Project title: European Federation for Cancer Images

Project acronym: EUCAIM

Grant Agreement: 101100633

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

D2.2 Training Plan

Partner: DKFZ

Authors: Heinz-Peter Schlemmer (DKFZ), Hanna Leisz (DKFZ), Laure Saint-Aubert (MEDEX), María Toboso Ortega (HULAFE), Luisa Cabonero Marco (HULAFE), Irene Marín (HULAFE), Philipp Seeböck (MedUniVienna, CIR), Ignacio Blanquer (UPV), Javier Blázquez Sánchez (SERMAS), Javier Soto (SERMAS), Ioanna Chouvarda (AUTH), Marco Aiello (SYNLAB), Mélanie Sambrès (LIMICS), Marta Martinez, Francesca Pia Caputo (UNIPI)

Reviewers: Janos Meszaros (KUL), Ignacio Blanquer (UPV), Sara Zullino (EATRIS), Amelia Suarez (MAT)

Date of delivery: 2023-12-22

Version: 1

Table of contents

Executive summary 3

1. Introduction and Scope of this Document, Target Groups and Relevant Terminology 3

2. Learning Objectives 5

3. Accessibility 7

4. Tier Structure and Demands for Training According to Target Group 7

4.1. Tier Structure 7

4.2. Data Holders, Tool Providers, Research Communities and Data User-Researcher's Demands for Training 10

5. Integration of Training Activities into the on-boarding processes for Data Holders, Tool Providers, Research Communities and Data User-Researchers 14

5.1. The On-boarding process for Data Holders, Tool Providers and Research Communities 14

5.1.1 The On-Boarding Process for Data Holders, Tool Providers and Research Communities as Stakeholders 15

5.1.2 On-boarding through the Internal and External Open Call 16

5.2. The On-boarding and Registration Process for Data User-Researchers 17

5.3 Timeline for Training Activities 18

6. Training Catalogue 19

6.1. Training for Data Holders, Tool Providers, Research Communities and Data User-Researchers 19

6.2. Individual Training Sessions 22

6.3. Legal and Ethical Training for Internal Consortium Members 22

6.4. Training on Updates 23

6.5. External Training Materials 23

7. Monitoring and Evaluation of Training 23

Executive summary

This document outlines the Training Plan for the EUropean Federation for CAncer IMages (EUCAIM) project. It provides an overview of relevant terminology, introduces the target groups for training and their demands for training according to their level of compliance in EUCAIM as well as the learning objectives. Based on this analysis, the integration of all training activities into the on-boarding process is described and the training catalogue is introduced.

1. Introduction and Scope of this Document, Target Groups and Relevant Terminology

The goal of EUropean Federation for CAncer IMages (EUCAIM) is to build a pan-European digital federated infrastructure of cancer-related radiological and nuclear medicine images and other related digital information, which will be used to develop Artificial Intelligence (AI) tools for Precision Medicine. EUCAIM aims at preserving the data sovereignty of providers and will address the fragmentation of existing cancer image repositories throughout Europe by building on repositories of the AI4HI initiative, European research infrastructures and national/regional repositories and thus include clinical images, pathology, molecular and laboratory data.

This document describes the Training Plan for data holders (DH) as well as tool providers (TP), research communities (RC) and data users-researchers (DU-R), and aims at facilitating the navigation through the EUCAIM platform for these target groups (for an illustration of the on-boarding process of theses target groups see 5.1). The Training Plan has been developed through the discussions of the participants in Work Package 2 and experts from the work packages related to the training activities, as well as a training survey that the Training Team conducted in May 2023 with all work packages.

The training plan mirrors the current status of the project towards the end of 2023 and may be subject to further changes as the project evolves and the training activities are monitored and evaluated. Consequently, we expect our training materials to mature over the course of the evolving project.

The Training Plan will make use of the terminology as agreed upon in the Glossary of EUCAIM[1]. Relevant terms and definitions for training are:

Common Data Model (CDM): A CDM is a standardised framework that defines both the structure and semantics of diverse datasets using ontologies, coding systems, and formal documentation. Clinical data standards provide a common structure and content of observational data, enabling interoperability and more efficient analyses that can produce reliable evidence. Within the EUCAIM project, two potential candidates for the CDM have been identified: HL7 FHIR and OHDSI-OMOP.

Data altruism: Voluntary sharing of data on the basis of the consent by data subjects to process personal data pertaining to them, or permissions of other data holders to allow the use of their non-personal data without seeking a reward, for purposes of general interest, such as scientific research purposes or improving public services.

[1] EUCAIM Glossary https://docs.google.com/document/d/16_QqvV-qwVpwdAiY4I2Z540DVKpoQ_EPInuj8iwQkpI/edit

Data Federation Framework (DFF): refers to a numerical high-level architecture through which users and providers may access various federated services.

Data Holders (DH): A Data Provider/Data Holder/Data Controller (also 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 for research purposes. This may include registering, providing, restricting access to, or exchanging the data. Examples of Data Holders include data repositories, regional biobanks, clinical centres, cancer screening programs, public entities, pharmaceutical companies, data altruism initiatives, and publication repositories. These infrastructures may host one or more datasets for discovery and retrieval, and the exposure and access to data in the Dashboard will be provided at the dataset level.

Data User-Researcher (DU-R): a EUCAIM end-user that is any person or entity that explores the public catalogue of available metadata, eventually requests access to data, and processes them using either the tools available in the platform or their own AI tools. An example of a DU-R can be a Principal Investigator conducting a research project on prostate cancer, with one of its objectives being the prediction of the best treatment allocation based on the analysis of baseline Magnetic Resonance Images (MRI) at the time of diagnosis.

FAIR Principles: Principles to define the Findability, Accessibility, Interoperability, and Reuse of resources for humans and computers at the source. For example, the principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.

- Findable: Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.
- Accessible: Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository.
- Interoperable: Metadata uses a formal, accessible, shared, and broadly applicable language for knowledge representation.
- Re-usable: Data and collections have a clear usage licence and provide accurate information on provenance.

Federated Catalogue: Metadata catalogue that stores the clinical and imaging metadata within the different federated nodes of the Atlas of Cancer Images, as a federated search endpoint compliant with the EUCAIM federated query requirements.

Federated Node/Local Node: Infrastructure deployed in a Data Provider that meets the hardware and software requirements of the EUCAIM project and that has been configured and connected to the federated network, being able to access Hyper-ontology compliant federation-exposed collections (under research project approval) and execute federated processing, including federated learning.

Hyper-ontology: The Hyper-ontology is a superset of the metadata of the catalogue and refers to the actual searchable data. Hyper-ontology refers to the creation of a comprehensive and unified semantic representation of fundamental knowledge defined in the local AI4HI projects. The main purpose of the Hyper-ontology is to facilitate the development of validated clinical decision-making systems that support diagnosis, treatment, and predictive medicine, to benefit citizens. The Hyper-ontology serves as a

standardised framework that enables interoperability between different projects using OMOP and FHIR standards. It allows for the expression of federated queries, enabling the analysis of distributed data sources. Additionally, a subset of the Hyper-ontology will be dedicated to describe datasets.

Public catalogue: Metadata catalogue available to anonymous users, offering a limited view of the catalogue and basic search options. This catalogue stores metadata, offering the Data User-Researchers basic descriptive information about the available datasets and data access conditions.

Research Communities (RC): 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.

Tool: A software application that assists, enhances or executes an action or process.

Tool Provider (TP): entity (e.g. startups, enterprises, research institutions, government agencies, non-profit organisations) that would like to make their already developed data processing tools, services, or applications available in EUCAIM's marketplace for EUCAIM users to utilise them for federated processing or data pre-processing purposes of the platform. An example of a tool provider would be a start-up company that is willing to provide an Artificial Intelligence (AI) explainability platform that helps explain, analyse, and monitor the behaviour of AI models in real-time.

Use case: A use case would refer to a description of a specific scenario or situation in which EUCAIM is intended to be used to address a particular problem related to cancer data for clinical purposes. It would outline the steps and interactions involved in the process, and the expected outcomes of applying EUCAIM in a real-world setting. Examples of use cases in this context could include the development of AI-based diagnosis tools or the use of federated data sources to improve patient outcomes. Use cases will be used to drive the design of the EUCAIM infrastructure and help define requirements, test functionality, and validation process for a clinical improvement.

2. Learning Objectives

The aim is to comprehensively inform about the EUCAIM platform concerning technical requirements as well as tools and rules for operation. The activities will focus on EUCAIM's different target groups, i.e. data holders (DH), tool providers (TP), research communities (RC) as well as data user-researchers (DU-R). The training activities will provide an overview of the platform's functionalities and operation guidelines which need to be understood and complied with.

The training catalogue consists of five thematic areas which are considered vital for all users to be able to adequately learn about the platform structure and its operation:

(1) Training to comply with legal and ethical aspects regarding data provision and data usage
   → Understand the legal and ethical requirements for data provision and data usage, terms and conditions
   → Understand relevant GDPR regulations
   → Understand safety and privacy in EUCAIM
   → Understand the authorisation process

(2) Training for technical requirements
   → Understand how to set up a local node (both hardware and software)
   → Understand how to connect to the central infrastructure

(3) Training for data and tool provision
   → Understand how to federate data
   → Understand the harmonisation and standardisation requirements in EUCAIM
   → Learn about the tools for data anonymization, segmentation, annotation and ETL in EUCAIM

(4) Training on FAIR data management
   → Understand how to make use of the platform in a FAIR compliant way
   → Understand how to access data and tools
   → Understand who can access data and tools

(5) Training of platform usage
   → Understand platform operation
   → Understand how to use the catalogue
   → Understand how to use the federated query
   → Learn about clinical and AI development use cases

Legal and ethical training is mandatory for DHs, TPs and RCs willing to join the federation. This recommendation is based on the results of the survey performed by WP2 together with all other work packages (see 6.). For the existing partners of the EUCAIM consortium, an internal training session on legal and ethical aspects (see 6.3) has already been conducted. This training is considered necessary to ensure that the data protection by design strategies in EUCAIM will be understood and complied with. All relevant legal aspects such as those in the context of the European Health Data Space (EHDS), security aspects in data privacy as well as the future regulations of the AI Act will be considered.

All other training activities are optional and aim at facilitating the connection to EUCAIM. They provide practical information on how to operate the platform operation and how to prepare FAIR compliant data and tools.

3. Accessibility

EUCAIM's training sessions and materials will be accessible through the training catalogue (see 6.). This catalogue will be provided to all target groups (DH, TP, RC, DU-R) after registering and providing the necessary documents as outlined in the on-boarding process (see 5.).

The training materials will provide an overview of the platform and its functionalities. The information is provided in addition to the comprehensive documentation provided by the consortium. Reference will be given to those documentation sources that provide in depth information on relevant topics. The training catalogue is structured according to topics and target groups. It allows for selecting individual training sessions or training materials used also for public dissemination through the website (e.g. training session on general platform use or data accessibility). This option will support EUCAIM in providing information to outside entities with an interest in EUCAIM:

- industry partners will be informed about the platform and its potential use for R&D;
- patients' organisations will be informed about the secondary use of health data for research;
- ethical committees, DPOs and other experts in legal and ethical use of AI in health applications will be informed about GDPR compliance and integration into the European Health Data Space (EHDS).

4. Tier Structure and Demands for Training According to Target Group

The following section illustrates EUCAIM's tier structure, i.e. the different levels of compliance for data in order to be made available through the platform. Varying demands for training are expected for the different tiers. A collection of questions and demands for training based on the tier structure has been developed and is featured in 4.2.

4.1. Tier Structure

The main objective of the on-boarding process for new DPs, TPs and RCs, is to facilitate the provision of data and tools. In this context, EUCAIM uses a three-tier structure representing increasing levels of compliance with the Data Federation Framework[2] (Figure 1). These tiers are scalable, allowing upgrading as the datasets and tools are used in new research projects. Within the EUCAIM network, achieving compliance with data standards and the DFF across the three technical tiers is an essential goal to ensure the seamless integration of diverse datasets.

[2] For a more detailed description of the tier structure see D4.3 First rules for participation report.

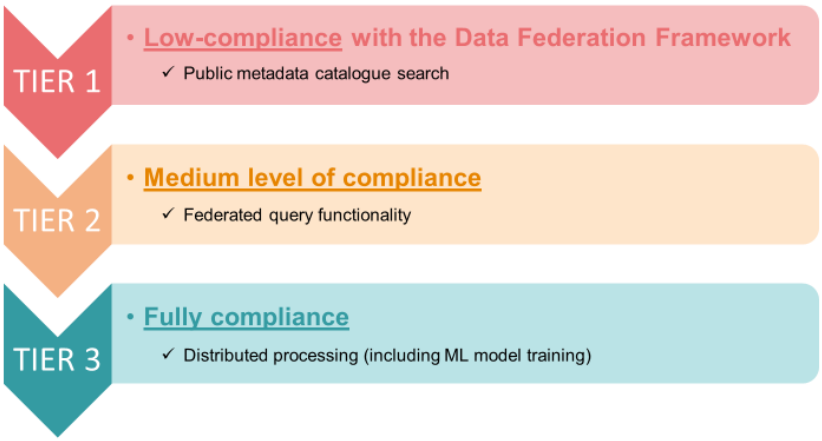

Figure 1: EUCAIM Data Federation Framework compliance tiers

Tier 1 Minimal Compliance with the Data Federation Framework

Data will be accepted by the Federation without compliance requirements concerning the source repository (mainly linked to an existing research project in the European framework) and data quality specifications established by the clinical centre of origin (in case of a clinical environment). The functionalities offered by the EUCAIM platform will be limited accordingly: only publication and visualisation of the dataset in the public metadata catalogue will be possible, allowing basic centralised filtering. The data does not have to comply with the common data formats (EUCAIM's Hyper-ontology). Neither federated/distributed processing capabilities nor a homogeneous framework for research will be available. The datasets in the public catalogue will be listed and made accessible (under the defined data request process). The DU-Rs will be warned that the data use is under these conditions.

Summary Tier 1:

- Datasets are registered in the public catalogue according to the metadata specification for the collections (D5.1)
- EUCAIM Access Committee manages contact points for the negotiation of the access requests in cooperation of the federated data holder.
- Specific services at the data holder's side are not required.

Tier 2 Compliance with the Federated Query

Compliance with EUCAIM's Federated Query service requires a stronger involvement of the DHs. It allows for improved visibility and usability of the data. Federated queries require provision of (meta)data according to EUCAIM's Hyper-ontology, and/or operating a local "mediator" service in order to execute the federated queries and report back the aggregated results. Datasets adhere to the standardised EUCAIM common data model (CDM), making it easier for researchers to use multiple datasets from Tier 2 in their projects compared to those from Tier 1.

Summary Tier 2:

- Datasets are registered in the public catalogue as in tier 1.
- The data holder is also integrated into the federated search. This requires:
    - ++ To develop the mapping component between the data structure and the hyperontology (see D5.1).
    - ++ To install a mediator service that is accessible from the central services of EUCAIM.
    - ++ To extract a snapshot of the metadata of the data available in the research repository.
- This service is lightweight, and it requires a standard computer (i7 or equivalent,16 GB RAM, 1 TB data, Docker installed).

Tier 3 Full Compliance with the Data Federation Framework

Full data compliance entails alignment with a wide set of requirements including data harmonisation, annotation and quality assessment. Tier 3 compliance should be the ultimate goal for DHs and RCs in order to achieve best possible usability and impact of their datasets within the Federation. It goes beyond the Federated Query capabilities of Tier 2 and enables federated processing, including Machine Learning (ML) and other advanced data processing techniques.

Summary Tier 3:

- Fulfil requirements of tier 2.
- Installation of the processing service, which requires:
    - ++ Development a “materialisation” service that copies and prepares on the fly the local data according to the hyperontology scheme.
    - ++ Provision of a service for running the federated processing (a Docker container).
    - ++ Provision of enough resources for processing the data (minimum resources to be defined)
    - ++ 24/7 availability of the service with an SLA of 95% and human support on working hours.

The following table illustrates availability and findability of data according to tiers.

		Tier 1 (minimal compliance)	**Tier 2 (medium compliance)**	**Tier 3 (full compliance)**
Catalogue	Datasets can be registered manually in the federated catalogue	*	*	*
	Datasets are automatically fed into the federated catalogue			*
Hyperontology	Dataset metadata should be compliant with the hyperontology	*	*	*

	specification.			
	Image data are compliant with the hyperontology specification		*	*
	Clinical data are compliant with the hyperontology specification			*
Negotiator	Contact points are included in the Negotiator process	*	*	*
Federated Search	A federated search service running on the provider's side is available		*	*
Distributed processing	Data can be provisioned in a compatible format.			*
	A processing service is included with the provider			*
	A marketplace of application synchronised with the general catalogue is available			*
AAI	Compatibility with EUCAIM AAI is provided			*

Table 1: Overview availability and findability of data in the EUCAIM tier structure

4.2. Data Holders, Tool Providers, Research Communities and Data User-Researcher's Demands for Training

EUCAIM intends a low-threshold entrance gate for data provision with subsequent option for upgrading. At Tier 1 level data can be provided and registered in the public metadata catalogue. Consecutive upgrading of these datasets to tier 2 and 3 is possible.

High demands for training activities are accordingly expected particularly for DHs and RCs in tier 2 and 3, or for DHs and RCs willing to upgrade their data and will facilitate smooth provision of data. Since the first on-boardings of new DHs, TPs and RCs will take place after submission of this deliverable, we will review the demands for training in the context of D2.4 Training Evaluation. We expect that DHs, TPs, RCs and DU-Rs will be challenged with the different requirements which have to be met and will need training activities and materials answering their questions. A collection of central questions is summarized in the

table below. The training catalogue introduced in section 6 will refer to these questions. The list of questions should be updated according to the monitoring and evaluation of the training plan.

Frequently asked Questions of DH, TP, RC and DU-Rs (initial expectation):

Topic	Target Group	Demands for training
Legal & Ethical Data Provision	DHs, TPs, RCs	▪ What are the legal and ethical requirements I have to meet? ▪ Do these requirements differ according to the tier/level of compliance with the Data Federation Framework? ▪ Which agreements (DTA, DSA, etc.) and related documentation do I have to provide? ▪ Which de-identification methods are used?
Legal & Ethical Data Usage	DU-Rs	▪ What are the legal and ethical requirements I have to meet? ▪ Do these requirements differ according to the tier/level of compliance with the Data Federation Framework? ▪ Which agreements (DTA, DSA, etc.) and related documentation do I have to provide? ▪ Which de-identification methods are used? ▪ What information about my project do I have to provide (research protocol, IRB approval)?
Technical (Data)	DHs, RCs	▪ Will there be resources (e.g. GPUs, RAM….) available for AI training or do I have to train deep learning models on my own machine? ▪ How does federated ML training work in EUCAIM? ▪ What technical (hardware, software, connection capabilities), data quality, interoperability, requirements must be met in each tier? ▪ Is there a support team to assist me during the process?
Technical (Tools)	TPs, RCs	▪ Will there be resources (e.g. GPUs, RAM….) available for AI training or do I have to train deep learning models on my own machine? ▪ How does federated ML training work in EUCAIM? ▪ Is there a support team to assist me during the process?

Technical (Data)	DU-Rs	▪ Can I download data to my own storage?
Data Provision	DHs, RCs	▪ How do I provide data? ▪ How is the data secured? ▪ What kind of data can be provided? ▪ Which data formats are allowed? ▪ Does the data have to be pseudonymised/anonymized? ▪ Is data curation required? ▪ Will I have to take care of the harmonisation, annotation and labelling of the images? ▪ What does the data provision process look like? ▪ Who should I contact within my organisation (IT dept., legal dept.) ▪ Is there any funding for data provision?
Data Provision	DU-R	▪ How can I provide feedback (i.e. wrong annotation, critical artefacts, incongruences, etc.) about the EUCAIM data I am using?
Tool Provision	TPs, RCs	▪ How do I provide tools? ▪ How is the tool secured? ▪ What kind of tools can be provided? ▪ What are the technical requirements I have to meet? ▪ What documentation do I need to provide? ▪ What does the tool provision process look like? ▪ Is there any funding for tool provision?
FAIR Data & Data Access	DHs, RCs	▪ What can my data be used for? ▪ How is it retrieved? ▪ Will I be informed about the usage of my data? ▪ Is there specific acknowledgment (in addition to generic reference to EUCAIM) about the use of the data I provided? ▪ How does EUCAIM ensure the adherence to the FAIR data principles?

FAIR Data & Tool Access	TPs, RCs	▪ How do I access data? ▪ What kind of data is available? ▪ How can I be informed about updated data applicable to my tool? ▪ How do I access results? ▪ How can I compare my tool against others? ▪ Which data formats are available? ▪ Which are the tools available through EUCAIM? ▪ Can I use and combine them with my tools? ▪ Do I have to provide source/open code for my tools? ▪ Do I have to publish the final models (in addition to the source code)? ▪ How does EUCAIM ensure the adherence to the FAIR data principles?
FAIR Data & Data/Tool Access	DU-Rs	▪ How do I access data/tools? ▪ What are the main steps in the process of getting access to the data? ▪ What kind of data/tools are available? ▪ How can I be informed about updated data/tools? ▪ What are the FAIR data principles? ▪ How can I ensure to use the data according to the FAIR data principles?
General platform use	DHs, RCs	▪ What is the tier system and how is it structured in EUCAIM? ▪ How can I increase my level of integration/move up a tier (support from the consortium, funding)? ▪ What are the main differences between uploading data to the central repository and being a federated node? ▪ What are the quality standards for the data available on the platform? ▪ What is the difference between the public catalogue and the federated query? ▪ How can I use EUCAIM for the development of AI tools? ▪ How can I use EUCAIM in the context of a clinical study?
General platform use	TPs, RCs	▪ What are the quality standards for the tools available on the platform? ▪ What is the difference between the public catalogue and the federated query? ▪ How can I negotiate with the TPs?

		▪ How can I use EUCAIM for the development of AI tools?
General platform use	DU-Rs	▪ How can I register on the platform? ▪ What is the difference between the public catalogue and the federated query? ▪ How do I use the public catalogue? ▪ How do I do a federated query? ▪ What are the quality standards for the data available on the platform? ▪ How can I negotiate with the DHs? ▪ How can I use EUCAIM in the context of a clinical study? ▪ How can I use EUCAIM for the development of AI tools?

Table 2: Expected demands for training

5. Integration of Training Activities into the on-boarding processes for Data Holders, Tool Providers, Research Communities and Data User-Researchers

The following section introduces the on-boarding processes for all target groups and illustrates the integration of training activities into these processes.

5.1. The On-boarding process for Data Holders, Tool Providers and Research Communities

The onboarding of DHs, TPs and RCs consists of three main approaches to incorporate data and/or tools into the infrastructure: (1) on-boarding as new stakeholder and data/tool donor without becoming a partner of the consortium; (2) on-boarding through the internal call for existing consortium partners and (3) on-boarding through the external open call for new beneficiaries. The main difference between the on-boarding as stakeholder and on-boarding through the internal and external open call is that the latter require the involvement of the Access Committee (AC) into the on-boarding process. The three approaches are described below. For further reference see also D2.1 On-boarding Invitation Package, Terms of Reference for the EUCAIM Access Committee, D7.1: Rules for evaluation and prioritising of use case applications from the open call.

1) On-boarding through collaboration as **stakeholder**. Entities not affiliated in the consortium may:
    a) participate in the open calls to become partners (everyone can participate in these open calls, understanding that not all participants will become partners);
    b) become data donors (data altruism), entailing non-economic benefits (other benefits instead).

2) On-boarding through the **internal open call**: consortium partners collaborating with data/tools/data users are offered a "mock-up" of the external open call. It requires a project proposal/use case. AC approval is needed for data/tool deposition in the platform, as outlined in D7.1.

3) On-boarding through the **external open call**[3]: New partners outside the consortium can join the EUCAIM consortium and receive funding. Similar to the internal calls, this requires a project proposal/use case, AC approval is needed for data/tool deposition in the platform, as outlined in D7.1.

5.1.1 The On-Boarding Process for Data Holders, Tool Providers and Research Communities as Stakeholders

EUCAIM invites Data Holders, Tool Providers and Research Communities to donate data even without participation in the external open call. This data altruism approach involves the following steps in terms of on-boarding. The training activities offered in the process are highlighted.

a) Initial contact is facilitated through e.g., that data holders can contact EUCAIM directly through the website (cancerimage.eu) or at contact@cancerimage.eu, or can be contacted by the Evangelization Team.

b) Data holders will receive the on-boarding invitation package, the form for the Expression of Interest and the Interoperability Checklist. The Engagement Team will follow up on their status and activities through the EUCAIM CRM tool.

c) The Expression of Interest must provide details on contact data as well as the organisation, experience and capabilities.

d) The functional and technical requirements of EUCAIM's infrastructure will be compared to the infrastructure and in-place settings at the data holder's site through the Interoperability Checklist.

e) Once the data holder's application is accepted, they will receive and sign a Data Collaboration Agreement and the Data Sharing Agreement/Data Transfer Agreement. The DH will be provided with a **mandatory training module on legal and ethical aspects for data provision**[4] (Module 1).

f) Successively, the DH will receive access to the training catalogue consisting of a number of **optional training modules**:

[3] For reference from D7.1: An Evaluation Committee (Access Committee, AC) for receiving and evaluating the proposals submitted for the open call has been first set-up. The AC has the responsibility of establishing the open call rules and evaluation criteria. Furthermore, it will evaluate the received applications based on the compliance with the established scientific, technical, ethical and legal requirements. The Access Committee will then score the applications and forward the evaluation results to the EUCAIM Management Board for final decision about acceptance or rejection. The Management Board decision about selected applicants will finally be confirmed by the EC office based on compliance with the Digital Europe Programme Regulation.

[4] For a description of the training catalogue see 6. Training modules.

- **Technical training** (Module 2)
- **Training for data and tool provision** (Module 3)
- **Training on FAIR data & data access** (Module 4)
- **General platform use** (Module 5)

g) The Technical Support Team and the ELSI (Ethical, Legal and Social Issues) Team will audit the on-boarding process, helping with any issues that may arise via a Helpdesk (e.g., adaptation or development of the necessary infrastructure and software, clinical CDM and data protection guidelines).

h) Once the training and data preparation (i.e., preprocessing, cleaning, harmonisation, de-identification, annotation, etc.) are finalised, the data holder will be able to provide EUCAIM with a first dataset. They will be in iterative contact with the Engagement Team that will supervise any activity, including the data compliance with FAIR principles.

i) Actions taken by the data holder will periodically be reviewed by the Engagement Team. The EUCAIM consortium (including all new data holders, among others) will publish papers promoting EUCAIM in the Q1 relevant literature (e.g., European Radiology, Insights into Imaging, etc.). The Evangelization Team may also ask data holders to participate in online interactive workshops to explain the impact that data had on the development of new radiomics and AI tools in cancer research.

j) The Data Monitoring Team will periodically update the data provision statistics and will present them on the EUCAIM webpage. Permission of the data holder's is required as part of the dissemination and communication strategy.

5.1.2 On-boarding through the Internal and External Open Call

Testing of the technical infrastructure and the workflow for data provision should be performed at an early stage of the EUCAIM project. To this end, an internal open call will be launched to incorporate pre-existing data and tools provided by the partners of the EUCAIM consortium.

In addition, an external open call will be launched for new data/tools holders willing to join the consortium and to add new cancer image databases in the federation. Successful applicants of the call will be included in the EUCAIM consortium as project beneficiaries.

The call will also target researchers interested in using the EUCAIM platform to address real life clinical questions, for which large numbers of datasets and possibly technical/ethical/legal support are required.

This open call will pursue the onboarding of i) new data holders, increasing the geographic dimensions, data modalities or cancer targets, and ii) the uptake of new trustworthy AI algorithms trained on the data of the repository.

Supported by the Technical Board and by the Ethical and Legal Board, the EUCAIM Access Committee (established in June 2023) has set up the rules for application and evaluation in alignment with the European Commission, and will receive and evaluate the proposals and finally prioritise them. The final acceptance/rejection decision will be taken by the Management Board taking into consideration the indications from the Access Committee.

Figure 2: Workflow of the Access Committee

Access to the training catalogue will be provided as soon as all required documentation[5] have been submitted, i.e. the Data Sharing Agreement, ethics approval, technical requirements documentation.

5.2. The On-boarding and Registration Process for Data User-Researchers

DU-Rs are considered the end-users of the platform. They will be able to either explore the public catalogue for available metadata without registering in the platform, or perform a federated query on Tier 2-3 datasets in case of registration. They can request access to data and tools and, if access is granted, process the data using the tools available on the platform or their own AI tools. In order to be granted access, they have to comply with the corresponding requirements[6]. Most importantly, DU-Rs will have to request access based on an approved research project, which can range from final undergraduate projects to large funded grants.

Registration on the platform is required via the Life Science AAI. Several options are available for this, such as affiliation to organisations, using an ORCID, a LS Hostel account. Authenticaton will be required before using the tools and services of EUCAIM. As part of this authentication process, the user has to accept several usage conditions as defined in the Acceptable Use Policy of EUCAIM services.

Common Conditions of Use:

The specific Conditions of Use will be included in the documents defined by WP3. Some of the Common Conditions of Use will be:

- The User agrees to be a bona fide researcher with (1) an intention to generate new knowledge and understanding using rigorous scientific methods, (2) an intention to publish the research findings and share the derived data in the scientific community, ideally without restrictions and with minimal delay, for wider scientific and eventual public benefit, and where (3) the intended activities are not inconsistent with legal and ethical requirements or widely recognised good research practice.
- The user will avoid any attempts to reverse privacy enhancing technologies (i.e., pseudonymization, anonymization) applied to the data.

[5] See also draft procedure for the open call.
[6] See also D4.3 First rules of participation report.

- If possible, any incidental findings will be reported back to the corresponding body within the EUCAIM consortium.

After completion of this process and agreeing to the Conditions of Use, DU-Rs will also receive access to the training catalogue and will be able to visit tutorials and access any other training materials such as FAQs or Howto-Manuals.

5.3 Timeline for Training Activities

As illustrated in the sections above, the training will be integrated into the on-boarding process for stakeholders as well as the processes of the open internal and external calls. The development of the training materials and organisation of tutorials (available also as recordings) will start in the beginning of 2024. We expect that the complete training materials will be available by mid-2024. At this point we will be able to perform an evaluation based on the internal use cases that require training. The following figure illustrates the timeline for training until deliverable D2.4 Training Evaluation is due by the end of 2024.

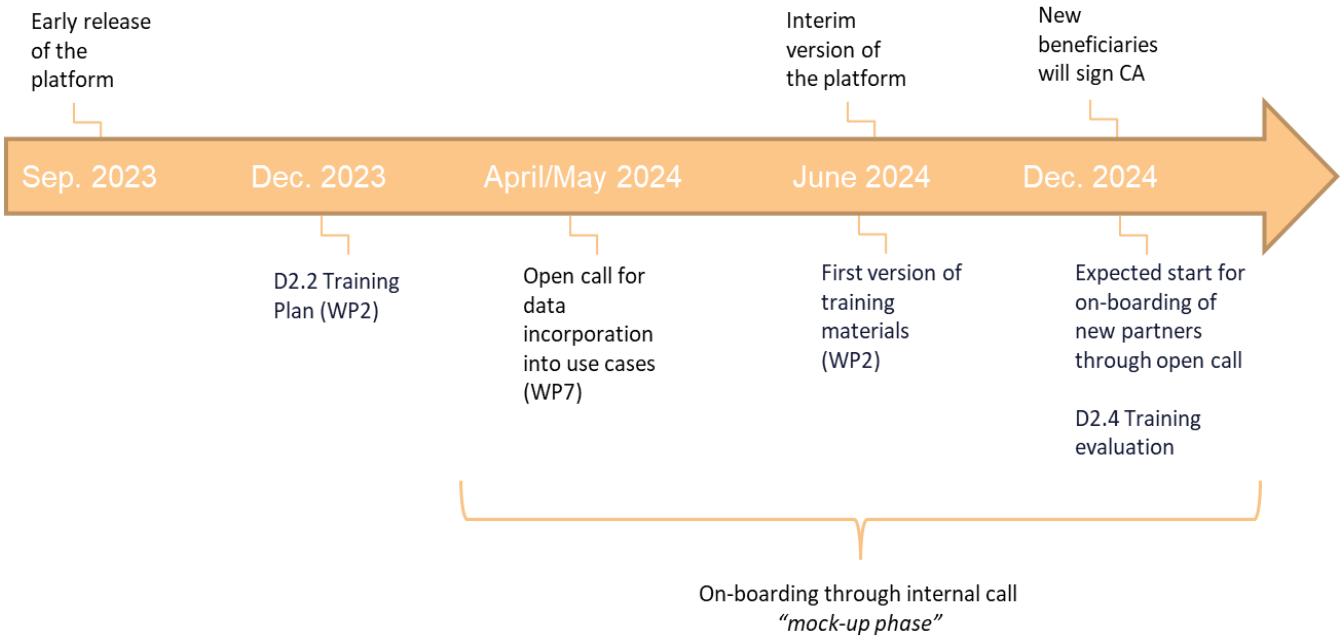

Figure 3: Timeline for on-boarding, training and internal/external open calls

6. Training Catalogue

The following section describes the training catalogue consisting of the different thematic modules offered in EUCAIM. The modules can be selected according to target group and contain different sessions (such as tutorials, workshops, Q&A sessions) and training materials (such as FAQs, Howto-manuals).

6.1. Training for Data Holders, Tool Providers, Research Communities and Data User-Researchers

The training in EUCAIM will consist of a training catalogue addressing all target groups. It will consist of one mandatory training module on legal and ethical aspects (Module 1) including a quiz and four optional training modules with information focusing on technical provisions (Module 2), data and tool provision (Module 3), FAIR data and data access (Module 4) and general platform use (Module 5):

Training Catalogue	
Module 1 (mandatory): **Training on legal and ethical aspects for:** **(a) Data provision** **(b) Data usage**	
▪ GDPR ▪ safety/privacy ▪ authorisation process ▪ terms and conditions	
Module 2 (optional): **Technical training**	**Module 3 (optional):** **Training for data & tool provision**
▪ set up of local node (hardware/software) ▪ connection to central infrastructure	▪ federating data ▪ tools for data harmonisation, standardisation ▪ tools for data anonymisation, segmentation, annotation, ETL tools
Module 4 (optional): **Training on FAIR data & data access**	**Module 5 (optional):** **General platform use**
▪ how to make use of the platform in a FAIR compliant way ▪ how to access data/tools ▪ who can access data/tools	▪ overview of platform ▪ how to use the catalogue ▪ how to use the federated query ▪ explicatory use cases: ➔ platform use for clinical studies ➔ platform use for development of AI tools

Figure 4: Training Catalogue

Upon accessing the catalogue, the respective target group according to DH, TP, RC or DU-R can be selected.

Subsequently, the user will be able to access all according sessions and training materials. Participation in the training activities is optional, except the mandatory legal and ethical training. The users will be able to select further optional training activities/materials which are considered relevant.

The following table illustrates the available training activities and materials for the different target groups. Intervals for internal revision are given to ensure that relevant updates can be included.

Module	Target Group	Brief description of contents	Potential materials, tools	Intervals for revision/ expert feedback
1a) Legal & Ethical: Data Provision	DHs, TPs, RCs, DU-Rs	**Tutorial**: Legal and ethical requirements for data/tool provision in a nutshell **FAQs**: Legal and ethical requirements, agreements and documentation, the tier system, defining anonymization, pseudonymisation, de-identification	Tutorial, FAQs	Feedback every 6 months
1b) Legal & Ethical: Data Usage	DHs, TPs, RCs, DU-Rs	**Tutorial**: Legal and ethical requirements for data/tool retrieval in a nutshell **FAQs**: Legal and ethical requirements, agreements and documentation	Tutorial, FAQs	Feedback every 6 months
2a) Technical (Data)	DHs, RCs	**Tutorial**: Technical requirements and guidance for data provision (required resources, federation, set-up of local node, central storage) **FAQs**: required hardware/software/ connection, set-up of local node, connection to central infrastructure **Howto**: Brief description of technical set-up	Tutorial, FAQs, Howto	Feedback every 3 months
2b) Technical (Tools)	TPs, RCs	**Tutorial**: Technical set-up and guidance for tool provision	Tutorial, FAQs, Howto	Feedback every 3 months

		FAQs: required resources, federated training in EUCAIM **Howto:** Brief description of technical set-up		
3a) Data Provision	DHs, RCs	**Tutorial:** Data preparation (data de-identification, data annotation, data quality and data harmonization) **Tutorial:** EUCAIM hyper-ontology and ETL tools (how to adapt clinical data to EUCAIM hyper-ontology) **FAQs:** on data pre-processing (data standards, pseudo-/anonymization requirements, etc.), data security, curation, funding	Tutorial, FAQs, Howto	Feedback every 3 months
3b) Tool Provision	TPs, RCs	**Tutorial:** tool dockerization and registration in bio.tools. **Tutorial:** tool validation (description of tool validation and benchmarking process) **Tutorial:** Good practices (version control, monitoring and tool usage) **FAQs:** tool provision, security, documentation, funding	Tutorial, FAQs, Howto	Feedback every 3 months
4 FAIR Data & data access	DU-Rs, TPs, DHs, RCs	**Tutorial**: FAIR data and data access (FAIR principles, FAIR-support in EUCAIM, how will the data/tool be used, who can access it) **FAQs:** FAIR data, data/tool access in EUCAIM, publication	Tutorial, FAQs, Howto	Feedback every 3 months

5 General platform use	DU-Rs	**Tutorial:** How to use the catalogue/federated query **Tutorial:** federated learning and distributed analysis **Tutorial**: EUCAIM for developing AI tools **Tutorial**: EUCAIM in the context of clinical studies **FAQs**: Tier system, central repository vs. federated nodes, quality standards, public catalogue vs. federated query, developing AI tools with EUCAIM, using EUCAIM for clinical studies	Video, Tutorial, FAQs, Howto	Feedback every 3 months

Table 3: Training modules according to target group

6.2. Individual Training Sessions

Individual training sessions will be available for data holders, tool providers as well as research communities if necessary in individual cases. This includes:

- Online live tutorials with Q&A for larger groups/institutions/research communities
- On-site training for cases where this is specifically required (e.g. because of complicated technical issues that may arise)

Due to limitations in staff availability and resources dedicated to traveling on-site for training purposes, these individual training sessions will be offered by EUCAIM only to those partners or stakeholders that require specific attention. The candidates for dedicated individual training sessions will be determined in close consultation with the Technical Board and the Technical Support Team. The WP2 training team will support the organisation of these sessions and identify the corresponding experts to implement the training sessions.

6.3. Legal and Ethical Training for Internal Consortium Members

For members of the EUCAIM consortium, WP3 has offered an internal training session on all legal and ethical aspects relevant for EUCAIM. This training session was conducted on November 2, 2023, and is available in recording to all EUCAIM consortium members. It consists of four main sections:

1. Data protection by design strategies
2. Requirements for European Health Data Space (EHDS) implementation
3. Security aspects in data privacy, AI and EHDS
4. Future requirements on AI

Section 1 (Data protection by design strategies) illustrates the workflow for GDPR compliance and each partner's contribution to EUCAIM's compliance model.

Section 2 (Requirements for EHDS implementation) informs about the relation between EHDS and the already existing GDPR regulations. In this context, the different configurations of a data space and the different modes to share data are explained and the resulting challenges for EUCAIM are addressed.

Section 3 (Security aspects in data privacy, AI and EHDS) outlines the security by design approach, which entails the adoption of a security framework for EUCAIM as well as in relation to AI and EHDS.

Section 4 (Future requirements on AI) introduces the proposed EU AI Act and necessary requirements in terms of technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental well-being; as well as accountability.

6.4. Training on Updates

In addition to the modularized training activities, ad hoc online workshop sessions will be offered in accordance with new platform developments or other updates.

These sessions will focus on the individual target groups and address the updated content. Thus, we can ensure that all target groups are informed about the latest updates.

All live sessions will be listed in the training calendar and announced through EUCAIM's communication channels. Recordings of these sessions will be made available in the training catalogue.

6.5. External Training Materials

In addition to the training catalogue developed by the EUCAIM consortium, training materials from TPs sharing their software applications and tools through the platform will be collected to create a training library with external training materials.

7. Monitoring and Evaluation of Training

The scope of D2.4 is the evaluation of all training activities, best-practices, lessons learned, including a recommendation for future continuation of training activities. Therefore, the following measures will be implemented to ensure that the training offered efficiently supports the different target groups. The measures conducted in this regard are:

- evaluation of training sessions and materials through survey (starting with the internal use-cases);
- periodic revision of training materials with experts from the corresponding WPs (legal, technical/operational, data-related, etc.);
- collection of lessons-learned based on the survey results as well as the feedback from the EUCAIM staff in conducting the training sessions.