EUCAIM proposal to stakeholders in Cancer Screening Programs

This document has been prepared with Work Package (WP) 2.

Aim of the proposal

The purpose of this document is to promote the collaboration between the European Federation for Cancer Images (EUCAIM)[1] and Cancer Screening Programs.

Europe's Beating Cancer Plan (EBCP) works to prevent cancer while ensuring the highest possible quality of life for cancer patients, survivors and families. The plan is structured around key areas, such as early detection, diagnosis, treatment, and quality of life of cancer patients and survivors. Following the European Commission's proposal to strengthen cancer prevention through early detection, the Council of the European Union adopted a new approach on cancer screening as one of the main initiatives of the EBCP. This new approach aims to introduce the latest available scientific developments and population coverage in cancer screening programs for breast, cervical and colorectal cancer screenings, extending to prostate and lung cancer in a stepwise approach.

Following the Council Recommendation on Cancer Screening, at present there are several national and regional projects and pilots on image-based cancer screening, including the national Breast Cancer Screening Programme from Slovenia (DORA)[2] and the Lung Cancer Screening in French women using low-dose CT and Artificial intelligence for Detection (CASCADE-LUNG) from France[3]. Furthermore, several other governmental projects are well underway as summarized in the Organization for Economic Cooperation and Development (OCDE)[4] country cancer profiles.

To enhance the role of medical imaging in cancer screening studies, the EUCAIM project, funded by Horizon Europe under the call for a Public Sector Open Data for AI and Open Data Platform (ref. DIGITAL-2022-CLOUD-AI-02-OPEN-AI), is deploying the Cancer Image Europe infrastructure, which stands as the largest cancer imaging infrastructure in Europe of FAIR (Findable, Accessible, Interoperable and Reusable), de-identified, imaging and related data that is recruited from research projects, scientific programs and daily clinical practice (real-world data), incorporating the latest advancements in artificial intelligence and digital technologies. EUCAIM provides a

[1] EUCAIM: https://www.egi.eu/project/eucaim/

[2] DORA: https://dora.onko-i.si/english

[3] Revel, M.-P., Abdoul, H., Chassagnon, G., Canniff, E., Durand-Zaleski, I., & Wislez, M. (2022). Lung CAncer SCreening in French women using low-dose CT and Artificial intelligence for DEtection: The CASCADE study protocol. *BMJ Open*, *12*(12), e067263. https://doi.org/10.1136/bmjopen-2022-067263

[4] OECD: https://www.oecd.org/

service for permanent archiving and sharing of all types of healthcare data, including already existing repositories, through the consolidation of fragmented datasets into a comprehensive Atlas of Cancer Images, while also ensuring data sovereignty and compliance with ethical standards. The final aim is to unlock the potential of imaging and big data in oncology, advancing a future in which personalized and effective care is provided to every patient.

EUCAIM will provide a comprehensive Dashboard for data discovery, metadata harvesting, federated search, and distributed pre-processing and processing, including federated and privacy-preserving machine learning techniques. This infrastructure will stimulate the European market's innovation with new tools and services. EUCAIM will also contribute to shaping the legal grounds for such operation on a pan-European scale, adapting to the particularities of different countries in the management of clinical data. To do so, EUCAIM will implement a federation of stakeholders compliant with commonly agreed legal grounds, defining common data models, ontologies, quality standards, FAIR principles, and de-identification procedures. EUCAIM will align with the European Health Data Space (EHDS) initiative toward a sustainable flagship repository of high-quality data (including outcomes) and tools.

As stakeholders, the cancer screening programs that aim to participate will be allowed to join in a federated way, where the data will not be exposed to the federation, and specific datasets will be created for approved projects. Project by project, the cancer screening programs will decide whether to transfer the specific datasets to the Central Repository or keep them within their Federated Node.

EUCAIM is willing to participate in the different European Cancer Screening Programmes by fostering the use of medical imaging and Artificial Intelligence in several ways, including:

- Harmonizing acquired images by processing them towards a common framework (resizing, normalization, denoising, vendor-protocol agnostic), allowing image quality assurance, image comparison and reproducible image processing.

- Detecting probabilities to develop tumors, pre-tumor niches and early tumors is pivotal. Pre-tumoral niche can be defined as those changes in the stromal microenvironment that precede macroscopic tumor development and morphological detection. Early AI based depiction will allow to either reschedule faster follow-up studies or change image modalities to more sensitive ones (such as contrast enhanced tomosynthesis or MR for breast cancer). This early detection would greatly increase the chances for survivorship and would allow physicians to take measures years before the tumor develops and becomes macroscopic. (Figure 1).

- Decreasing the false positive by improving the characterization of lesions using Convolutional Neural Networks (CNN) to a more precise diagnosis of benign conditions. The precise characterization of lesions will decrease false results at the screening level.

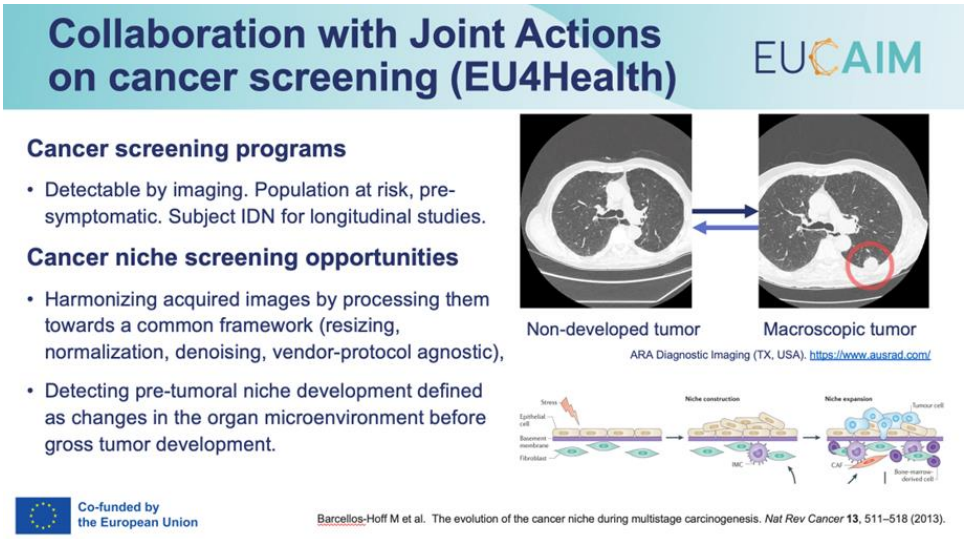

Figure 1. Proposal of collaboration with cancer screening programs

Benefits of collaboration with EUCAIM

Becoming a stakeholder in EUCAIM not only helps in the advance of cancer research but also offers a multitude of benefits that significantly contribute to healthcare research and innovation. Here's what it can be expected:

- **Efficient data management and quality enhancement**
  - EUCAIM provides robust tools and guidance that facilitate data management processes. This ensures the standardization, easy to access, and interoperability of the data for research and analysis.
- **Collaboration and recognition**
  - Involvement in new large cancer studies: EUCAIM collaborates with Member States and other stakeholders to facilitate large-scale, multicentric cancer-fighting studies, fostering collaboration through Research Communities (RC) to address new challenges in cancer research and innovation. It allows the promotion of collaboration and knowledge exchange among institutions.
  - Enhanced visibility and reputation: As an active participant in EUCAIM, your institution becomes an integral part of Europe's Beating Cancer Plan. This affiliation increases your institution's recognition, attracting additional funding and fostering partnership opportunities.
  - Opportunity for revenue: Stakeholders can benefit from sharing data through partnerships within the European Data Governance Act (DGA) and EHDS framework.
- **Utilization of EUCAIM's resources**
  - Training program: EUCAIM will support stakeholders by setting up a training program, that will include a training curriculum and specific training materials (on-site, online as well as self-training) and interactive workshops.
  - Technical support: Comprehensive technical support and assistance for efficient data management processes, ensuring data standardization, accessibility, and usability for research.
  - Facilitated compliance with EHDS regulation: According to the current regulatory proposal, data holders would be required to make health data available to third

parties, for secondary use for the permitted purposes. EUCAIM provides you with the functional, operational and legal framework to facilitate compliance with this new regulation.

- **Socioeconomic impacts**
  - Health benefits for citizens: EUCAIM increases data visibility, benefiting citizens through health research. In addition, it will improve healthcare by aiding in making diagnoses more accurate and cancer prevention more effective, supporting the development of personalized medicine and customized treatments.
  - Smart use of resources: EUCAIM empowers decision-makers to wisely allocate healthcare resources, ensuring efficient utilization.
  - Accelerating research and development: Promoting faster discoveries and innovations through access to standardized and quality data. That also encourages responsible data use, building trust among stakeholders, and ensuring secure and transparent access to health data.

EUCAIM infrastructure

For joining the federation, the cancer screening programs can become a Federated Node, having autonomous control over the data, or they can upload the anonymized data to the Central Repository. In both cases, they are required to provide the (meta)data of their dataset with additional documentation requirements and benchmarking information to make it available to the central services as part of the metadata catalogue shared through the EUCAIM infrastructure.

- Federated Node

A federated provider hosts the data locally, exposes it in accordance with EUCAIM's interfaces and specifications, registers their prepared datasets of specific projects in the EUCAIM's metadata catalogue and eventually offers local processing capabilities. In the specific case that processing capacity is not available, a dedicated space can be enabled for it within EUCAIM. In all cases, data quality and integrity are ensured as well as compliance with privacy and security regulations to facilitate collaborative data sharing and analysis within the federated system. Once the projects in which the Cancer Screening Program has participated have been concluded, the specified datasets created may be moved to the Central Repository, through a specific agreement. Cancer Screening Programs may join EUCAIM as a Federated Node using the infrastructure available on-site after signing the Data Sharing Agreement (DSA).

- Central Repository

This case occurs when the Cancer Screening Program transfers the datasets to the Central Repository, accomplishing a Data Transfer Agreement (DTA), to make them sustainable and long-term available. EUCAIM will guide through data de-identification and will guarantee that the archived datasets will be accessible to other researchers and used for research in a legal, safe and privacy-concerned way.

- Research Community

Research Communities are groups or entities with a common research goal, typically formed through the course of already finalized, 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.

Data provision application framework

The data provision application process consists of the provider submitting an application to join EUCAIM, highlighting its capacity to provide valuable datasets while adhering to stringent data privacy and de-identification. The submission will be evaluated by the corresponding Governing Body. The procedures that need to be followed by the provider will be negotiated and a contractual document will be signed before the actual integration in the federation.

The specific datasets that are meant to be re-used for research will be shared/transferred between the parties. To this end, the access to the datasets is based on various main pillars:

- **Collaboration agreement (CA)** to foster symbiotic relationships built upon collaboration, co-governance, and long-term sustainability.

- **Data Sharing or Transfer Agreements** between the participating entities. The DSAs will apply in the case of a federated node, while DTAs will apply in the other cases of providers. The data made available is fully de-identified and can be re-used for research purposes, under Open Data rules, or in accordance with conditions specifically provided for in national or EU legislation. The data will be either uploaded to the Central Repository, managed directly by the EUCAIM's Central Hub or stored at the Data Holder's premises and connected through a federated node.

- **A repository governance model** that ensures a legitimate use of the data, an ethical and regulatory compliance framework in which access decisions are made by the Data Access Committee, defined by the common agreements. The repository governance model should also address the potential exploitation models.

- **Memorandum of Understanding (MoU).** EUCAIM will facilitate the creation of dedicated Research Communities (RCs) that will establish a research environment with their own dedicated spaces, complete with repositories, tools, and enhanced

communication features where the different groups or entities specialized in a particular study area could join as a community. Thus, EUCAIM will give the members the opportunity to initiate new projects within the EUCAIM infrastructure as well as to continue working with their data and tools. In return, their data and tools will be usable by other researchers, upon request.

The agreements between entities, repositories and infrastructures, will follow guidelines, best practices, and standards for building and operating an infrastructure that promotes responsible data sharing in accordance with the EU Privacy and Security Policy.

Data Compliance

Low-compliance with the Data Federation Framework (DFF) defined by EUCAIM will not impede the participation. Instead, EUCAIM will incentivize data curation and facilitate data transformation, if necessary, to align with the DFF. In addition, the Observational Medical Outcomes Partnership (OMOP) will be used to enable the integration of the multiple data sources and to support the analysis and exploitation capabilities of the datasets.

To accommodate diverse levels of data compliance, it has been established three technical tiers. These tiers are scalable for the Data Holders, allowing upgrading as the datasets are used in new research projects. EUCAIM will provide support and adaptation plans to assist the achievement of Tier 2 and Tier 3 compliance.

- **Tier 1:** Acceptance of data with low DFF compliance, suitable for entry-level participation.
- **Tier 2:** Medium compliance with the DFF, enabling federated search functionality.
- **Tier 3:** Full compliance with the DFF, allowing distributed processing, including machine learning model training.

Conclusion

The utilization of Artificial Intelligence (AI) has demonstrated its capacity to significantly aid in cancer risk assessment, detection, and characterization. This involves processing vast amounts of data to identify patterns and markers crucial in early cancer detection and individual risk evaluation. To achieve this, an extensive and diverse dataset is pivotal. The EUCAIM initiative aims to provide such comprehensive datasets for AI-based research and development, serving as a critical resource to enable more accurate and effective AI applications in cancer risk assessment and diagnosis. On the other hand, EUCAIM aims to foster an environment that not only facilitates the integration of new stakeholders but also maintains stringent data quality standards.

We invite you to join EUCAIM, contribute your valuable data, and be part of our mission to advance cancer research and innovation.