



**Project title:** European Federation for Cancer Images

**Project acronym:** EUCAIM

**Grant Agreement:** 101100633

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

## D4.1: First EUCAIM Operational Platform

**Partner(s):** HULAFE

**Author(s):** Ignacio Gómez-Rico, Leonor Cerdá, Irene Marín, Ana Miguel, Miguel Ángel Herrero, Pedro Mallol, Luis Martí Bonmatí (HULAFE)

**Contributors:** Ignacio Blanquer (UPV), Gianna Tsakou (MAG), Ricard Martínez (UV), Janos Meszaros (BBMRI), Amelia Suárez (MAT), Sara Zullino (EATRIS), Linda Chaabane (EURO-BIOIMAGING), Valia Kalokyri (FORTH), Sara Colantonio (CNR), Raul Costa (ITI), Katrine Riklund (UmU), Peter Gordebeke (EIBIR)

**Date of delivery:** [30/06/2023]

**Version:** 1

## Table of contents

1. Introduction	3
Aim and scope of the deliverable	3
Relation with other deliverables and WPs	4
2. User roles	5
Management Profiles	9
3. User stories	9
4. User Actions	26
User Actions enabled for all Roles	27
User Actions enabled for Data Holders	27
User Actions enabled for Tool Providers	28
User Actions enabled for Data Users-Researchers	29
User Actions enabled for the Governing Body	30
User Actions enabled for the EUCAIM Platform Manager	31
5. Operational procedures	31
Data provision	31
Tools / service provision	33
Data Access Requests	35
Other considerations	37
6. Operative boards managing the Central Hub	39
7. Rules for Participation	43
Minimum requirements in terms of data and infrastructure	43
Expected response times	45
Data Quality	45
Compliance framework design	45
8. Templates for Service Level Agreements	47
9. Conclusions	50

## 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  
DPO = Data Protection Officer  
ELB = Ethical and Legal Board  
ETL = Extract Transform Load  
EU = European Union  
FAIR = Findable, Accessible, Interoperable, Reusable  
GA = Grant Agreement  
GDPR = General Data Protection Regulation  
GPU = Graphics Processing Unit  
IPR = Intellectual Property Rights  
MB = Management Board  
MR = Magnetic Resonance  
MS = Milestone  
R&D = Research and Development  
RI = Research infrastructures  
SC = Steering Committee  
SCo = Scientific Coordinator  
SLA = Service Level Agreement  
TB = Technical Board  
WP = Work Package  
WPLs = Work Package leaders

# 1. Introduction

## Aim and scope of the deliverable

The EUCAIM project aims to deploy a pan-European digital federated infrastructure of FAIR, de-identified, cancer imaging data from daily clinical practice (real world data) that will be used to develop, validate and benchmark AI tools towards precision medicine. The EUCAIM platform will address the fragmentation of existing cancer image repositories by building on repositories including data from various sources such as clinical imaging, pathology, molecular and laboratory data. Ultimately, the platform aims to support clinicians, researchers, and AI innovators in building validated clinical decision-making systems to benefit citizens.

The present document provides a comprehensive description of the various user roles that will interact with the EUCAIM platform, outlining the specific actions expected from each of them within the infrastructure and through the EUCAIM Dashboard, which is the website intended for users who want to use EUCAIM data 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 as well as to the 'Architecture of the Federated Repository' document, both of which are presently still being finalised.

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, D4.1 provides clear guidance on how to navigate and utilise the EUCAIM infrastructure 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, D4.1 encompasses a comprehensive set of definitions for specific terminology used within the EUCAIM project. 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 project's life span. These terms will serve as the foundation for the project's glossary of terms, which is currently under construction.

Along with the operational procedures defined, D4.1 presents an overview of the governance bodies responsible for the operation of the Central Hub (comprising the Central Repository/Storage and the services and tools 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 has been established within this deliverable towards MS23. These rules specify the minimum requirements in terms of data and infrastructure that participants must meet, as well as an outline of the compliance framework under which holders are expected to operate. The document also establishes the foundation for the Service Level Agreement (SLA) template, which will 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 participants.

Please note that the content of this document is subject to potential changes based on the project evolution and on the final development of two proposed legislations: the "Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space"<sup>1</sup> and the "Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS."<sup>2</sup> Our objective is to ensure maximum alignment with these forthcoming regulations.

The guidelines and procedures outlined in this document may need revisions to ensure full compliance with the approved legal texts once they are officially adopted by the European Parliament and Council, but also to reflect the evolution of the EUCAIM project over time and accommodate updates on user requirements as they evolve. The corresponding EUCAIM partners will incorporate any necessary changes into updated versions of this document promptly. This approach guarantees conformity and consistency with the new binding rules at the earliest possible opportunity.

We kindly request all readers of this document to be aware of the likelihood of these changes and remain attentive to any updates communicated regarding this matter.

## Relation with other deliverables and 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, D4.1 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 this document will be further documented in D4.3 (First rules of participation report) and finalised in D4.4 (Final rules of participation report). Likewise, operational procedures herein defined will be further refined and documented in deliverables D5-12, all of 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

---

<sup>1</sup> Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM/2022/197 final of 3 May 2022. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52022PC0197>

<sup>2</sup> Proposal for a Regulation of the European Parliament and of the Council Laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. COM/2021/206 final of 21 April 2021. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>

(D4.12). All necessary updates derived from the above-mentioned deliverables, where applicable, will be documented in D4.2 (Final EUCAIM Operational platform, due m24).

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's 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 Tools, etc.), which starts an interaction flow (User Actions) supported by the components of the final architecture (section 3). The user roles identified so far are as follows:

1. **Data Holder<sup>3</sup>/Data Provider/Data Controller:** A Data Holder/Data Provider/Data Controller (from here onwards in the present document referred to as "Data Holders" only) refers to any natural or legal person, including entities, bodies, and research organisations in the health or care sectors, as well as European Union institutions, bodies, offices, and agencies, who has the right, obligation, or capability to make certain data available, including registering, providing, restricting access, or exchanging the data. 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.

In a nutshell, Data Holders have two options for joining the federation: they can become a federated node or upload anonymised data to the Central Repository. In both cases, they are required to prepare the (meta)data of their collection following EUCAIM's specifications (guidelines will be provided by WP5) and make it available to the central services as part of the metadata catalogue shared through the EUCAIM infrastructure. The different types of Data Holders are defined in the D2.1. Onboarding invitation package. The following bodies can be invited as Data Holders via the D2.1. Onboarding invitation package of the project:

1. Repositories with data storage and processing capabilities. These repositories will be federated, and data processing will be performed in a distributed way, respecting the access conditions of each repository, and also supporting federated learning.
2. Repositories with data storage but no processing capabilities. Data storage will be federated and intensive data processing on anonymised data will be performed in the Central Hub or external computational infrastructures or through third party agreements and temporary data transfer.
3. Repositories with neither data storage nor processing capabilities. This could be the case of imaging repositories created in the framework of a research project that is terminated or near termination. In order to make their data

---

<sup>3</sup> Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space COM/2022/197 final

collections sustainable, they may upload anonymised data in the EUCAIM Central Storage.

4. Clinical Data Holders that are willing to set up a federated node that will connect to the federation. Data processing will be performed in a distributed way, supporting federated learning.
5. Clinical Data Holders that are not willing or do not have the resources to set up a federated node and will upload anonymised data in the Central Storage.
6. Research infrastructures (RIs) providing distributed federated services that link Data Holders) with existing anonymised and minimally annotated research data, which could be linked to the EUCAIM infrastructure in a federated way. RIs should expose or register their catalogues and interact with EUCAIM regarding data access requests. Ideally, RIs should also support data discovery through federated query and distributed processing, if possible.
7. Data altruism organisations and patient associations that will either connect to the federation or upload anonymised data in the Central Storage.

Additionally, it is important to note that Data Holders may offer different access conditions for the data: Authorisation to download the datasets, authorisation to access remotely, view and process in-situ the datasets and authorisation to remotely process the datasets without the ability to access and visualise data, even remotely.

Another important aspect to note is that the hospitals affiliated with EUCAIM will not expose their data warehouses to the federation. These data warehouses may already exist or may have been created for EUCAIM. Instead, they will be approached individually each time there is a new research project on specific clinical cases. If they choose to participate in the project, the hospitals will prepare the necessary anonymised collections within their data warehouses to be shared with the federation through a federated node or by uploading them to the Central Storage. Therefore, hospitals will only expose metadata for specific datasets from projects in which they have chosen to participate upon request.

Data Holders are also responsible for exposing clinical data and imaging metadata in a standardised manner, using the agreed Common Data Model (CDM) schema and standard vocabularies to be defined during the project, so that the data can be queried and used in federated learning processes to train AI models requested by the researchers, acting as a node within the federated network. In case of Data Holders that do not have their data ready following the agreed CDM schema, they will need to use the tools and guidelines provided by EUCAIM to make their data interoperable, or use a mediator, to match the terms in the query. To comply with relevant data privacy and security regulations, including the GDPR, Data Holders should ensure that the clinical and imaging metadata they expose meets all applicable standards and legal requirements. Furthermore, they must ensure that their data is of high quality, accuracy, and up to date to maximise its utility in the federated learning processes.

As part of the Data Holder role, a Local Data Manager, who is an authorised technical expert (or a team of experts) on the provider's site, is the responsible individual for installing, configuring, operating, and maintaining the local services that support the federation or handling the uploading of the data to the central storage, as well as the

metadata in both cases. The authorised technical expert will have the full support of the EUCAIM helpdesk and all corresponding teams.

Example of a **Data Holder**: One or more research institutes that will provide imaging and clinical data to the platform through a research project. The entity is willing to contribute data for research on an altruistic basis and as part of its business model. The site has a large amount of data gathered over the last 10 years and can create specific datasets according to the GDPR. Its only restriction is that all the data must be fully anonymised and the users should request permission before accessing the data.

2. **Tool Provider**: The Tool Provider refers to the entity (startups, enterprises, research institutions, government agencies, non-profit organisations) that would like to contribute with processing tools, services, data preprocessing tools, or applications they have developed to the EUCAIM's marketplace for use in the federated processing or data pre-processing purposes of the platform. The Tool Provider must ensure that their developed tools comply with the network's technical requirements and standards. When required, the Tool Provider must prove that the tools 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 tools, 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. The tool 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, Tool Providers should offer their tools 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 tools 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 Tool 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 **Tool Provider**: A startup that provides, after agreement, an AI explainability platform that 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 tools.

3. **Data User-Researcher**: The Data User-Researcher refers to any person or entity that wants to explore the public catalogue of available (meta)data and eventually request access to data and process them using either the tools available in the platform or their own AI tools. This data access request by the Data User-Researcher should be made under a Research and Development (R&D) project (approved by a competent Ethics Committee of the country where the project is going to be conducted), which will be evaluated by the Access Committee. Within the different purposes for requesting data access, we can find the specific profile of a Data Scientist, who is anyone willing to develop / improve AI tools, and will be provided with tools, processing environment and computational resources. If the request by the logged-in user is accepted, the Data User-Researcher gains data access.



Data Users-Researchers 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 User-Researchers may require specialised access to certain datasets or tools that enable complex data analysis.

Data Users-Researchers with a Data Scientist profile typically request data to develop and test statistical or machine learning models, algorithms, or predictive models. Their focus is on deriving insights from data that can be used to improve business decisions, optimise processes, or develop new products or services. Data scientists often require specialised data mining and analysis skills to clean, process, and analyse data to uncover meaningful insights.

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

Example of a **Data User-Researcher with a Data Scientist profile**: A Data Scientist is developing an AI tool to analyse health images and related clinical and molecular data on the most prevalent cancers in Europe. The scientist has an initial model they want to improve with new data. They are seeking quality and labelled data, and do not accept unstructured data or data without a logical folder structure.

4. **Research Communities**: Research Communities refer to groups or entities with a common research goal, typically formed through the course of already finalised, currently ongoing or newly emerging projects, that would like to make use of EUCAIM's research environment to continue the research their original project facilitated in the first place. With this, the community taking part in that project (e.g. consortium) will need to agree to share the data collected (together with the tools developed through the project lifespan where applicable) to EUCAIM's Central Repository. The expectation is that the Research Community (RC) will remain connected via EUCAIM and will as a result be able to continue and further expand the work done in the scope of such a project via EUCAIM. In return, EUCAIM will include the related datasets in its catalogue, providing the RCs with a secure and highly interoperable environment and enabling them to initiate new projects within the EUCAIM infrastructure, while establishing new collaborations with other partners connected to EUCAIM.

Example of **Research Communities**: AI4HI is a European group of researchers, data scientists, and healthcare professionals focused on applying artificial intelligence techniques to improve healthcare outcomes. As a Research Community within EUCAIM, AI4HI contributes entire finished projects, such as PRIMAGE, that involve developing AI algorithms for medical image analysis, clinical decision support systems, and predictive models for two types of paediatric cancer diagnosis, prognosis, and therapies follow-up. Thus, the entire finished project could be transferred to the Central Hub, so that the data can still be used and they can further contribute to the project. They share their expertise, data, and tools with other members of the EUCAIM

federation, promoting collaboration and innovation in the field of AI-driven healthcare solutions through new research projects, engaging new partnerships.

## Management Profiles

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

1. **EUCAIM Platform Manager:** The EUCAIM Platform Manager is a technical expert or team of experts who will operate the core services of the EUCAIM platform. The Platform Manager is responsible for managing and maintaining the underlying EUCAIM technical infrastructure, including the central storage, servers, databases, and other resources. The Platform Manager manages user accounts and access permissions, deploys applications, and services, uploads new applications to the marketplace (provided by Tool Providers), and ensures their proper integration into the platform. They support the orchestration of federated processing, working with Data Holders and Tool Providers to integrate metadata, tools, and services, and ensuring that Data User-Researchers 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 provides user support, responds to inquiries, provides documentation, and troubleshoots issues that arise with the platform.

Example of an **EUCAIM Platform Manager:** A system administrator employed by the institution that operates the central storage 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 should 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 a 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.

Example of **Governing Body:** The EUCAIM Access Committee (the responsible of accepting or rejecting data access requests) receives access requests to specific datasets of the Atlas of Cancer Images from Data Users-Researchers. 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.

## 3. User stories

User Stories are descriptions of full interactions of a User Role with 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 25 user stories that identify actions that could be performed by the actors from the above user profiles. These user stories are listed in Table 7 and Table 8.

*Table 7. List of User Stories related to the Dashboard.*

<b>#</b>	<b>User Story</b>	<b>User Roles</b>	<b>Mapping</b>
usDH1	Application to join the federation	Data Holder	Local Data Manager
usDH2	Setup of the data node to connect to the federation	Data Holder	Local Data Manager
usDH3	Quality control (legal / technical) and data preparation from a provider (on premises)	Data Holder	Local Data Manager
usDH4	Apply pre-processing tools to on-premises data	Data Holder	Local Data Manager
usDH5	Uploading metadata to the Dashboard	Data Holder	Local Data Manager
usDH6	Ingest data on the Central Storage	Data Holder	Local Data Manager
usDH7	Monitoring the access of data collection	Data Holder	Local Data Manager
usTP1	Application of a tool to the federation marketplace	Tool Provider	Data Scientist/ Bioengineer
usTP2	Monitoring the access of tools	Tool Provider	Data Scientist/ Bioengineer
usDU1	Exploration of collections from the public catalogue	Data User- Researcher	Researcher/ Data Scientist
usDU2	Federated search of aggregated data in the collections	Data User- Researcher	Researcher/ Data Scientist
usDU3	Request access to the collections from the User's Catalogue	Data User- Researcher	Researcher/ Data Scientist
usDU4	Get an overview of the datasets to which they have been granted access	Data User- Researcher	Researcher/ Data Scientist
usDU5	Select the data to be used from the accessible data collections	Data User- Researcher	Researcher/ Data Scientist
usDU6	Exploring the tools available in the federation and their provenance	Data User- Researcher	Researcher/ Data Scientist/ Processing Requester
usDU7	Configure tool settings and computation and storage requirements for data	Data User- Researcher	Researcher/ Data Scientist

#	User Story	User Roles	Mapping
	processing		
usDU8	Process (distributed) the data from the federation by using a tool of the catalogue	Data User-Researcher	Researcher/ Data Scientist
usDU9	Run a Federated AI model training	Data User-Researcher	Data Scientist
usDU10	Monitoring job status in the processing	Data User-Researcher	Data Scientist
usDU11	Retrieve and review obtained results	Data User-Researcher	Researcher/ Data Scientist

Furthermore, other User Stories have been identified that outline actions that can be performed by EUCAIM management profiles, who will interact directly or indirectly with the EUCAIM platform.

*Table 8. List of User Stories related to the Dashboard performed by management profiles.*

#	User Story	User Roles	Mapping
usGB1	Evaluate and accept/reject data collections from Data Holders	Governing Body	Access Committee
usGB2	Evaluate and accept/reject tools from Tool Providers	Governing Body	Access Committee
usGB3	Evaluate and accept/reject a data access request application	Governing Body	Access Committee
usPM1	Uploading of a tool to the federation marketplace	Platform Manager	
usPM2	Platform services administration	Platform Manager	

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

*Table 9. Application to join the federation.*

<b>Story name:</b> Application to join the federation	<b>Related WP:</b> WP2
<b>Triggered by User Role (Fictitious Name):</b> Data Holder (Raul)	<b>Related User Roles (Fictitious Name):</b>
<b>Overview Description:</b> Raul is a member of a research organisation that seeks to join the EUCAIM federation and to contribute imaging and clinical data to the platform. The organisation, as a Data Holder, is willing to contribute data for research on an altruistic basis and as part of its business model.	

The provided data will be used in federated learning processes to train AI models by researchers, within the federated network.

Raul's organisation has accumulated a large amount of imaging and clinical data over the past 40 years. Compliance with data privacy and security regulations, including GDPR, is a top priority for them. Thus, they ensure that all clinical data exposed will be fully anonymised and that users must obtain permission before downloading the data.

To apply for membership in the federation, Raul accesses the Dashboard and gets informed about the process and conditions of data sharing. Then, he follows the documentation available in the Data Holders area to register his organisation as a potential Data Holder for the federation. Raul submits an application that highlights the research organisation's capacity to provide valuable datasets while adhering to stringent data privacy, anonymisation, accuracy, and data quality standards. The application demonstrates their commitment to maintaining high-quality and up-to-date data that will be aligned with EUCAIM's agreed CDM schema and standard vocabularies, maximising interoperability, and usefulness in federated learning processes. Moreover, his entity has (or will purchase) 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 researchers to access and utilise their available clinical and imaging data 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.

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

*Table 10. Setup of the data node to connect to the federation.*

<p><b>Story name:</b> Setup of the data node to connect to the federation</p>	<p><b>Related WP:</b> WP2, WP4, WP5</p>
<p><b>Triggered by User Role (Fictitious Name):</b> Local Data Manager (Carol)</p>	<p><b>Related User Roles (Fictitious Name):</b> Platform Manager (Sergio)</p>
<p><b>Overview Description:</b></p> <p>Carol, as 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), is responsible for the configuration of the node that will be connected to the network, which will be able to execute federated processing (accessing only previously anonymised collections selected within the scope of a research project).</p> <p>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 and instructions on how to install and configure the necessary software components and dependencies to connect the node to the network. She accesses the tool marketplace and searches for the software needed to set up the connection, downloads it and installs it in the node. In addition, within this authenticated workspace she can find the references and contacts of the Technical Support team and the access links to the Helpdesk, to receive support for any of the processes that she needs to perform.</p> <p>Once she finishes the software installations and configurations, she contacts Sergio (Platform Manager) to test the integration of the node with the rest of the platform, validating the</p>	

connectivity, and launching multiple test jobs with validation/test collections (datasets containing imaging + clinical data).

Example of **Quality control (legal / technical) and data preparation from a provider (on premises)** User Story:

*Table 11. Quality control (legal / technical) and data preparation from a provider (on premises).*

<b>Story name:</b> Quality control (legal / technical) and data preparation from a provider (on premises)	<b>Related WP:</b> WP2, WP3, WP5
<b>Triggered by User Role (Fictitious Name):</b> Local Data Manager (Carol)	<b>Related User Roles (Fictitious Name):</b>
<p><b>Overview Description:</b></p> <p>Carol is the Local Data Manager in a Data Holder center that has recently joined the EUCAIM federation (meaning that the center 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 and the Central Hub.</p> <p>In order to prepare the dataset (which will be extracted from a Data Warehouse already modelled using a Common Data Model and standard vocabularies) she has to apply some pre-processing tools for it to be compliant with the EUCAIM data collections requirements (the main aspects to be processed/evaluated are FAIR compliance, anonymisation of the collections, data curation, quality assurance and data harmonisation).</p> <p>To do it, after reviewing the related guidelines available in the dashboard, 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 tools that she wants to run over the data collections that she selected. She receives the necessary information to be able to deploy the tools locally, avoiding federated processing or any kind of orchestration coming from the Central Hub.</p> <p>Once she applies these pre-processing tools she is able to explore, perform a quality control process and validate the data using tools that may or may not come from the set of tools available at the marketplace.</p>	

Example of **Apply pre-processing tools to on-premises data** User Story:

*Table 12. Apply pre-processing tools to on-premises data.*

<b>Story name:</b> Apply pre-processing tools to on-premises data	<b>Related WP:</b> WP5
<b>Triggered by User Role (Fictitious Name):</b>	<b>Related User Roles (Fictitious Name):</b>

<b>Name):</b> Data Holder (Raul)	
<b>Overview Description:</b> <p>Raul, a member of a hospital that has recently accepted to participate in a research project in the EUCAIM federation, wants to pre-process his already curated and anonymised collection of multiparametric MR imaging data of prostate cancer and a collection of lung-nodule CT data, by using some pre-processing steps. In particular, he wants to annotate his prostate dataset and co-register T2-weighted and ADC modalities.</p> <p>He logs via the authentication and authorisation service (AAI) and accesses the Data Holders working space. He then accesses the marketplace looking for the pre-processing applications that may serve his needs. He is presented with an initial set of pre-processing tools, such as data annotation, data visualisation, data exploration, ... and an open text field for filtering the tools. This time, he selects the one for annotating the data, which opens a dedicated frame/window offering him several menus and functionalities.</p> <p>In a federated manner, he can utilise this tool with his own resources. Subsequently, he chooses a T2-weighted MR series and employs the annotation tool to manually draw or review the suggested automated segmentation of the prostate gland.</p> <p>Once he has processed the whole dataset, temporarily stored in this local workspace/vault/environment, he saves the newly pre-processed (i.e., annotated) dataset in his node within the platform. This dataset, with its metadata reporting the information about the pre-processing apps applied, the time and actors involved (i.e., Raul) is registered into the dataset Catalogue.</p>	

Example of **Uploading metadata to the Centra Dashboard** User Story:

*Table 13. Uploading metadata to the Dashboard.*

<b>Story name:</b> Uploading metadata to the Dashboard	<b>Related WP:</b> WP2, WP7
<b>Triggered by User Role (Fictitious Name):</b> Local Data Manager (Carol)	<b>Related User Roles (Fictitious Name):</b>
<b>Overview Description:</b> <p>Carol, as the Local Data Manager in a clinical partner of EUCAIM, is responsible for uploading metadata to the Dashboard. Recently, Carol's hospital agreed to participate in a research project within the EUCAIM federation. As part of this collaboration, the hospital has prepared a curated and anonymised collection, ready to be shared with the platform through their local federated node.</p> <p>Carol now focuses on preparing the metadata collections. The goal is to contribute to the metadata catalogue, which showcases metadata from diverse datasets and collections provided by different Data Holders within the federation. It also aims to make the available data interoperable with the rest of the EUCAIM federated data.</p> <p>To ensure data security, the EUCAIM platform implements an access system with different levels of privileges. A public catalogue is available to anonymous users, offering a limited view of the metadata and basic search options. However, advanced searching and the complete catalogue view are restricted to authenticated users. This approach guarantees that only registered individuals have access to the detailed catalogue information.</p>	

Once the metadata collections are prepared, Carol proceeds to upload them to the EUCAIM Dashboard. She takes care of organising and structuring the information, ensuring that the descriptions of the datasets are comprehensive and accurate. This includes details about the dataset's provenance, access conditions, and the processing services available at the provider's side.

Example of **Ingest data on the Central Storage** User Story:

*Table 14. Ingest data on the Central Storage.*

<b>Story name:</b> Ingest data on the central storage	<b>Related WP:</b> WP2, WP7
<b>Triggered by User Role (Fictitious Name):</b> Local Data Manager (Carol)	<b>Related User Roles (Fictitious Name):</b> Data Holder (Raul) / Clinical Staff (Anna)
<p><b>Overview Description:</b></p> <p>Member of a research institute, Carol wants to populate the central storage of EUCAIM with an approved research project. The central storage hosts data (both medical imaging and associated clinical data) coming from the partners. To do so, data must be curated and fully anonymised before being uploaded (manually or in-bulk) into the repository.</p> <p>Before Carol starts uploading the data into the central storage, her colleague Raul, as a Data Holder, has received permission from Anna (Clinical Staff role) to extract all information from the different clinical databases involved in the research project and perform the actions of curation and anonymisation. Once everything is prepared and they have the approval of the EUCAIM Governing Body, Carol starts uploading studies (both images and clinical data). Every study uploaded is signed by Carol.</p> <p>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 central storage through the Data Access/Ingestion Service. The second one, she can directly access the Case Explorer Application provided by the repository for manual ingestion (case by case).</p>	

Example of **Monitoring the access of data collection** User Story:

*Table 15. Monitoring the access of data collection.*

<b>Story name:</b> Monitoring the access of data collection	<b>Related WP:</b> WP4
<b>Triggered by User Role (Fictitious Name):</b> Local Data Manager (Carol)	<b>Related User Roles (Fictitious Name):</b>
<p><b>Overview Description:</b></p> <p>A member of the clinical partner's staff, Carol, has successfully added data to the central storage system of EUCAIM. Now, she is interested in identifying the individuals or parties that have been accessing her uploaded data.</p> <p>To achieve this, Carol accesses the EUCAIM's application dashboard by logging in with her user id through AAI.</p>	



Once logged in, Carol navigates to the Collection Explorer 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.

Example of **Application of a tool to the federation marketplace** User Story:

*Table 16. Application of a tool to the federation marketplace.*

<b>Story name:</b> Application of a tool to the federation marketplace	<b>Related WP:</b> WP4, WP6
<b>Triggered by User Role (Fictitious Name):</b> Tool Provider (Paolo)	<b>Related User Roles (Fictitious Name):</b> Platform Manager (Sergio)
<b>Overview Description:</b> Paolo, an application developer, aims to contribute to the EUCAIM's marketplace by applying a specialised tool. His goal is to share a tool that enhances image processing capabilities within the federated ecosystem. Once Paolo has diligently developed the tool, he submits a registration request to the Governing Body. After reviewing and accepting the tool, ensuring its compliance and functionality, the Governing Body entrusts Sergio, the EUCAIM Platform Manager, with the task of uploading it to the Marketplace. By registering the application in the federation marketplace, Paolo aims to contribute to the collaborative and innovative environment of federated learning, fostering further advancements in the field. Paolo's tool becomes a valuable asset within the federation marketplace, enabling researchers and data scientists to leverage its capabilities in their own federated learning projects.	

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

*Table 17. Monitoring the access of tools.*

<b>Story name:</b> Monitoring the access of tools	<b>Related WP:</b> WP4
<b>Triggered by User Role (Fictitious Name):</b> Tool Provider (Paolo)	<b>Related User Roles (Fictitious Name):</b>
<b>Overview Description:</b> Paolo, a tool developer, has previously uploaded a tool onto the EUCAIM's marketplace and is now seeking to ensure that the platform's users are using it without encountering any issues. To accomplish this, Paolo gains access to the EUCAIM's application dashboard by logging in using the AAI. Once successfully logged in, Paolo proceeds to navigate through the Collection Explorer and locate his uploaded tool in order to access detailed information about it. Within the expanded menu of the tool, he can assess how it is being used by the platform's users.	

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

*Table 18. Exploration of collections from the Public Catalogue.*

<b>Story name:</b> Exploration of collections from the federated providers	<b>Related WP:</b> WP5
<b>Triggered by User Role (Fictitious Name):</b> Data User-Researcher (Alice)	<b>Related User Roles (Fictitious Name):</b>
<p><b>Overview Description:</b></p> <p>From a research group, Alice is a researcher willing to prepare a research project based on medical images of prostate cancer. As an anonymous user of the EUCAIM platform, Alice wants to conduct exploration of diverse metadata collections provided by federated Data Holders. Her goal is to perform an initial evaluation of the potentially available datasets and assess their relevance and suitability for her research project.</p> <p>Alice explores the Public Catalogue within the EUCAIM platform and navigates through the public partial view of the metadata collections, which are a combination of generic metadata and metadata defined in the Hyperontology (as a static world). She uses basic search and filtering options to identify collections that could align with her research interests and objectives. For example, Alice wants to know which collections have breast cancer cases and the cohort's age is between 40 and 50. As a result, she finds that INCISIVE and CHAIMELEON collections satisfy her criteria. Additionally, Alice can view the access conditions for each data collection offered by the Data Holders.</p> <p>In order to view the complete metadata collections available in the user's catalogue, Alice will need to register into the platform via the Life Science AAI and accept the Terms of Usage and Privacy.</p>	

Example of **Federated search of aggregated data in the collections** User Story:

*Table 19. Federated search of aggregated data in the collections.*

<b>Story name:</b> Federated search of aggregated data in the collections	<b>Related WP:</b> WP4, WP5
<b>Triggered by User Role (Fictitious Name):</b> Data User-Researcher (Alice)	<b>Related User Roles (Fictitious Name):</b>
<p><b>Overview Description:</b></p> <p>After signing up in the EUCAIM platform, Alice wants to explore the complete available metadata catalogue and make an advanced search on the multiple sources executing a federated query.</p> <p>For this purpose, Alice logs into the EUCAIM Dashboard and uses the available filters (cancer type, imaging modality, annotation availability, etc.) to perform a federated query over the different federated nodes, including the central repository which acts as another node. This federated query is based on the hyperontology concepts and is more detailed than the previous</p>	

search she performed. Now, as a result, she can not only determine the collections but also the number of cases (as aggregated numeric results) that meet her criteria. For example, Alice searches for collections that have breast cancer with a specific TNM stage value and with the MR series from a specific manufacturer. As a result, she discovers that INCISIVE collection has 123 cases, and CHAIMELEON collection has 400 cases that satisfy these criteria. Alice downloads a report of the numeric aggregated results obtained about the collections, to be attached to the official request of the research proposal towards the EUCAIM access board, which will include the objectives and methodology of her study.

Example of **Request access to the Collections from the User’s Catalogue** User Story:

*Table 20. Request access to the Collections from the User’s Catalogue.*

<b>Story name:</b> Request access to the Collections from the User’s Catalogue	<b>Related WP:</b> WP3
<b>Triggered by User Role (Fictitious Name):</b> Data User-Researcher (Alice) / Data User-Researcher with a Data Scientist profile (Leo)	<b>Related User Roles (Fictitious Name):</b> Access Committee
<b>Overview Description:</b> <p>In the EUCAIM platform, roles like Alice and Leo often need access to federated data for their research projects. They are both Data Users-Researchers, but Leo with a Data Scientist profile, who have identified the specific federated datasets they require for their research or analysis. As described above, they have explored the user’s catalogue and used a federated query to search for potentially available federated datasets meeting their specific criteria. Once they have identified the relevant datasets, Alice or Leo initiates the process by submitting an access request to the Access Committee. The Access Committee will review the access request and evaluate its alignment with the project’s purpose, data privacy regulations, and any other applicable policies to provide a final decision on the approval/rejection of the data access request. In some cases, depending on the agreement signed between the Data Holder and EUCAIM, the request will also be sent to the corresponding Data Holder, who will be able to provide this final decision, accepting or rejecting the request. Upon approval of the access request, Alice and Leo are granted access to the federated data. By getting access to the federated data, Alice and Leo can conduct their research or analysis using a rich and diverse dataset, expanding the scope and accuracy of their findings.</p>	

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

*Table 21. Get an overview of the datasets to which they have been granted access.*

<b>Story name:</b> Get an overview of the datasets which they have access to	<b>Related WP:</b> WP4
<b>Triggered by User Role (Fictitious</b>	<b>Related User Roles (Fictitious Name):</b>

<p><b>Name):</b> Data User-Researcher (Alice) / Data User-Researcher with a Data Scientist profile (Leo)</p>	
<p><b>Overview Description:</b></p> <p>Following the successful evaluation and acceptance of their access request applications in the EUCAIM platform, Leo (Data Scientist) and Alice (Researcher) now seek to obtain an overview of the datasets to which they have been granted access. This allows them to explore their User library with the collections to which they have access for their research projects.</p> <p>Having received confirmation from the Access Committee regarding their approved access, Leo, and Alice log into the User view of the Catalogue and within the service, they navigate to the section dedicated to accessing their approved datasets (User library).</p> <p>In this section, Leo and Alice can review the metadata and details of each dataset they have been granted access to. This overview enables them to fully assess the suitability of each dataset for their specific research objectives.</p> <p>Leo and Alice can also access information on any restrictions or conditions applicable to the datasets, ensuring compliance with privacy regulations, data usage agreements, or licensing requirements.</p> <p>By obtaining a comprehensive overview of the datasets they have access to, Leo and Alice can make informed decisions regarding the selection and use of the available data resources. They can select for use datasets that align closely with their research goals, ensuring the relevance and quality of their analyses and findings.</p>	

Example of **Select the data to be used from the accessible data collections** User Story:

*Table 22. Select the data to be used from the accessible data collections.*

<p><b>Story name:</b> Select the data to be used from the accessible data collections</p>	<p><b>Related WP:</b> WP4</p>
<p><b>Triggered by User Role (Fictitious Name):</b> Data User-Researcher with a Data Scientist profile (Leo)</p>	<p><b>Related User Roles (Fictitious Name):</b></p>
<p><b>Overview Description:</b></p> <p>Leo, a Data Scientist, plays the role of a Data User-Researcher within the EUCAIM platform. Her objective is to explore the user's catalogue and choose specific datasets for image analysis and processing using the platform's available tools.</p> <p>As a Researcher, Leo has used advanced search and filtering functionalities, so she could efficiently identify datasets that align with her research objectives and requirements.</p> <p>Upon successful access approval and exploring the access-granted collections in the User's Area, Leo can select the ones she wants to use. With them, she will utilise the Processing Services to develop complex models, conduct in-depth data analysis, and derive valuable insights to advance her research projects.</p>	

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

Table 23. Exploring the tools available in the federation and their provenance.

<b>Story name:</b> Exploring the tools available in the federation and their provenance	<b>Related WP:</b> WP6
<b>Triggered by User Role (Fictitious Name):</b> Data User- Researcher (Alice)	<b>Related User Roles (Fictitious Name):</b>
<b>Overview Description:</b> <p>As a Data User-Researcher of the EUCAIM’s platform, Alice has the opportunity to explore the diverse range of federated processing tools available within the federated environment. Her goal is to identify suitable tools for her specific research needs while understanding their provenance.</p> <p>Alice navigates the federated Tools Repository, a comprehensive catalogue of tools contributed by federated providers. These tools 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.</p> <p>Additionally, the platform provides traceability and provenance information of the tools. This valuable insight allows Alice to understand the datasets used for training the tools, the performance metrics, and the overall popularity and reputation of each tool within the federated community.</p> <p>Throughout her exploration, Alice can leverage the Marketplace interface to browse, search, and filter the available tools based on various criteria, such as tool type or functionality. The platform ensures a seamless and efficient experience, facilitating Alice’s exploration of the tools and enabling her to discover innovative solutions for her research projects.</p>	

Example of **Configure tools settings and computation and storage requirements for data processing** User Story:

Table 24. Configure tools settings and computation and storage requirements for data processing.

<b>Story name:</b> Configure tool settings and computation and storage requirements for data processing	<b>Related WP:</b> WP6
<b>Triggered by User Role (Fictitious Name):</b> Data User-Researcher with a Data Scientist profile (Leo)	<b>Related User Roles (Fictitious Name):</b>
<b>Overview Description:</b> <p>Leo, a Data User-Researcher with a Data Scientist profile, wants to be able to configure the settings of the tools used for data processing and define the computation and storage</p>	

requirements, so that she can optimise the data analysis process according to her specific needs.

In order to do so, Leo has access to an accessible and user-friendly configuration interface or settings panel where she can define the parameters and preferences for the data processing tools. The configuration interface allows her to specify the computational resources required for the data processing tasks, such as the number of CPU cores, memory allocation, and GPU usage. She can adjust the storage settings, including the destination folder for processed data, file formats, compression options, and retention policies.

Leo would also like the interface to provide options to select and customise the algorithms, libraries, and frameworks used for data processing, allowing her to leverage the most suitable tools for her analysis. This way, Leo can set up any necessary data pre-processing steps or transformations, such as data cleaning, feature scaling, or data normalisation, within the configuration interface. The system is able to support a wide range of data processing tools and frameworks commonly used in data science, providing flexibility and compatibility with Leo's preferred technologies. The system also offers the ability to save and manage multiple configurations, enabling Leo to switch between different settings profiles based on different analysis scenarios or project requirements. Integration with cloud computing platforms or resource management systems may be considered to facilitate the allocation and scaling of computational resources based on the defined requirements.

During this process, Leo 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. She would like the system to validate the configuration settings and provide relevant feedback or error messages if any conflicts or inconsistencies are detected.

Finally, the configuration interface should have appropriate security measures in place to ensure that only authorised users, such as Leo, can modify the settings and access sensitive information.

Example of **Process (distributed) the data from the federation by using a tool of the catalogue** User Story:

Table 25. *Process (distributed) the data from the federation by using a tool of the catalogue.*

<p><b>Story name:</b> Process (distributed) the data from the federation by using a tool of the catalogue</p>	<p><b>Related WP:</b> WP6</p>
<p><b>Triggered by User Role (Fictitious Name):</b> Data User-Researcher (Alice)</p>	<p><b>Related User Roles (Fictitious Name):</b></p>

**Overview Description:**

After exploring the catalogue of available tools within the EUCAIM federated environment, Alice, a Data User-Researcher, has selected the appropriate tools for her prostate cancer research project. Her main objective is to analyse baseline MR at the time of diagnosis to allocate treatment accurately.

Once Alice has chosen the available tools that meet her project's specific requirements, she leverages the distributed processing capabilities of the EUCAIM platform to process data from the federation. By utilising the power of distributed computing, she can analyse a large volume of MR data in parallel, speeding up the research process and enabling timely treatment allocation decisions.

Additionally, Alice can rely on the support services offered by EUCAIM, such as technical assistance, for any help she may require during tool selection and data processing. The user-friendly interface and documentation resources of the platform enable Alice to effectively utilise the selected tools, even without extensive experience in AI development.

By integrating the chosen tools 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 tools, contributes to advancing the field of precision medicine and improving outcomes for patients in the context of prostate cancer management.

Example of **Run a Federated AI model training** User Story:

*Table 26. Run a Federated AI model training.*

<b>Story name:</b> Run a Federated AI model training	<b>Related WP:</b> WP6
<b>Triggered by User Role (Fictitious Name):</b> Data User-Researcher with a Data Scientist profile (Leo)	<b>Related User Roles (Fictitious Name):</b>
<p><b>Overview Description:</b></p> <p>A team of data scientists led by Leo aims to publish a research project focused on developing an innovative federated AI model using the EUCAIM platform, a powerful infrastructure supporting federated learning.</p> <p>For this purpose, Leo accesses the EUCAIM Dashboard and uses the user's catalogue to identify suitable datasets for their research. Leo explores the metadata collections through a federated query. Once she has identified the datasets to use, she requests access to the collections through the Access Negotiator, which manages the access requests.</p> <p>With the datasets and processing tools ready, they begin training the model using EUCAIM's distributed federated learning approach. Each federated node securely retains its local data, ensuring privacy while contributing to the model's intelligence. Leo stores aggregated models, intermediate results and evaluation metrics in EUCAIM's storage, promoting collaboration within the team.</p> <p>After successfully training and evaluating the model, Leo 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, Leo is accepted to provide their</p>	

developed tool to the platform's marketplace, adhering to recommended standards for interoperability and reusability.

Example of **Monitoring job status in the processing** User Story:

*Table 27. Monitoring job status in the processing.*

<b>Story name:</b> Monitoring job status in the processing	<b>Related WP:</b> WP6
<b>Triggered by User Role (Fictitious Name):</b> Data User-Researcher with a Data Scientist profile (Leo)	<b>Related User Roles (Fictitious Name):</b>
<p><b>Overview Description:</b></p> <p>Leo, a Data User-Researcher with a Data Scientist profile, wants to be able to monitor the status of her job in the processing pipeline, so that she can track the progress and ensure timely completion of her data analysis tasks. In order to do so, Leo would like to access a user-friendly dashboard that displays the status of her jobs in the processing pipeline with a clear and intuitive interface that enables easy navigation and understanding of the job status information. She would like the dashboard to be scalable and capable of handling a large number of concurrent jobs without compromising performance, providing real-time updates on the progress of each job, including information such as job name, start time, estimated completion time, and current status (e.g., running, queued, completed). She should be able to filter and search for specific jobs based on criteria like job name, start time, or status to easily locate and track the desired job. Leo would like to be able to visualise the overall progress of the pipeline, such as a progress bar or a percentage completion display. The system should have appropriate security measures in place to ensure that only authorised users, such as Leo, can access the job status information.</p> <p>In case of any errors or issues encountered during job processing, Leo would like the dashboard to provide appropriate error messages or alerts, indicating the nature of the problem. Leo would also like to receive notifications or alerts when a job completes or encounters an error, allowing her to take prompt action if required. In addition, she would appreciate being able to maintain a log of completed jobs, including their details and any relevant metrics or output generated, for future reference and analysis. Finally, Leo would like this monitoring interface to be integrated with other tools or systems used in the data analysis workflow, such as data storage platforms or visualisation tools, to enhance user experience and workflow efficiency.</p>	

Example of **Retrieve and review obtained results** User Story:

*Table 28. Retrieve and review obtained results.*

<b>Story name:</b> Retrieve and review obtained results	<b>Related WP:</b> WP6
<b>Triggered by User Role (Fictitious Name):</b>	<b>Related User Roles (Fictitious Name):</b>



Data User-Researcher with a Data Scientist profile (Leo)	
<p><b>Overview Description:</b></p> <p>As a data scientist, Leo 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 to the list of her finished processing jobs and accesses them (or one of them) to check the results. There, she explores the available options for retrieving the trained model's results. Leo utilises the tools provided by EUCAIM to obtain the results from the aggregated AI model. She analyses and evaluates some aggregated information displayed in charts, which have been retrieved from the federated nodes. The EUCAIM platform integrates data visualisation tools within the Dashboard, allowing her to effectively interpret the results. To fine-tune the obtained results, Leo applies filters to the charts, enabling her to focus on specific aspects of the model's performance. She iteratively reviews the charts, gaining insights into the model's accuracy, precision, recall, and other relevant metrics. In addition to reviewing the visualised results, Leo exports and downloads a comprehensive report generated by the EUCAIM platform as a CSV/XLSX file. This report includes detailed information about the trained model's performance, including statistical analysis, graphs, and any other relevant data. She plans to attach this report to the official research project report, providing a detailed evaluation of the model's outcomes.</p>	

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 data collections from Data Holders** User Story:

*Table 29. Evaluate and accept/reject data collections from Data Holders.*

<b>Story name:</b> Evaluate and accept/reject data collections from Data Holders	<b>Related WP:</b> WP2
<b>Triggered by Profile (Fictitious Name):</b> The Access Committee (AC)	<b>Related User Roles (Fictitious Name):</b> Data Holder (Raul)
<p><b>Overview Description:</b></p> <p>Raul sends an application for providing data collections to be incorporated within EUCAIM. The AC makes a decision of acceptance or rejection, supported by the internal governance bodies on ethics and legal compliance. Raul receives a reply as soon as possible .</p>	

Example of **Evaluate and accept/reject tools from Tool Providers** User Story:

*Table 30. Evaluate and accept/reject tools from Tool Providers.*

<b>Story name:</b> Evaluate and accept/reject tools from Tool Providers	<b>Related WP:</b> WP2, WP3
--	--------------------------------

<b>Triggered by Profile (Fictitious Name):</b> The Access Committee (AC)	<b>Related User Roles (Fictitious Name):</b> Tool provider (Paolo)
<b>Overview Description:</b> Paolo sends an application for providing tools to be incorporated within EUCAIM. The AC makes a decision of acceptance or rejection, supported by the internal governance bodies on ethics and legal compliance, as well as those on technical matters. Paolo receives a reply as soon as possible .	

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

*Table 31. Evaluate and accept/reject a data access request application.*

<b>Story name:</b> Evaluate and accept/reject a data access request application	<b>Related WP:</b> WP3
<b>Triggered by Profile (Fictitious Name):</b> The Access Committee (AC)	<b>Related User Roles (Fictitious Name):</b> Data Scientist (Leo) / Data User-Researcher (Alice)
<b>Overview Description:</b> The AC receives a request of access to specific datasets of the Atlas of Cancer Images. The request has been issued by Leo (Data Scientist) or Alice (Researcher). After using the Public Catalogue, they have identified the datasets they need to carry out their research Project. They have logged into the User's Area, checked the instructions related to the type of documents they need to submit for requesting access to datasets and the maximum time to AC response. They prepare the required documentation and submit it. The AC receives through the Negotiator service an email notification that a new request has been submitted and initiates the established procedure to perform an evaluation. 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.	

Example of **Uploading of a tool to the federation marketplace** User Story:

*Table 32. Uploading of a tool to the federation marketplace.*

<b>Story name:</b> Uploading of a tool to the federation marketplace	<b>Related WP:</b> WP4
<b>Triggered by Profile (Fictitious Name):</b> Platform Manager (Sergio)	<b>Related User Roles (Fictitious Name):</b> Tool provider (Paolo)
<b>Overview Description:</b> As a member of the EUCAIM Platform Manager team, Sergio is responsible for uploading tools provided by the tool providers to the federation marketplace.	

Recently, Sergio 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, Sergio gets access to the tool from Paolo and can build the containerized tool. Sergio 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. Sergio registers this information through a user interface on the platform.

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

*Table 33. Platform services administration.*

<b>Story name:</b> Platform services administration	<b>Related WP:</b> WP4
<b>Triggered by Profile (Fictitious Name):</b> Platform Manager (Sergio)	<b>Related User Roles (Fictitious Name):</b> Data Scientist (Leo) / Tool Provider (Paolo)
<p><b>Overview Description:</b></p> <p>Sergio, as part of the EUCAIM Platform Manager team, is responsible for managing platform services in an efficient and secure manner. He receives requests from Paolo, an Application Developer, or Leo, a Data Scientist, to provide them with execution environments that meet their specific tool requirements for research purposes. With administrative credentials to the platform, including access to the Dashboard, Sergio ensures seamless administration of platform services.</p> <p>Sergio also proactively identifies the need for updates to existing tools, either due to security concerns or at the request of Alice or Leo. Recognizing the importance of maintaining a secure environment, Sergio employs DevOps tools 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.</p> <p>As a EUCAIM Platform Manager, Sergio plays a crucial role in ensuring the smooth functioning, security, and accessibility of services within the EUCAIM ecosystem. By leveraging automation and proactive updates, Sergio contributes to the overall efficiency and user satisfaction, ultimately supporting the research efforts of both Alice and Leo, as well as Paolo and other platform users.</p>	

## 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 1 shows the User Actions that can be carried out by all user roles (any User).

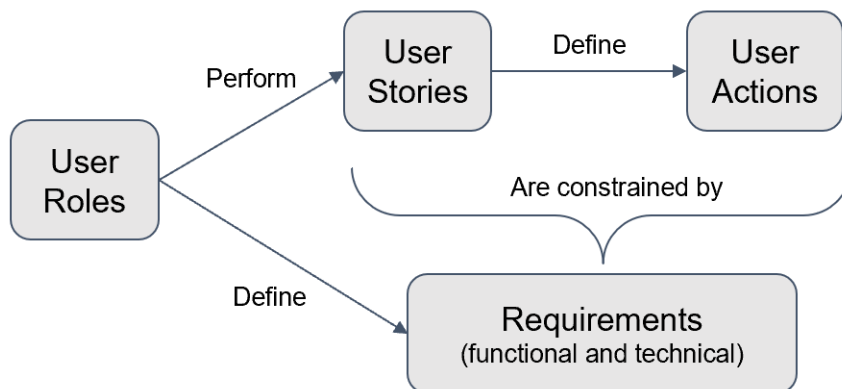


Figure 1. User workflow.

## User Actions enabled for all Roles

Table 1. User Actions enabled to all Roles.

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

## User Actions enabled for Data Holders

Table 2. User Actions enabled to Data Holders.

User Action	Description	User Story
Join the federation	Submit an application to join the federation	usDH1, usDH2, usDH3
Request collection registration	Submit an application to be authorised by the Governing Body	usDH2, usDH3, usGB1

User Action	Description	User Story
	(Access Committee) to register collections in the Federated Catalogue	
Prepare (meta)data	Prepare the (meta)data of the collection following EUCAIM's specification	usDH3, usDH4
Create a collection	Upload their already prepared data to the Central Storage and create a collection	usDH5
Upload metadata	Make their data discoverable, by pushing the metadata of the collections into the Public Catalogue	usDH6
Trace a collection	See information about the usage of their collections	usDH7

### User Actions enabled for Tool Providers

*Table 3. User Actions enabled to Tool Providers.*

User Action	Description	User Story
Request tool registration	Request approval of the Governing Body to upload the tool to the platform, accepting the terms and conditions	usTP1
Upload the tool to the platform	Upload the developed tool following the EUCAIM guidelines and recommendations to the platform, along with the related usage documentation	usTP1
Track the usage of the tool	Review the usage of the tool and some metrics about it	usTP2

## User Actions enabled for Data Users-Researchers

Table 4. User Actions enabled to Data Users-Researchers.

User Action	Description	User Story
Browse and filter collections	Browse collections in the public catalogue from different repositories and filter the collections of interest	usDU1
Federated query to filter collections	Search for collections 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
Add collection references to the data user-researcher library	Add the references of collections selected from the public catalogue (either filtered using the federated query mechanism or not) to the user library	usDU1, usDU2, usDU3
Request access to collections	Requests access to the collections by providing a project description, the institutional endorsement of the project, external ethical review (where applicable) and a specification of resources requested	usDU3
Follow-up on the status of the application	Retrieve information about the status of the request through the request tracking mechanisms integrated in the Dashboard.	usDU3
Browse the collections to which the user has access	List and filter the collections to which the data user-researcher has been granted permissions	usDU4, usDU5
Browse tools in the marketplace	Inspect, filter and select the available tools to be executed by the federated processing service	usDU6
Federated processing	Once the proposal is accepted, the researcher gets access to the	usDU5, usDU8, usDU9

User Action	Description	User Story
	processing service which will orchestrate the execution of the chosen tools over the authorised and selected datasets (either in the central node or in a federated node)	
Define the job parameters	Configure the computation and storage needs and the settings of the tools that are going to be executed	usDU8, usDU9
Monitor job status	Monitor the progress of a federated processing job and eventually cancel it	usDU10
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	usDU11

## User Actions enabled for the Governing Body

Table 5. User Actions enabled to the Governing Body - Access Committee.

User Action	Description	User Story
Ethical and legal review of applications	Review that data access requests meet the appropriate ethical and legal conditions	usGB3
Decide on the concession of permissions to access the data	Depending on the type of agreement with the Data Holder and the location of the data (federated node or central repository) the Governing Body will approve or reject the data access request	usGB3
Forward the request to the Data Holder Access Committee	If the data collections are located at federated nodes (not the central repository), the Governing Body, using the Negotiator tool, will forward the request to the Access Committee of the correspondent	usGB3

User Action	Description	User Story
	Data Holder	
Interact with the applicant	The applicants are informed about the status of the request through the request tracking mechanisms integrated in the Dashboard, which could be used by the Governing Body to provide feedback about the application	usGB1, usGB2, usGB3

## User Actions enabled for the EUCAIM Platform Manager

*Table 6. User Actions enabled to the EUCAIM Platform Manager.*

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

## 5. Operational procedures

### Data provision

#### Application process:

A Data Holder submits an application to join the federation using a standardised form.

The proposal will be evaluated by the corresponding Governing Body (AC). The procedures that need to be followed by the Data Holder will be negotiated with the corresponding technical, legal, ethical and scientific boards. A contractual document will be signed before the actual integration of the node.

EUCAIM works with two main categories of Data Holders:

#### 1. Data Holder in the Federated node of the infrastructure:

##### Description



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

## **Main procedures**

### **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.

### **2. Legal engagement:**

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).
- Pan-European entity (aggregating several federated nodes in different Member States).

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

3. Data de-identification will be done in accordance with the local regulation and the internal procedures at each node, under their direct management.
4. 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 Central Storage:**

### **Description**

A Data Holder in the Central Storage 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 may join EUCAIM by transferring data to the Central Storage of the infrastructure. This requires a series of steps regarding legal matters, such as signing a data sharing agreement and data de-identification, as well as on data quality assurance and compliance monitoring, as explained below.

## **Main procedures**

1. Data sharing agreement (delivered in WP3) will be signed between the Data Holder and EUCAIM.
2. Data anonymization. If needed by the Data Holder, EUCAIM may facilitate access to services/tools which streamline the anonymisation and batch uploading to the EUCAIM Central Hub.
3. Data quality assurance. Data quality verification will be performed (e.g., image, minimum set of clinical variables per cancer type). Quality standards for each type of image, and minimum clinical variables for each type of cancer may be defined.
4. Compliance monitoring guidelines will be distributed amongst Data Holders in order to ensure that EUCAIM procedures are followed.

Clinical sites participating in EUCAIM and the AI4HI projects: PRIMAGE or PROCANCER-I will constitute the first pilot of Data Holders to the Central Storage. These image repositories will serve to refine a set of operational procedures which will then be validated in additional use cases as part of the Open Call.

## **Tools / service provision**

### **Application process**

A tool provider may submit an application to incorporate a tool or service into the EUCAIM marketplace using a standardised form.

The proposal will be evaluated by the corresponding Governing Body (AC). The procedures that need to be followed by the tool provider will be negotiated with the corresponding technical, legal, ethical, and scientific boards.

### **Description**

A Tool Provider refers to the entity that would like to contribute with processing tools, services, or applications they have developed to the EUCAIM's marketplace for use in the federated processing purposes of the platform.

Development guidelines for tools and services 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 tools and services should be done through containerization. Containers provide a consistent and isolated environment for deploying and running applications. Developers should package their tools and services into container images to ensure portability and ease of deployment across different environments.

2. Container Image Requirements:

When creating container images, certain requirements should be met. These may include specifying the base image, ensuring the inclusion of all necessary

dependencies and libraries, 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 tools or services are deployed, they should be verified by a group of security experts of the Consortium to assess any potential security risks or concerns. They will evaluate the container image, review the source code for vulnerabilities, and ensure compliance with established security standards. This verification process ensures that the tools and services meet the required security criteria.

4. Usage Instructions and examples:

To facilitate the adoption and usage of the tools and services provided, clear instructions and examples should be provided. These guidelines should 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 tools and services, a specific usage licence should be defined. This licence outlines the terms and conditions for the use, distribution, and modification of the tools or services. It ensures that the legal rights and restrictions associated with the tools or services are clearly communicated and respected by users.

6. Integration Tests:

To ensure the compatibility and seamless integration of the tools or services with the existing infrastructure, comprehensive integration tests should be performed. These tests validate the interoperability of the tools 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 tools 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 should be in place to review and update dependencies, ensuring the long-term stability and security of the provided tools or services.

By following these development guidelines, organisations can ensure that the provision of tools and services is standardised, secure, and well-documented, enabling smooth adoption, integration, and maintenance within the EUCAIM marketplace.

## Data Access Requests

### Application process

A Data User-Researcher may submit a request for access to data from one or more collections that he/she has selected using the Public Catalog, to be used only within a research project. To do so, he/she must provide the required documentation (e.g. the study protocol) and the explicit approval of an Ethics Committee in the country where the research project is going to be conducted, when applicable. The request process will be orchestrated through the Negotiator tool which will allow the interaction of the Access Committee and the requester. The request will be evaluated by the Access Committee, and, depending on the agreements signed with the Provider that owns the data, it may also be evaluated by the provider's Access Committee (or an equivalent entity). The approval of the request will imply the granting of authorization to use the datasets.

## **Description**

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

## **Main procedures**

### **1. Data Discovery as an anonymous user**

To initiate their research project, the researcher begins by engaging in the Data Discovery process. They explore the infrastructure's Public Catalog, which provides access to a partial view of a metadata catalogue containing basic information about the available data collections in the Atlas of Cancer Images. The researcher carefully evaluates the catalogue, searching for datasets that align with the objectives of their research.

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

Before proceeding further, the researcher 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 discovery as an authenticated user**

As a registered user, the researcher can now explore the user's view of the catalogue (or user's catalogue), which provides access to the complete view of the public metadata catalogue containing extensive information about the available data collections in the Atlas of Cancer Images. Additionally, as an authenticated user, the researcher can execute a federated query and make an advanced search on the federated catalogue.

### **4. Data Access request, for data in the Atlas of Cancer Images**

As a registered user, the researcher is now prepared to request data access from the Central Storage and/or federated nodes, depending on the specific requirements of their research. There are three distinct processes that the researcher can follow based on their needs:

- If the desired data collections are available in the Central Storage, the researcher can identify them through the User's catalogue and request access. This straightforward process involves submitting a data access request for the relevant data collections.
- If the desired datasets are part of the federation, the researcher can identify them in the User's catalogue and proceed with a similar access request. The researcher understands the significance of leveraging data from diverse sources within the federation, enabling them to draw comprehensive conclusions.
- Lastly, in situations where the researcher's research project necessitates other data collections that are not included in the search results returned to them, they can send a request to the federation for new data collections if available at affiliated hospitals. This process involves signing special agreements with the respective clinical Data Holders, ensuring the compliance of the newly available data collections.

A local ethics committee's approval (either from the data user-researcher's institution or external if not available, but always from the country where the research project is going to be conducted) is needed at this stage to request data access for the preparation of a research project, where this is applicable.

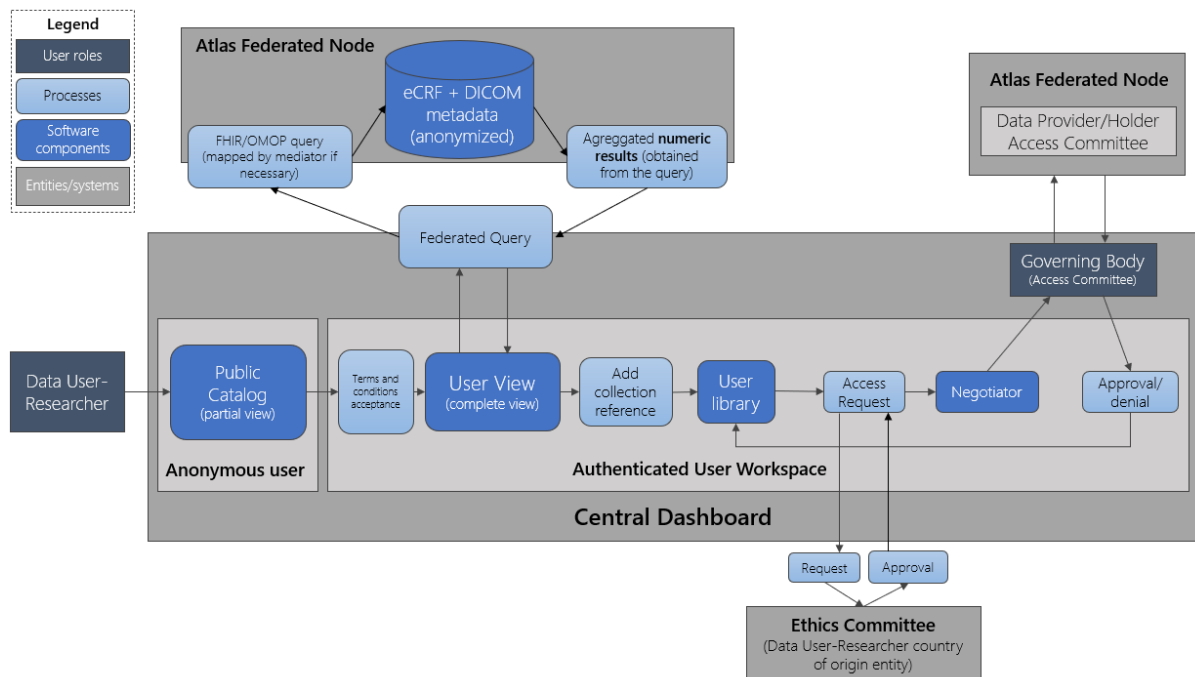


Figure 2. Data Access Request workflow.

## 5. Access to data collections and tools in the Atlas of Cancer Images

### 4.a Access to datasets of the Central Storage:

Once the researcher gains access to the datasets located in the Central Storage, they may have two options for data utilisation. Based on the specific agreements with the Data Holders, they could either download the datasets for local analysis or employ the cloud platform provided by the infrastructure to perform their analysis within the platform itself. The researcher may weigh the

advantages of both approaches and data available and select the one that best suits their research requirements.

#### 4.b Access to federated data

Once the researcher gains access to federated data within the research infrastructure, they are presented with a range of options for data utilisation. Federated data refers to data that is distributed across multiple nodes within the infrastructure, allowing researchers 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.

#### 4.c Access to datasets of the Central Storage and to federated data

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

#### 4.d Access to tools, managed by the Central Hub

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

#### 5. Invoicing and payment, when applicable

In certain cases, access to specific datasets or processing services may come with a fee. The researcher 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.

### Other considerations

#### **Important dates for the involvement of Data Holders and users (Figure 3)**

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

- The Open Calls in 2025: the Data Holders and users winners of the Open Call have access to funding and this implies the need for them to become project partners, under the Digital Europe programme rules. Therefore, they 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.
- 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.

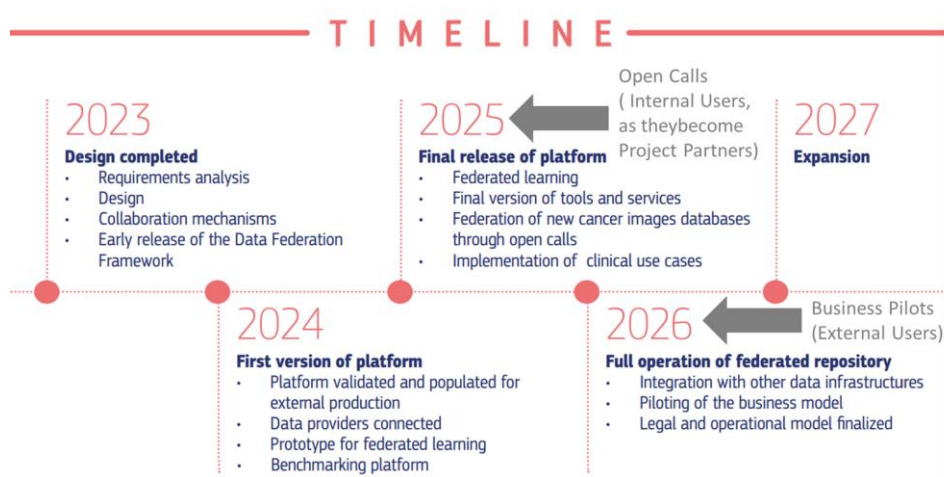


Figure 3. The EUCAIM platform implementation timeline.

The business model pilots will imply the need to assess the incorporation of new Providers (both Data Holders and Tools Providers), as well as new Data Users-Researchers. The managerial structure for the assessment of candidates is depicted in Figure 4 and Figure 5 for Data / Tools providers and Data users-Researchers, respectively.

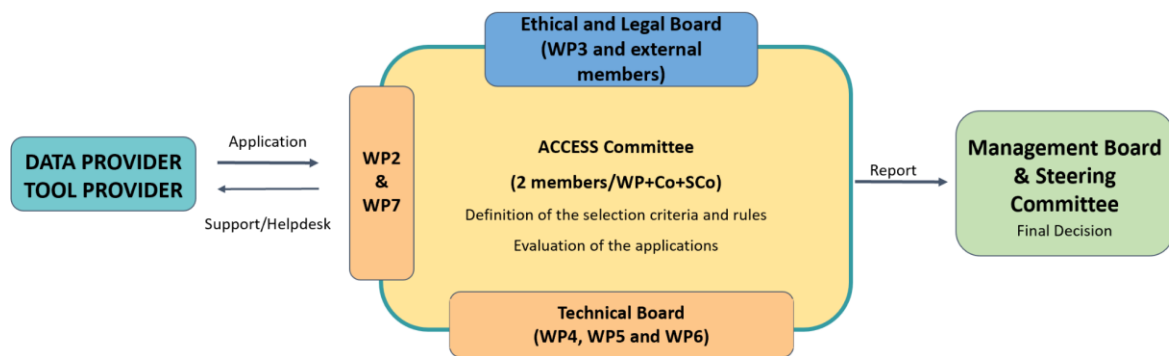


Figure 4. Data / Tools Providers flow.

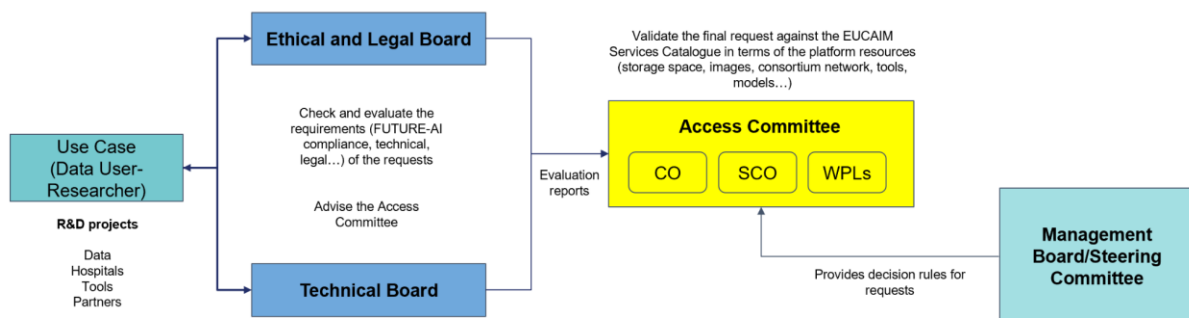


Figure 5. Data User-Researcher flow.

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 a first proposal for the minimum management structure needed to operate the Central Hub, which is currently envisioned to be an evolution of the management structure of the EUCAIM Project. 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.

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

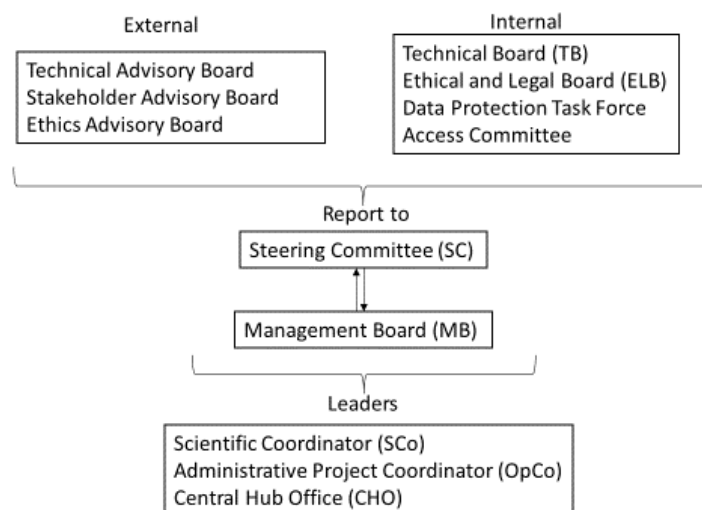


Figure 6. Operative boards managing the Central Hub.

### 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” in MS1: Project Management Handbook), being chaired by the Scientific Coordinator. 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 Steering Committee 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 should 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
- Ethics Advisory Board - more particularly dedicated to advising the consortium on ethical and related legal issues, as well as on exploitation and regulatory matters.

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.

The stakeholder AB is currently being defined as different stakeholders are expressing their interest and contacting MB members to be included.

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).

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 experts in cloud infrastructure maintenance, technical support, legal matters, fundraising and project management, IPR, dissemination and promotional actions, as well as administrative and financial management. 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.

#### **7. Technical Board (TB)**

The Technical Board (TB) is currently tasked with the review of the potential engagement of tools 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: Luisa Cabornero Marco (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 (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 tools 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 tools 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 recently created Access Committee is to ensure responsible and secure access to the infrastructure data and services, promoting valuable research while upholding ethical and privacy standards. To do so, the AC is being designed as a dedicated body responsible for overseeing and managing the access requests to the infrastructure's data issued by authenticated users. Its primary role will be to review and approve data access requests, ensuring compliance with legal, ethical, and privacy regulations (following guidelines from the TB and ELB). The committee will initially define the set of documents that the Data Users-Researchers need to submit, the submission channels and the timeline of the evaluation process. With these, it will assess the scientific validity and feasibility of the research proposal and evaluate the data protection measures implemented by the requester. The AC will also be in charge

of evaluating and accepting/rejecting Data collections from Data Holders, as well as tools from Tool Providers.

The current expectation is that this committee will be appointed by the Management Board, although the operative rules for this body are currently under definition.

Being EUCAIM 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 AC 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. Rules for Participation

This section includes a first approach to technical implementation needs. More detailed information on the Rules of Participation can be found in deliverable D4.3 Rules for Participation.

### Minimum requirements in terms of data and infrastructure

The minimum requirements for both the case of the Central Storage and of a Federated node are defined in the D2.1. Onboarding invitation package of the project:

#### 1) For the case of the Central Storage:

- Access to a dedicated machine for uploading and administering data in the central node.
- Network configuration: allow outgoing network connection over HTTPS in order to connect to the central services of the Federated Learning platform.
- Technical support: Staff availability for technical support is required for installation of software and hardware.
- Compliance with the technical guidelines through the deployment and execution of various EUCAIM tools (e.g. tools for data anonymization, image annotation tools, data quality/cleaning tools) and transformation of data to the common data model of the EUCAIM infrastructure (e.g., through the Extract Transform Load (ETL) tools).

#### **Functional Requirements:**

- Minimally-annotated and anonymised data availability.
- Data sharing mechanism in place.
- Data mapped to the hyperontology and transformed to the common data model.

## 2) For the case of a Federated node:

- Procurement or ownership of the needed infrastructure.
  - Each organisation is expected to obtain the technical and management infrastructure needed to host a federated node.
- Acquire and install dedicated hardware.
  - If the local nodes will support federated processing, at least a large server equipped with graphics processing units (GPUs) allowing for intensive learning processes. Alternatively, data sites may set their own agreements with third-party trusted institutions (e.g., regional or national computational facilities, commercial providers) to provide the indicated computational infrastructure on their behalf. In those cases, Data Holders will be responsible to set the necessary legal and technical agreements with such institutions.
  - Data redundancy measures (e.g. redundant array of independent disks; RAID) will have to be implemented to mitigate the risks associated with data loss, deletion or corruption, both during and after the project's lifespan. This includes the establishment of regular data backups, replication processes, and robust failover systems to guarantee the continuity and reliability of data storage and retrieval.
- Install operating system and deploy image pre-processing and EUCAIM repository connectivity software provided in the form of software containers.
- Local technical support.
  - Staff availability for technical support is required for installation of software and hardware.
  - Compliance with the technical guidelines through the deployment and execution of various EUCAIM tools (e.g. tools for data anonymization, image annotation tools, data quality/cleaning tools, clients for federated analysis) and transformation of local data to the common data model of the EUCAIM infrastructure (e.g. Extract Transform Load (ETL) tools).
  - Application programming interfaces (APIs) for data sharing should be available according to EUCAIM specifications.
- Network configuration.
  - Site should allow outgoing connections to selected ports. The possibility of opening inbound ports is desirable but optional.

### **Non-Functional Requirements:**

- Security measures should be enforced so that access is regulated. All communications among services will be performed using HTTPS and REST APIs. Authentication and Authorization will use OpenID and SAML.
- Traceability and auditing mechanisms should be in place.
- Integration tests should be performed.
- Minimum uptime should be ensured for Data Holders.
- Provide the Service Level Agreement (SLA) agreed in terms of availability, scalability, and maximum number of requests per second for the federated data search and processing.

## Expected response times

The Access Committee operates for both the Central Hub and the Federated Nodes and will make a decision of acceptance or rejection of Data Holders, tools providers, or data access requests by researchers.

The federated nodes will have a period to decide whether they accept or not the use of their data regarding a specific data access request. Hospitals will always be asked for this decision while other Data Holders may or not be asked depending on their agreement with EUCAIM.

The evaluation of applications and the expected response times (in terms of provision of data and tools, as well as data access requests) will undergo a structured process to ensure that participation to the EUCAIM infrastructure is granted based on rigorous scientific, technical, legal and ethical criteria. This process is an ongoing work by the time of this current deliverable and will be set up as per T7.1: Data incorporation use cases (Open Call). The EUCAIM Access Committee and the EUCAIM Management Board will oversee the entire process to maintain transparency and fairness. Comprehensive details of the evaluation procedure, including relevant timelines, will be made available through D7.1: Rules for the Evaluation and Prioritization of Use Case Applications from the Open Call (related to T7.1a), scheduled for release in December 2023 (M12).

## Data Quality

The D2.1. Onboarding invitation package of the project states that: “Data quality and cleaning tools will be applied to the data for addressing quality issues such as correcting data that is incomplete, incorrect, inaccurate or biased, etc.”. Such tools will be documented in D5.1 Early release of the Data Federation Framework by month 9 of the project, and in D5.4 Data Pre-processing Tools and Services (outcome of T5.3) by month 18 of the project.

In order to prepare the datasets, Data Holders will apply some pre-processing tools for datasets to be compliant with the EUCAIM data collections requirements, with the main aspects to be processed/evaluated being FAIR compliance, anonymisation of the collections, data curation, quality assurance and data harmonisation.

## Compliance framework design

The operational design proposed by the Proposal for a Regulation of The European Parliament and of The Council on the European Health Data Space should inspire the Operational Platform's approach for EUCAIM. In any case, during the ongoing procedure of the new Regulation, the GDPR offers sufficient criteria to achieve an outcome that is valid for the present and future-oriented. This implies the design of compliance models at least in the areas described below. The relevant bodies expected to govern the decision making in each area are indicated in brackets.

- **Definition of the roles of the different subjects in the Platform (Technical Board):**
  - Data Holder and assumptions of co-responsibility.
  - Tool Provider.
  - Data User-Researcher.
  - Research Communities.
- **Definition of governance models (Management Board):**
  - Internal Organisational Governance.
  - Access Committee.

- Steering Committee or board of directors (when legal entity status will be acquired).
- Ethical and Legal Board (with independent experts external to EUCAIM).
- Bodies related to the social environment (patients' associations, academia, industry, interest groups).
- **Procedural (Access Committee, Management Board):**
  - EUCAIM accession processes.
  - Data sharing agreements.
  - Research consortia with EUCAIM.
  - Requests and authorisations for access to data.
  - Cooperation with authorities and bodies: data protection authorities, specialised supervisory authorities in DGA, and future AI Regulation supervisory bodies and EHDS-R Health Data Access Bodies and supervisory authorities.
- **Organisational (Data Protection Task Force):**
  - Chief (information) data officer: CDO-CIO.
  - Chief Security Officer.
  - Data Protection Officer.
  - AI ethics advisor
  - Highly qualified technical positions (AI developers, data analysts, etc.).
- **Compliance (Ethical and Legal Board, Data Protection Task Force):**
  - Regulatory compliance framework and compliance by design in AI.
  - Compliance audit.
  - Security audit.
  - Instrumental process of data protection risk analysis and auditing.
- **Design of legal documentation to support regulatory compliance (Ethical and Legal Board, Data Protection Task Force):**
  - Joint controllership agreement.
  - Data Processor Agreement.
  - Data Sharing/Data Transfer Agreement.
  - Terms and Conditions.

The deployment of all these objectives must guarantee the following values and principles that constitute a core and inalienable element of the European Union's Common Acquis:

- People's (patients') rights are at the core: Human centric approach, fundamental rights based development.
- Ensuring digital sovereignty.
- A high ethical standard incorporating the new ethics of AI into the ethics of biomedical research.
- A high degree of transparency on the availability and nature of the data sets, the conditions of registration, access and use of the platform, and the results achieved.
- A risk based approach ensuring that:
  - The risk to fundamental rights has been assessed.
  - Ethical risks have been taken into account in all their dimensions, including digital ones.
  - All security risks have been assessed.
  - Analysis of specific risks in the fields of anonymisation and pseudonymisation, and AI with particular attention to specific aspects such as explainability, bias or adversarial effects.

A model focused on traceability and the adoption of appropriate safeguards to ensure:

- The legitimate origin of the data.
- Legality in the transfer, disclosure, sharing and use of data.
- The definition of binding commitments for all participants in the system and, in particular:
  - On the guarantee of the quality, interoperability, accessibility and availability of the data. With special attention to the complexity of the level of service in cases of federated data processing (Data Holders).
  - The registration, traceability and definition of the conditions to be applied to users with a data access permission (data users).
- An operational compliance framework aligned at EU and national level, traceable and auditable.

## 8. 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 EUCAIM 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 indicates the terms that should be filled into bracket keys (“{ term }”) Optional terms (e.g. the provider may or may not dedicate processing services) are indicated with square brackets (“[ optional term ]”). Terms separated by a vertical pipe define different options from which only one should be included (e.g., “centralised | distributed”). Comments are enclosed between “<<” and “>>”.

It is important to state that some of the clauses may not apply to a specific provider (e.g. a pure Data Holder may include “0” processing resources or vice-versa for a compute provider site).

### Objective of the document

The {Federated Node Name | Repository provider} services are provided by {Entity Name} (“the Provider”). This document defines the Level Service Level Agreement (SLA), valid for the EUCAIM Federated infrastructure to support the access [and processing] of the data exposed by the Provider.

The Provider will provide access to the anonymised local Data Warehouse [and the processing resources associated] according to the conditions of the following clauses.

### Clause 1: Data provisioning through the local node.

The Provider will set up a local node that will comply with the technical interoperability principles described in the technical interoperability report. <<A reference to the document will be added here>>.



The local node will: <<Remove those that are not applicable>>

- Register collections in the EUCAIM Public catalogue following EUCAIM's collection metadata specification.
- Synchronise the local node catalogue with the EUCAIM Public catalogue supporting EUCAIM's collection metadata specification.
- Expose a federated search endpoint compatible with the EUCAIM Federated Query.
- Provide access to the data under the conditions described in clause 3.

The provider will commit to expose at least:

- {Num} cases, {Num} studies and {Num} images related to cancer management.
- Images should fulfil the Quality principles defined in <<a reference to the quality principles>>

## **Clause 2: Processing resources provided**

The Provider commits to the EUCAIM Federated infrastructure the following processing resources:

- {Num} GPUs of type {model}.
- {Num} nodes with a total of {Num} cores, {Num} GB of RAM.
- {Num} TB of Persistent Storage.

The resources will be shared among the users and provided "as available".

## **Clause 3: Access conditions**

The use of the resources should be proportional and reasonable and could be revoked by the provider in case of urgent needs or user abuse.

The access to the repository will include: <<Remove those that are not applicable>>:

- Access to a virtual environment to browse the data that the user is authorised to access.
- Access to processing facilities [ including GPUs], subject to the availability of the resources.
- Downloading of anonymised data.

## **Clause 4: Service Support**

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., 24 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 organisation providing the service.

The Quality of Support level is defined as follows:

- Less urgent (5 working days)
- Urgent (5 working days)
- Very Urgent (1 working day)
- Top Priority (1 working day)

### **Clause 5: Service Availability**

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.

### **Clause 6: Contact points**

Please direct any questions, complaints or claims related to the Services or your use of the Services by email to {<non\_personal\_alias\_email\_address>}. The official address of the provider is {Provider Address}.

## 9. Conclusions

This deliverable provides a first approach to the 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 preliminary operational procedures and rules for participation. The project consortium hopes this deliverable will set the first grounds on how to use the EUCAIM infrastructure securely and effectively. This document also provides an initial overview of the governance bodies responsible for the operation of the Central Hub, offering definitions about their roles, responsibilities, and decision-making processes. These definitions are expected to be subjected to change during the infrastructure's life span.

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 under construction and aims to provide vocabulary standards for use throughout the infrastructure's implementation and beyond.

Looking ahead, we anticipate that this initial version of the operational framework will undergo further updates by the end of the project's 18th month. During this period, the central hub will be operational for internal clinical validation (MS27). These updates will incorporate valuable insights gained from the validation process, thereby refining and enhancing the operational framework to ensure its effectiveness and efficiency.