious Name):** Data User (Alice) / Software Provider (Paolo)
Overview Description: Ignacio, as part of the Platform Manager team, is responsible for managing platform services in an efficient and secure manner. He receives requests from Paolo, an Application Developer, or Alice, a Data Scientist, to provide them with execution environments that meet their specific tool requirements for research and innovation purposes. With administrative credentials to the platform, including access to the Dashboard, Ignacio ensures seamless administration of platform services. Ignacio also proactively identifies the need for updates to existing software, either due to security concerns or at the request of Alice. Recognizing the importance of maintaining a secure environment, Ignacio employs DevOps software to automate the building and uploading process for these applications. He meticulously codes the application environment, conducts thorough verification, and proceeds to upload it to the Application Registry. By doing so, the updated applications become readily available to all users within the EUCAIM platform. As a EUCAIM Platform Manager, Ignacio plays a crucial role in ensuring the smooth functioning, security, and accessibility of services within the EUCAIM ecosystem. By leveraging automation and proactive updates, Ignacio contributes to the overall efficiency and user		

satisfaction, ultimately supporting the research efforts of both Alice, as well as Paolo and other platform users.

Preconditions:
- Administrative credentials to the EUCAIM platform are available.
- The need for updates to existing software is identified, driven by security concerns or user requests.

Postconditions:
- Execution environments satisfying specific tool requirements are provided to users for research purposes.
- Updates to existing software are meticulously coded, verified, uploaded and accessible to all users within the EUCAIM platform.
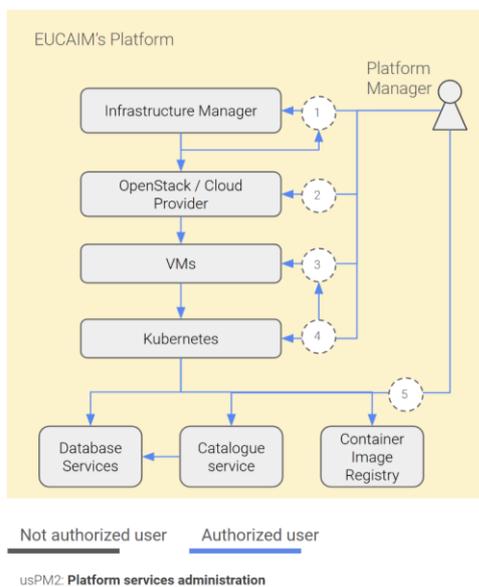
NOTE: The platform manager should have credentials for managing their resources.
NOTE: The Core Services Platform is operated through services running on SW containers orchestrated by a Kubernetes cluster deployed on top of Virtual Machines (VMs) that are managed by an OpenStack Cloud platform and which have been instantiated through an DevOps tool (Infrastructure Manager).
1: The platform manager deploys or updates the SW platform by means of the Infrastructure Manager (IM) DevOps Tools using the IaC blueprint of a Kubernetes Cluster that is deployed on the VMs of OpenStack. For this action, IM and OpenStack credentials are needed. As a result, the Platform Manager receives the credentials and endpoints for accessing the VMs and the Kubernetes cluster.
2: The Platform manager administers the lifecycle of the VM deployed by IM directly on the Cloud Provider (e.g. graceful restarts, planned shutdowns).
3: The Platform manager administers the VMs (eg. SW patching) by logging on the VMs with the credentials provided by IM.
4: The Platform Manager deploys and operates the Central Hub Services in Kubernetes through IaC manifests, either directly to the Kubernetes Dashboard or through the Kubernetes command line tool by logging in the VMs.
5: The Platform Manager accesses the core services of the Central Hub with privileged credentials to operate the Database services, the catalogue services and the container registries. These credentials are defined in the deployment stage (4).

Figure 22. UML diagram of usPM2

4. User Actions

For each User Role identified in the previous section, a detailed description is presented, including the specific list of User Actions (related to the use of the platform from a technical perspective) the user is enabled to carry out and the context of the User Story where these actions take place. Regarding the list of actions, Table 32 shows the User Actions that can be carried out by all user roles (any User).

The user actions defined in the initial version of this document (*D4.1*), were evaluated in *D4.10, Technical Evaluation of the Platform*. In this version, we have updated the user actions, taking into account the feedback received from *D4.10* as well as the project's progress. The final feedback from these evaluations will be incorporated into the platform's final implementation.
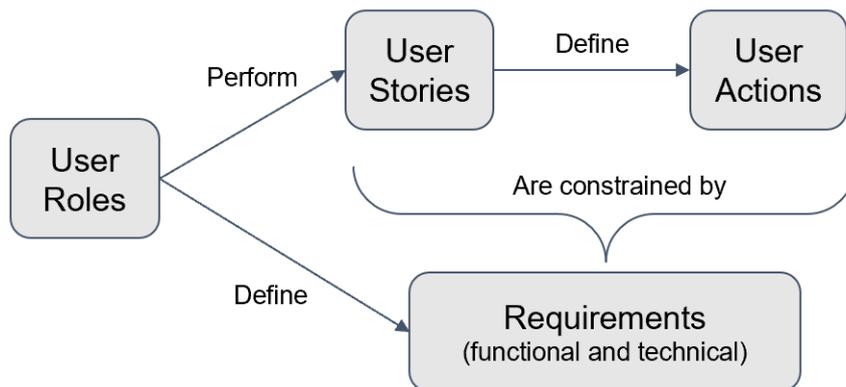
Figure 23. User workflow

User Actions common for all Roles

Table 32. User Actions common for all Roles.

#	User Action	Description	User Story
uaC1	Understand the processes	Find instructions and documentation, as well as links to use the platform depending on the user role	usDHx, usSPx, usDUx, usGBx, usPMx
uaC2	Register into the platform	Register and log into the platform through the Life Sciences (LS) AAI	usDHx, usSPx, usDUx, usGBx, usPMx
uaC3	Authenticate into the platform	Use LS AAI to authenticate users in the Dashboard	usDHx, usSPx, usDUx, usGBx, usPMx

User Actions enabled for Data Holders

Table 33. User Actions enabled to Data Holders.

#	User Action	Description	User Story
uaDH1	Join the federation	Submit the Expression of Interest to join the federation	usDH1
uaDH2	Application for new datasets to be included in the Federation	Submit an application to include new datasets in the Federation (e.g. through the Internal Open Call)	usDH1, usGB1
uaDH3	Prepare the datasets with their metadata	Prepare the datasets and their metadata following EUCAIM's specification	usDH4, usDH5, usDH7
uaDH4	Transfer or share a dataset	Ingest their already prepared data to the Reference Node or share it through a Federated Node	usDH2, usDH6, usDH7
uaDH5	Upload metadata	Make their data discoverable, by pushing the metadata of the datasets into the Public Catalogue	usDH5
uaDH6	Trace a dataset	See information about the usage of their datasets	usDH8

User Actions enabled for Software Providers

Table 34. User Actions enabled to Software Providers.

#	User Action	Description	User Story
uaSP1	Request tool registration	Request approval of the Governing Body to upload the tool to the platform, accepting the terms and conditions	usSP1, usGB
uaSP2	Upload the tool to the platform	Upload the developed tool following the EUCAIM guidelines and recommendations to the platform, along with the	usSP1, usPM1, usGB2

#	User Action	Description	User Story
		related usage documentation	
uaSP3	Track the usage of the tool	Review the usage of the tool and some metrics about it	usSP2

User Actions enabled for Data Users

Table 35. User Actions enabled to Data Users.

#	User Action	Description	User Story
uaDU1	Browse and filter datasets	Browse datasets in the public catalogue from different repositories and filter the datasets of interest	usDU1
uaDU2	Federated query to filter datasets	Search for datasets of the federated catalogue. Query will be expressed in a structured language. The query will be run on the repositories of different providers, returning the number of studies that match the search criteria	usDU2
uaDU3	Request access to datasets	Requests access to the datasets through the Negotiator providing the respective information	usDU3
uaDU4	Follow-up on the status of the application	Retrieve information about the status of the request through the request tracking mechanisms integrated in the Negotiator.	usDU3
uaDU5	Browse the datasets to which the user has access	List and filter the datasets to which the Data User has been granted permissions in the User's Dataset Library	usDU3, usDU4

#	User Action	Description	User Story
uaDU6	Browse software in the marketplace	Inspect, filter and select the available software to be executed by the federated processing service	usDU5
uaDU7	Federated processing	Once the proposal is accepted, the DU gets access to the Federated Processing Dashboard which will orchestrate the execution of the chosen software over the authorised and selected datasets (either in the reference node or in a federated node)	usDU6, usDU7
uaDU8	Monitor job status	Monitor the progress of a federated processing job and eventually cancel it	usDU8
uaDU9	Retrieve the results of the execution	Review the charts and tables containing the results of the federated processing and export them as a CSV/XLSX file	usDU9, usDU10

User Actions enabled for the Governing Body

Table 36. User Actions enabled to the Governing Body - Access Committee.

#	User Action	Description	User Story
uaGB1	Ethical and legal review of DUs applications	Review whether the data access requests meet the appropriate ethical and legal conditions	usGB3
uaGB2	Decide on the concession of permissions to access the data	Depending on the type of agreement with the Data Holder and the characteristics of the application, the Governing Body will approve or reject the data access request	usGB3

#	User Action	Description	User Story
uaGB3	Forward the request to the Data Holder Access Committee	If the datasets are located at federated nodes (not the reference nodes), the Governing Body, using the Negotiator service, will forward the request to the Access Committee of the correspondent Data Holder when needed	usGB3
uaGB4	Evaluate and accept/reject the inclusion of new datasets and software	The Access Committee and SW Validator will evaluate new applications	usGB1, usGB2
uaGB5	Interact with the applicant	The applicants are informed about the status of the request through the Negotiator	usGB1, usGB2, usGB3, usGB4
uaGB6	Interact with the RWDHs	Contact with the RWDHs to ask for their participation in new observational studies coming from the DUs through the Negotiator	usGB4, usDU3

User Actions enabled for the EUCAIM Platform Manager

Table 37. User Actions enabled to the EUCAIM Platform Manager.

#	User Action	Description	User Story
uaPM1	Integrate new software accepted into the platform	Upload the software supplied by the Software Providers to the platform, configuring them and integrating them with the rest of the services	usPM1
uaPM2	Maintain services and monitor jobs	Monitor the services and federated processing jobs	usPM2
uaPM3	Manage platform permissions	Register users in the platform and assign permissions based on the input of the Governing Body	usPM2

5. Operational procedures

5.1. Data provision

Application process:

A Data Holder submits an application to join the federation using the Expression of Interest, available from the 'Become a Data Holder' section of the Dashboard[14].

The proposal will be evaluated by the Access Committee in relation to its Scientific quality of the applications, taking into account its Relevance, Implementation and Impact, and its compliance with the EUCAIM legal requirements will be checked.

Prior to the integration of the data holders into the infrastructure, the corresponding regulatory contract must be signed.

EUCAIM works with two main categories of Data Holders:

1. Data Holder in a Federated node:

Description

A Data Holder in the federated node is an entity that contributes data assets, ensures data quality and integrity, and abides by privacy and security regulations to facilitate collaborative data sharing and analysis within the federated system.

A Data Holder joins EUCAIM as a federated node using infrastructure available on-site. This requires a series of steps regarding legal matters, such as signing a DSA and data de-identification, as well as on data quality assurance, interoperability and compliance monitoring, as explained below.

Main procedures

1.1 Functional requirements and IT implementations:

Each federated node will need to fulfil some minimum technical requirements to ensure quality of service in the federated data access and processing.

1.2 Legal engagement:

A DSA and all its annexes (delivered in WP3) will be signed between the Data Holder and EUCAIM.

Contractual documents will be signed between EUCAIM and the Data Holder. Different versions of these documents will be available to cover for:

- Each federated node.
- One national node (aggregating all federated nodes in the Member State).

---

[14] Expression of Interest: https://dashboard.eucaim.cancerimage.eu/expression-of-interest

- EUCAIM as an EDIC (Pan-European entity) may aggregate federated nodes in different Member States according to its Statutes.

1.3 Legal framework will define the requirements for the data to be made accessible via EUCAIM.

Within the DSA, a specific usage licence must be defined. This licence outlines the terms and conditions for the use and distribution of the data. It ensures that the legal rights and restrictions associated with the data are clearly communicated and respected by users.

1.4 Proper anonymisation or data de-identification will be done in accordance with the local regulation and the internal procedures at each node, under their direct management.

1.5 Data quality assurance. Data quality verification will be performed (e.g., image quality check, minimum set of associated clinical variables per cancer type).

1.6 Datasets registration in the Metadata Catalogue, following the FAIR standards defined by EUCAIM.

1.7 Compliance monitoring guidelines will be distributed amongst Data Holders in order to ensure that EUCAIM procedures are followed.

The first pilot of the federated infrastructure will be built on knowledge and technology developed in both a) AI4HI centralised imaging repositories (EUCANIMAGE, CHAIMELEON) and b) clinical sites participating in other AI4HI projects (INCISIVE).

2. Data Holder in the Reference Nodes:

Description

A Data Holder in the Reference Nodes is an entity that supplies and contributes data to the centralised storage system, ensuring data quality, adhering to privacy and security regulations, and enabling efficient data management, analysis, and retrieval.

A Data Holder joins EUCAIM by transferring data to the Reference Nodes of the infrastructure. This requires a series of steps regarding legal matters, such as signing a data transfer agreement and data de-identification, as well as on data quality assurance and compliance monitoring, as explained below.

Main procedures

2.1 A DTA (delivered in WP3) will be signed between the Data Holder and EUCAIM.

2.2 Within the DTA, a specific usage licence must be defined. This licence outlines the terms and conditions for the use and distribution of the data. It ensures that the legal rights and restrictions associated with the data are clearly communicated and respected by users.

2.3   Data anonymization. If needed by the Data Holder, EUCAIM will facilitate access to services/software which streamline the anonymisation and batch uploading to the EUCAIM Central Hub.

2.4   Data quality assurance. Data quality verification will be performed (e.g., image quality check, minimum set of associated clinical variables per cancer type).

2.5   Datasets registration in the Metadata Catalogue, following the FAIR standards defined by EUCAIM, and ingestion into the Reference Nodes, through the software provided by EUCAIM.

2.6   Compliance monitoring guidelines will be distributed amongst Data Holders in order to ensure that EUCAIM procedures are followed.

5.2.   Software provision

Application process

A software provider will submit an application to incorporate a software or service into the Marketplace using a standardised form following the steps described in the working document "*EUCAIM software on-boarding guideline*". The proposal will be evaluated and accepted/rejected by the EUCAIM SW Validator.

Description

A Software Provider refers to the entity that would like to contribute with pre-processing/post-processing software, services, or applications they have developed, to the EUCAIM's marketplace for use in the federated, distributed or centralised processing purposes of the platform.

It must be aligned with development guidelines for software provision, with a focus on containerization, image requirements, verification, usage instructions, licensing, integration tests, and periodic dependency updates.

Main procedures and development guidelines

1.   Container-based Development:

      To adhere to development guidelines, the provision of software and services must be done through containerization. Containers provide a consistent and isolated environment for deploying and running applications. Developers have to package their software and services, as described in the working document "*EUCAIM software on-boarding guideline*", into container images to ensure portability and ease of deployment across different environments.

2.   Container Image Requirements:

      When creating container images, certain requirements need to be met. These may include specifying the base image, ensuring the inclusion of all necessary dependencies and libraries, adding the necessary volumes to be mounted, defining

the entry point for the application or tool, and configuring any environment variables or runtime settings required for proper functioning.

3. Verification by a group of security experts of the Consortium:

Before the software or services are deployed they must be validated by the EUCAIM SW Validator to ensure compliance with privacy, security, and integrity standards. The process includes a security assessment to identify vulnerabilities, validation of encrypted communication (e.g., SSL/TLS), and sandboxed execution to prevent data corruption or resource exhaustion. The tool must demonstrate algorithmic fairness, avoiding biases, and provide comprehensive logging for transparency and traceability. This ensures that the tool operates securely, ethically, and aligns with the platform's requirements.

4. Usage Instructions and examples:

To facilitate the adoption and usage of the software and services provided, clear instructions and examples must be provided. These guidelines will include step-by-step instructions on how to deploy the container image, configure any necessary parameters, and effectively utilise the tool or service. Additionally, examples showcasing common use cases and demonstrating the functionality of the tool or service can help users understand its capabilities.

5. Specific Usage Licence:

When providing software and services, a specific usage licence must be defined. This licence outlines the terms and conditions for the use, distribution, and modification of the software or services. It ensures that the legal rights and restrictions associated with the software or services are clearly communicated and respected by users.

6. Integration Tests:

To ensure the compatibility and seamless integration of the software or services with the existing infrastructure, comprehensive integration tests need to be performed. These tests validate the interoperability of the software or services with other components, verify the functionality of APIs or interfaces, and detect any potential conflicts or issues that may arise during integration.

7. Periodic Dependency Updates:

As software ecosystems evolve, it is crucial to regularly update the dependencies used by the software or services. Periodic dependency updates help address security vulnerabilities, take advantage of new features and performance improvements, and maintain compatibility with other components in the infrastructure. A systematic and scheduled process needs to be in place to review and update dependencies, ensuring the long-term stability and security of the provided software or services.
By following these development guidelines, organisations can ensure that the provision of software and services is standardised, secure, and well-documented, enabling smooth adoption, integration, and maintenance within the Marketplace.

5.3. Data Access Requests

Application process

A Data User may submit a request for access to data, whether it is already available from existing repositories exposed in the Public Catalogue or they want to generate on-demand datasets within the RWD environment under a proposal for a new observational study. This request process will be orchestrated through the Negotiator service which will allow the interaction of the Access Committee and the requesters. The request will be evaluated by the Access Committee, and, depending on the licences signed with the Data Holder that owns the data, it may also be evaluated by the DH's Access Committee (or an equivalent entity). The approval of the request will imply the granting the usage of the datasets.

Description

The DU is fully committed to leveraging the research and innovation infrastructure's capabilities to drive their project forward. By adhering to the necessary procedures, accessing the relevant datasets, utilising federated data analysis and other processing software made available by EUCAIM, and considering the financial implications, the DU aims to achieve groundbreaking results in their project.

Main procedures

1. Exploring the Public Catalogue as an anonymous user

   To initiate their R&I project, the DU begins by exploring the Public Catalogue, which provides access to the metadata catalogue containing basic information about the available datasets in the Atlas of Cancer Images. The user carefully evaluates the catalogue, searching for datasets that align with the objectives or interests.

2. User registration /authorisation and their acceptance of the Infrastructure's Terms and Conditions of use

   Before proceeding further, the DU understands the importance of adhering to the infrastructure's rules and regulations. They complete the necessary User Registration and Authorization process by accepting the Infrastructure's Terms and Conditions (T&C) of use (to be delivered by WP3). This step is crucial to maintain the integrity and security of the infrastructure.

3. Data searching as an authenticated user

   As a registered user, the DU can now execute federated queries and make an advanced search across multiple sources. As a result, users will obtain aggregated results with the total amount of studies and subjects fulfilling their search criteria.

4. Data Access request

   As a registered user, the researcher or innovator is now prepared to request access to data from the Reference Nodes and/or federated nodes, depending on the specific requirements of their project. There are two distinct processes that the DUs can follow based on their needs:

   - If the desired datasets are available in the Public Catalogue, DUs can select them and send them to the Negotiator component, where they must complete and submit the required information and documentation about their project.

- If the DUs is looking for new data that are not included in the search results returned to them, they can send a request to the Federation for new data. For this, they can select the "Build an Observational Study with RWD" option in the Public Catalogue and send it to the Negotiator, where additional documentation will be required. This process may involve signing special agreements with the respective RWDHs, ensuring the compliance of the newly available datasets.

A local ethics committee's approval (either from the Data User's institution or external if not available, but always from the country where the project is going to be conducted) is needed at this stage to request data access, where this is applicable.
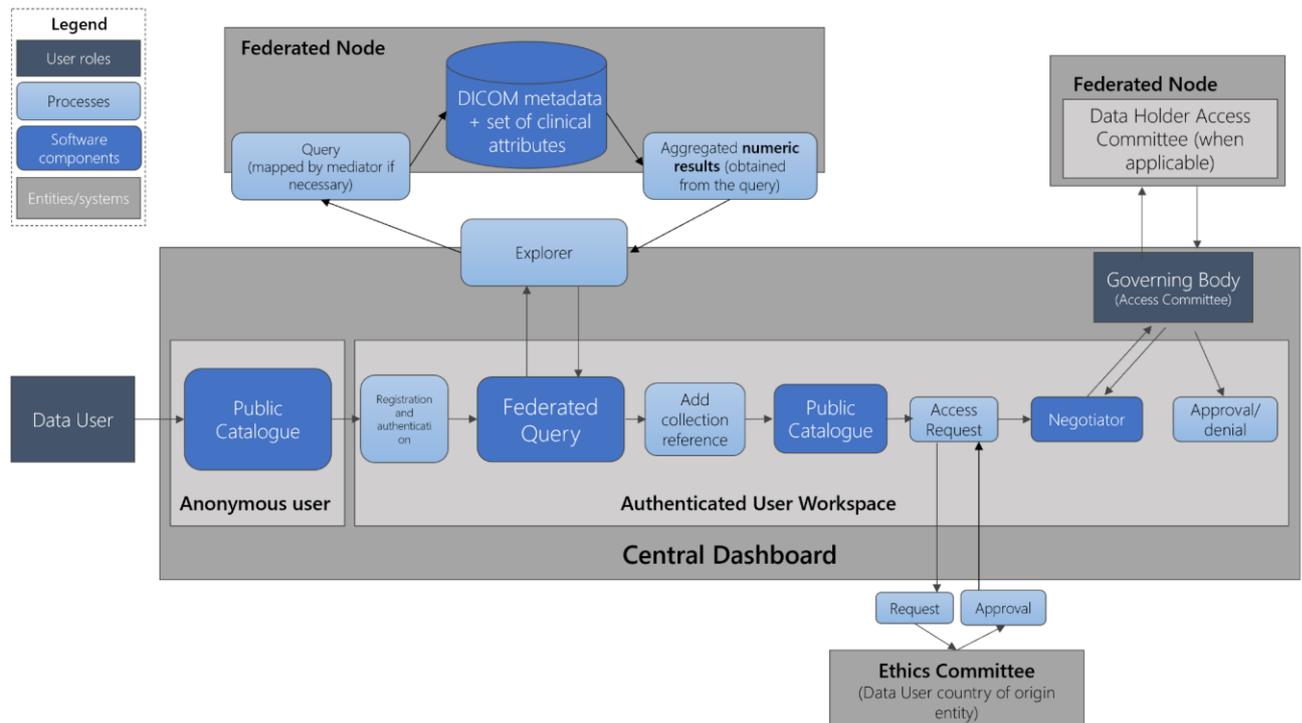

Figure 24. Data Access Request workflow

5. Access to datasets and software in the Atlas of Cancer Images

5.a Access to datasets of the Reference Nodes:

Once the user gains access to the datasets located in the Reference Nodes, they will be able to employ the cloud platform provided by the infrastructure to perform their analysis within the platform itself.

5.b Access to federated data

Once the DU gains access to federated data within the infrastructure, they are presented with a Federated Processing Dashboard for data utilisation. Federated data refers to data that is distributed across multiple nodes within the infrastructure, allowing DUs to analyse and gain insights from diverse sources while preserving data privacy and security. This approach offers

several benefits, including data security, collaboration and sharing, scalability and resource management and streamlined workflow.

5.c Access to datasets of the Reference Nodes and to federated data

Once the DU gains access to both datasets located in the Reference Nodes and to federated data, the workflow would be the same as the ones described previously for each type.

5.d Access to software, managed by the Central Hub

In addition to data access, the infrastructure offers pre-processing (data annotation, harmonisation, data quality, data FAIRness evaluation) and diagnostic, prognostic and therapeutic software for both centralised and federated data analysis and training / inference from AI models. The DU recognizes the value of these software in unlocking hidden insights and enhancing their project outcomes. They eagerly utilise these resources, harnessing the power of data analysis and AI models for their research.

6. Invoicing and payment, when applicable

In certain cases, access to specific datasets or processing services may come with a fee. The DU understands the potential costs involved and acknowledges the Invoicing and Payment process. Depending on the circumstances, the Central Hub may handle the pricing and payment, or the individual nodes within the infrastructure may manage these aspects directly.

6.a Future requirements.

The design of procedures consistent with two processes associated with the full entry into force of the EHDSR will be considered during the life cycle of the Project:

- The integration of data sets coming from data altruism entities.
- The procedures for automatic recognition of data permits for data cross-border processing on the basis of a data permit issued by an HDAB or HealthData@EU

5.4.   Other considerations

Important dates for the involvement of Data Holders and users (Figure 25)

During the execution of the project, the involvement of new Data Holders and data users is foreseen, at two important milestones:

- The Open Call: 10th of June 2024 was the deadline for submitting proposals to the Open Call. There have been 68 applications (Data Holder: 30; Data user: 19; Both: 15; Ineligible: 4). The Access Committee evaluated them and ranked the proposals, and the Coordination Team and the EC validated the new partners. The budget for all the applicants was limited to 214.000 € to all participants including all costs. The new partners will be founded under the Digital Europe Programme rules and will have the

same legal framework (GA and CA) as the rest of participants in the project and operate as "internal users" of the infrastructure. It is expected that the new partners sign the GA at the end of 2024.

- The Business Model pilots in 2026: the Data Holders and users to be involved in these pilots are not reimbursed and are "external users" of the infrastructure, therefore a legal framework will be needed to rule their use of the infrastructure. This will be the basis of the legal framework that the infrastructure will use in its Operational Phase, after project end, under the management of the EUCAIM Central Hub.
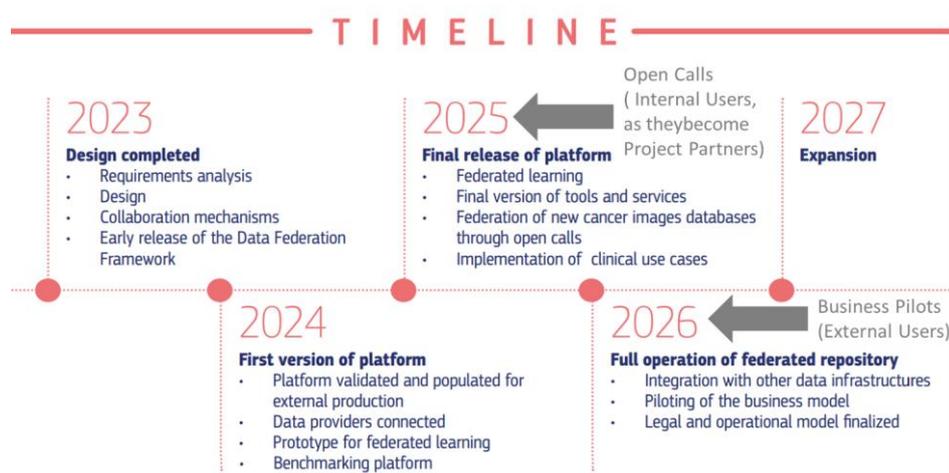

Figure 25. The EUCAIM platform implementation timeline

The business model pilots will imply the need to assess the incorporation of new Providers (both Data Holders and software Providers), as well as new Data Users. The managerial structure for the assessment of candidates is depicted in Figure 26 and Figure 27 for Data / software providers and Data Users, respectively.

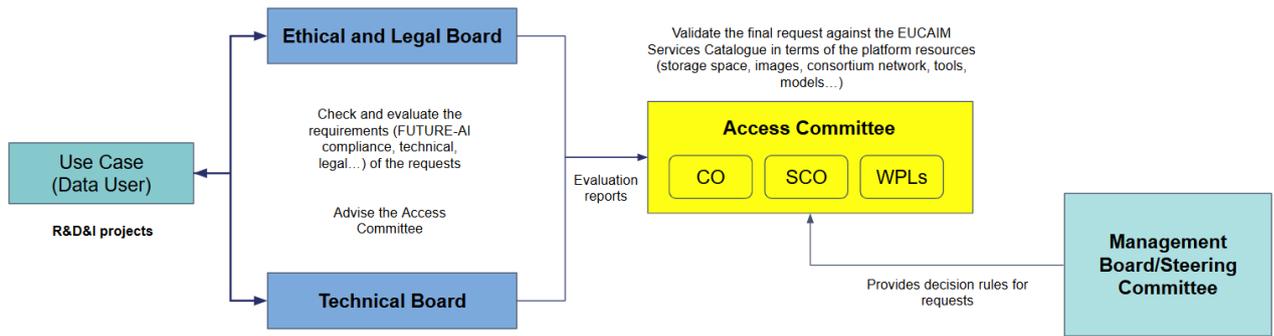*Figure 26. Data Holders / Software Providers flow*

Figure 27. Data User flow

These flows will be evaluated and refined as part of the EUCAIM project, in the assessment of the Open Calls in WP7 and the Business Model Pilots in WP8 candidates. This will set the basis for the flows to be implemented in the Infrastructure Operational phase, after project end, when the Project Managerial structures will be replaced by the Infrastructure Central Hub managerial structures.

6. Operative boards managing the Central Hub

The management and governing structure of the Central Hub will play a crucial role in facilitating the coordination, governance, and efficient operation of the federated infrastructure. At the heart of this hybrid model, the Central Hub will serve as a cohesive entity that oversees the onboarding of Data Holders and users, ensures compliance with technical standards and data quality metrics, promotes collaboration with synergetic initiatives and ensures financing mechanisms for its long-term sustainability. This section provides the minimum management structure needed to operate the Central Hub. As such, the bodies described below correspond to boards put in place for the decision-making during the project execution. These boards themselves along with the range of their decision making may be subjected to modifications upon project end, once the final infrastructure is fully operative. Therefore, these are the boards governing the pilot phase of the project, while the proposed governance for the EDIC phase has been included in D8.2 and further updated in D8.3

The different decision-making boards that have been proposed to date are the following:

Figure 28. Operative boards managing the Central Hub

1. Steering Committee (SC)

The Steering Committee is the highest-level decision-making body of the infrastructure and project consortium. It currently consists of one representative of each project partner entity (as indicated in Table 2 "Members of the EUCAIM Steering Committee" of the Project Management Handbook, related with MS1.). It is chaired by the Scientific Coordinator and the members of the SC are required to be duly authorised to deliberate, negotiate and decide on all matters which fall under the responsibility of the SC as laid out in the Infrastructure Statutes.

During the project duration, the SC will discuss and decide on major modifications of the consortium membership (e.g., entry of new partners, withdrawal of partners), as well as on the work plan, project budget, intellectual property rights, etc. A more detailed description of these matters are listed in the Article 6.3.1 of the project's Consortium Agreement.

Upon project end, the SC is envisioned to have the last word in the decision-making of any unresolved matter at a lower level (e.g. Technical board, Access Committee). In this context, the SC will be convened ad-hoc by its Chair – the Scientific Coordinator. It is expected that each project partner will be represented at the meeting by its designated representative or by their proxy if the former is not available.

2. Management Board (MB)

During project execution, the Management Board (MB) is the operational body responsible for the monitoring of the technical progress of the project, quality

assurance, and the ad-hoc coordination of scientific and technological activities. It comprises the Administrative Project Coordinator, the Scientific Coordinator (SCo) (chair), and all Work Package leaders (WPLs) (see Table 4. Members of the EUCAIM Management Board in MS1: Project Management Handbook).

Currently, the Management Board is the interface between work packages and the Steering Committee and the Central Hub Office in charge of the daily responsibilities. The Management Board meets monthly by teleconference to ensure close monitoring of progress and quality assurance.

Upon project end, the MB is also envisioned to be in charge of any decision making regarding any technical implementations and quality control of all operations regarding the day-to-day functioning of the infrastructure, including the coordination of scientific activities around it.

3. Advisory Boards (AB)

External Advisory Boards are being set up throughout the course of the project to advise the Management Board on technical, ethical and related legal issues as well as on exploitation and regulatory matters. To date, two main external advisory boards have been proposed:

- Stakeholder Advisory Board - composed of representatives from diverse stakeholder groups including patients, citizens, clinicians, researchers, innovators, and regulators. Stakeholder AB members are part of the Stakeholder Forum, representing stakeholders who are not active partners in the project but whom the consortium wishes to work closely with. Its aim is to provide feedback on the project's development and progress, ensuring the final infrastructure meets stakeholder expectations and gains acceptance within relevant communities. This includes providing feedback on documents the project may wish to publish, or documents the consortium is completing – i.e., consultation on deliverables that affect stakeholders.
- Ethics Advisory Board - more particularly dedicated to advising the consortium on ethical and related legal issues, as well as on exploitation and regulatory matters. The Ethics AB is already formed by 3 members, as follows:

    - Prof. Nikolaus Forgó (University of Vienna, Austria)

    - Dr. Valentina Colcelli (National Research Council (CNR) Institute CNR-IFAC, Italy)

    - Ms. Magdalena Kogut-Czarkowska (Timelex, Belgium).

These boards will involve participants that are not part of the consortium members, in order to provide a fresh-eye, unbiased view on the decision making of the rest of boards.

Similar to the boards described above, upon project end the AB is envisioned to continue to provide external, unbiased advice on any decision-making regarding the day to day operations of the infrastructure, both at the technical and legal level.

4. Scientific Coordinator (SCo)

The Scientific Coordinator (SCo) of the project – Prof. Luis Marti-Bonmati from Beneficiary 2 HULAFE – leads the Central Hub operations in all scientific and technical aspects and provides strategic scientific guidance. The Scientific Coordinator is a central figure in conflict resolution and decision-making in the project management bodies and plays a central role in the monitoring of the Project's overall progress and strategic plans. The current expectation is that this role will continue beyond project end, this position converting into the Coordinator for Scientific Operations of the Central Hub.

5. Administrative Project Coordinator (AdmCo)

The administrative project coordinator (EIBIR) is currently responsible for the mediation between the project consortium and the funding authority, the EC. It is the main point of contact with the funding body (European Commission) and is responsible for the overall administrative and financial management of the EUCAIM project. The AdmCo is also tasked with the technical review of deliverables and milestones and financial reporting. The administrative coordinator is supported by the project coordination team (EIBIR, HULAFE, UPV, MAT) forming together the **Coordination Team.**

It is currently envisioned that the administrative coordinator will oversee all managerial aspects of the Central Hub Office, which overall purpose will be to support the implementations of the activities planned in the periodic strategic plan for the maintenance of the infrastructure.

6. Central Hub Office (CHO)

The Central Hub Office (CHO) will be responsible for all functions necessary in accordance with the infrastructure´s statutes, the needs of its ordinary functioning and compliance with the legal requirements for an entity of its nature (still to be defined as part of the project). The CHO will comprise these units: Central administration and coordination, Compliance, Intellectual Property, Business development unit, Technical and user management unit, Training, communication and outreach unit, External relations and project management unit.

. A distributed office including a physical core office working in tight coordination with part-time employees contributing from different entities and geographical locations is an option currently being explored. The structure, roles and implementation of the CHO is defined in the D8.2 (Plan towards the creation of the Central Hub Office), submitted in M18.

7. Technical Board (TB)

The Technical Board (TB) is currently tasked with the review of the potential engagement of software and service providers to EUCAIM. Technical partners have the responsibility to adopt a responsible research and innovation attitude when designing and developing their solutions, by following the guides and requirements of

the ethical committees, with the lead and support of the Data Protection Task Force and Ethics Advisory Board.

Currently and during project execution, the main responsible of the TB is the Technical Coordinator and WP4 leader of the EUCAIM project, Prof. Ignacio Blanquer Espert from BEN 3- UPV. The list of TB members (ordered by WP) is the following:

- WP2: Patricia Serrano Candelas (HULAFE), Hanna Leisz (DKFZ), Laure Saint-Aubert (MEDEX).
- WP4: Ignacio Blanquer Espert (UPV), Marcel Koek (Erasmus MC), Leonor Cerdá Alberich (HULAFE).
- WP5: Gianna Tsakou (MAGGIOLI), Manolis Tsiknakis (FORTH), Melanie Sambres (LIMICS).
- WP6: Salvador Capella-Gutierrez (BSC), Josep Lluís Gelpi (UB).
- Research Infrastructure (Technical expert): Petr Holub (BBMRI).

Being EUCAIM an infrastructure that is aimed to be kept alive for as long as possible beyond project end, it is envisioned that as technology evolves new software will need to be incorporated in the infrastructure in the future. In this context, it is expected that the TB will continue to facilitate the decision-making regarding the addition of new software and technologies in this context.

8. Ethical and Legal Board (ELB)

The Ethical and Legal Board (ELB), during the duration of the project and beyond, will be the body in charge of ensuring that no EU rule is violated, while ensuring that the research conducted is up to the accepted EU standards. In this context, the term "Ethics" refers to questions of legal and regulatory compliance that constitute a part of the governance process. In EU-funded projects, ethics is deemed a transversal issue and Ethics Advisory Board a key oversight mechanism to ensure understanding of the Ethics Appraisal Procedure, proper implementation of the Ethics Requirements, addressing specific issues such as Privacy and Data Protection Impact Assessments or Artificial Intelligence and ensuring ethics compliance in general.

The ELB will act as a contact point for guidance on ethical issues that may arise during project execution and beyond project end, working in close connection with any party saddled with ethics-related responsibilities. During the project execution, the ELB will be chaired by the WP3 leaders and composed of legal experts in the participating entities. Beyond project end, the members of this board may be reselected based on availability.

9. Access Committee (AC)

The mission of the EUCAIM Access Committee is to ensure responsible and secure access to the infrastructure, its available data and services, promoting valuable research while upholding ethical and privacy standards. To do so, the Access Committee is being designed as a dedicated body responsible for overseeing and managing the access requests to the infrastructure´s data issued by authenticated

users. As shown in Figure 28, the Access Committee is under authority of the Steering Committee and the Management Board. The Steering Committee is the main decision-making body within the EUCAIM project while the Management Board is the main operational body. The Management Board is the interface between work packages and the Steering Committee and is responsible for monitoring the technical progress of the project, quality assurance, and the ad-hoc coordination of scientific and technological activities.

In turn, the Access Committee includes representatives of all different WPs for ensuring secure access to data and services that also complies on security and ethical levels. As the Access Committee is executing a practical activity within the project, the Management Board is responsible for it. Therefore, the Access Committee reports to the Management Board, which in turn reports to the Steering Committee as necessary.

During the project, the Access Committee establishes access rules and evaluates the applications received both for the internal and external open calls. The external open call has the objective to include more partners to the consortium and it was launched with a deadline 10th of June of 2024. On the other hand, the internal call aims to transfer use cases to the platform, it was open once in the first semester of the year 2024, and it is going to reopen in November 2024 and it will still open until the end of the project. The rules of the internal call have been redefined in the second round.

These are the main functions of the Access Committee:

1. Develop the rules and procedures to facilitate the access to the EUCAIM platform as a Data Holders, Data users and Software Providers, in compliance with the EHDS regulation [1] including GDPR [2], before the launch of the first open call for inclusion of new data providers which is foreseen to happen during 2024.
2. Provide strategic advice on the topics of open calls for new Data Holders / use cases.
3. Evaluate the scientific relevance, impact and innovation of proposals for new Data Holders/use cases / access requests.
4. Make recommendations to the EUCAIM Management Board on acceptance/rejection of new Data Holders or use cases requests, also based on the results of the ethical and technical evaluations provided by the Ethical and Legal Board and by the Technical Board.
5. Review the access requests based on the established technical, legal, organisational requirements and on the scientific value of the user project.
6. Perform its tasks in due cooperation and in consultation with the EUCAIM Management Board, Ethical and Legal Board, and, if necessary, seeking the advice of external/additional experts, and in compliance with the European GDPR regulation.

With respect to the publicity of the decisions of the AC, the possibility to appeal and the transparency of the decisions, there are two different scenarios:

- In the evaluation of the applications of the External Open Call, the decisions achieved are based on the rules indicated in the text of the Call (eligibility, scientific

relevance, impact and implementation. A summary of the results will be published on the website of EUCAIM. This will be an overview of the selected applications with broad descriptions, but no detailed scores for the applications. Further transparency will be achieved by sending the applicants the results of the evaluation of their applications, indicating the score in each part. These individual, detailed applications will however not be published on the website. In this process it is also not possible to appeal.

- In the applications of data users requesting for data, and for including new data holders and software providers, the decisions will be directly shared with the interested parties, as well as the results of the evaluations based on the scientific relevance, impact and innovation, ensuring the transparency in this decision. In this case, data users, data holders and software providers will have the possibility to appeal and establish a conversation with EUCAIM in order to improve their proposal.

The composition of the Access Committee is also already defined:

- The EUCAIM Access Committee is formed by two people from each of the project work packages and the Scientific Coordinator, unless otherwise decided by the EUCAIM Management Board.
- The EUCAIM Access Committee is composed of internal highly qualified, internationally recognized senior experts for medical imaging, image data management, research infrastructure management and ethical issues related to clinical data, each selected on the basis of their competence and covering all aspects of the activities of the EUCAIM platform.
- The EUCAIM Access Committee members were nominated by WP leaders and approved by the EUCAIM Management Board. A Chair of the Access Committee will be nominated by the Access Committee members and will be in charge for the duration of the appointment.
- The Access Committee members will be appointed for the entire duration of the project (from June 2023 to December 2026). In case some member becomes unavailable, his/her replacement must be approved by the Management Board.
- The Management Board should ensure a scientifically, geographically and gender-balanced composition of the Access Committee.

Regarding the conflicts of interest, it has been decided that:

- Access Committee members shall sign a 'Conflict of Interest' declaration and an agreement to comply with GDPR during the evaluation of the applications/requests.
- Access Committee members must indicate if they are in Conflict of Interest with whatever assigned applications that they are requested to evaluate.

More information on the process in the evaluation of the proposals in the External Open call can be found in section 2.4 (Evaluation of applications) of the D7.1, Rules for evaluation and prioritising of use case applications from the open call.

Being an infrastructure that is aimed to be kept alive for as long as possible, it is envisioned that access requests will continue to happen on a regular basis during project execution and beyond. Therefore, it is envisioned that an Access Committee board will be needed in both contexts and likely periodically renewed.

10. Data Protection Task Force

The Data Protection Task Force will be a body that plays the role of the Data Protection Officer (DPO) during both the project execution and beyond. It will monitor internal compliance, inform, and advise on data protection obligations, provide advice regarding Data Protection Impact Assessments and act as a contact point for all the partners and data subjects (the results of this task being documented in D3.6 - Data Management Plan). During the project execution phase, the main representatives of this task force will involve the DPOs of each consortium partner. Upon project end, the members of this board may need to be re-elected.

7.   Templates for Service Level Agreements

This section includes the bases of the Service Level Agreement (SLA) template that will be used to regulate the integration of a local node in the EUCAIM Federation. Other similar SLAs will be subscribed with the providers or services of the platform. We anticipate that this template will need to be tailored to each specific provider, but it may help to facilitate the discussion with the provider. The final acceptance of the SLA of a specific provider will be issued by the Governing bodies. The document does not regulate the Terms of Use of the users in the platform, but the obligations of the provider to join the federation. Moreover, this document is not the compulsory template to be used but a draft to be used as a basis for developing the specific SLAs with the providers.

The template of the SLA can be downloaded from the https://dashboard.eucaim.cancerimage.eu/eucaim-sla-template. The SLA of the UPV reference node is available in https://eucaim-node.i3m.upv.es/dataset-service/web/sla.pdf as a reference.

Service Level Agreement for the repository

The Service and Access

The services of the **<NAME OF THE PROVIDER NODE>** are provided by the **<NAME OF THE INSTITUTION>** ("the Provider"). This clause defines the Level Service Level Agreement (SLA), valid for the services of the **<NAME OF THE PROVIDER NODE>** provided to support the access and processing of the data held in this node in the context of the EUCAIM Infrastructure.

This SLA will enter in force in the signature of the document and will have a duration of **<DURATION>** months.

The provider commits to the EUCAIM Infrastructure through the **<NAME OF THE PROVIDER NODE>** the following resources:

- **<NUMBER AND TYPE OF GPUs>**
- **<NUMBER AND TYPE OF CPU CORES>**
- **<RAM MEMORY>**
- **<PERSISTENT STORAGE>**
- **<ANY OTHER RELEVANT EQUIPMENT>**

Resources that will be shared among the users and provided "as available".

The access to the repository will include **<REMOVE THE CONDITIONS THAT ARE NOT FULFILLED>:**

- Tier 1 compliance: Availability of a FAIR Data Point entry per dataset providing the metadata information of each dataset in RDF format and following the specification of the EUCAIM Hyper-ontology.
- Tier 2 compliance: Availability of a Federated Search mediator endpoint to query the information about the datasets in the **<NAME OF THE PROVIDER NODE>** and connected to the EUCAIM Federated Search service.
- Tier 3 compliance:
  - Access to a virtual environment to browse and process (including GPUs and subject to the availability of the resources) the data that the user is authorised to access.
  - Access to processing facilities through EUCAIM Federated Processing service.
- Authentication and Authorization of users through LS-AAI, including a VO sub-group named **<NAME OF THE SUBGROUP>** in the EUCAIM organisation. Authorisation could be delegated to the LS-AAI (LS-AAI will block users that do not belong to the subgroup) or implemented at the Data Holder's AAI service based on the entitlements of the LS-AAI.

The use of the resources should be proportional and reasonable and could be revoked by the provider in case of misuse or abuse, or temporarily if an urgent request due to a public health emergency would require those resources..

Service Conditions

The service is in general delivered 24 hours per day, 7 days per week basis (i.e. 365 days or 8,760 hours), to seamlessly support operations. Planned and announced interruptions may reduce the effective operating time of service.

The following exceptions apply:
- Users will be notified via e-mail in a timely manner, (i.e. at least 72 hours before the start of the outage by e-mail), about the planned maintenance windows or service interruptions ("scheduled downtimes").
- Unplanned service interruptions will be accompanied by an explanation.
- Downtime periods exceeding 48 hours need justification.
- Human services are provided only during support hours.

Support is available between

- Service issues: Monday to Thursday from 9:00 to 17:00 CET/CEST time, Friday from 9:00 to 14:00 CET/CEST.
- This excludes public holidays at the same time in the organization providing the service.

The Quality of Support level in terms of response time and resolution is defined as follows:

- Low priority, max 5 working days.
- Normal, max 2 working days.
- High, max 1 working day.

Any unsuccessful termination of a request should be clearly justified and documented.

Availability and Reliability

The Service level will be:

- Monthly Availability defined as the ability of a service or service component to fulfil its intended function at a specific time or over a calendar month.
  - Average service level target (as a percentage per month): 90%
  - Minimum (as a percentage per month): 80%
- Monthly Reliability, defined as the ability of a service or service component to fulfil its intended function at a specific time or over a calendar month, excluding scheduled maintenance periods.
  - Average service level target (as a percentage per month): 95%
  - Minimum (as a percentage per month): 85%

The provisioning of the service under the agreed service level targets is subject to the following limitations and constraints:

- Support is provided in the following languages: Spanish and English.
- Downtimes caused due to upgrades for fixing critical security issues are not considered Agreement violations.
- Force Majeure. A party shall not be liable for any failure of or delay in the performance of this Agreement for the period that such failure or delay is due to causes beyond its reasonable control. Means any Fire, flood, earthquake or natural phenomena, War, embargo, riot, civil disorder, rebellion, revolution which is beyond the Providers' control, or any other causes beyond the Providers' control.

Contact

The **<NAME OF THE PROVIDER NODE>** defines the following contact points:

- Responsible Person of the platform: **<Name, position and e-mail>**
- Responsible Person of the datasets: **<Name, position and e-mail>**
- Technical Contact point: **<Name, position and e-mail>**

- Security Contact point: **<Name, position and e-mail>**

Additionally, the users of the platform could contact the platform staff for general matters by email to **<contact-email>** and/or through the **<issue-tracker>** or by regular mail at **<regular mail address>**. This will be used only under the failure of the issue tracker services. The Quality of Service Response times do not apply to these alternative communication means.

8.   Conclusions

This deliverable provides the final description of the various user roles that will interact with the EUCAIM platform as well as the specific actions that are expected to be undertaken by them. This has led to the definition of the User Actions and User Stories, which, along, have defined the related operational procedures. The rules for participation have been moved to D4.3 and D4.4. The project consortium hopes this deliverable sets the first grounds on how to use the EUCAIM infrastructure securely and effectively. This document also provides an overview of the governance bodies responsible for the operation of the Central Hub during the pilot phase of the project, offering definitions about their roles, responsibilities, and decision-making processes. The proposed governance for the EDIC phase has been included in D8.2 and further updated in D8.3.

It is worth noting that the elaboration of this document has been integral to the establishment of the project's glossary of terms, which is currently being updated and aims to provide vocabulary standards for use throughout the infrastructure's implementation and beyond.