Welcome to the session
AI and Cancer: Unleashing Opportunities, Overcoming Challenges

Clustering event powered by INCISIVE and the AI4HI cluster

INCISIVE

Introduction to cancer medical imaging repositories and their importance in research and clinical practice

10:30 - 11:10

EUCAIM | AI4Health Imaging

Introduction to cancer medical imaging repositories and their importance in research and clinical practice.

Mrs. Gianna Tsakou
Gruppo Maggioli
INCISIVE Project

Mr. Oliver Díaz
University of Barcelona
RadioVal project.

Mr. Kostas Marias
HMU, Head CBML-ICS-FORTH
ProCAncer-I project

Mr. Ignacio Blanquer
Polytechnic University of Valencia
PRIMAGE and CHAIMELEON projects

Dr. Luis Martí-Bonmatí
La Fe Polytechnic University Hospital
EUCAIM, PRIMAGE and CHAIMELEON projects

INCISIVE | EUCAIM

Mrs. Gianna Tsakou

Gruppo Maggioli. Senior Project Manager

INCISIVE Coordinator

Key facts (1)

A multimodal AI-based toolbox and an interoperable health imaging repository for the empowerment of imaging analysis related to the diagnosis, prediction and follow-up of cancer

Duration: 42 months
From October 2020 to March 2024

Call: H2020-SC1-FA-DTS-2019
Topic: DT-TDS-05-2020 AI for Health Imaging

Funding: 9.995.727,50 €

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Key facts (2) – Our consortium

- **9** Universities / Data providers
- **7** SMEs
- **6** Research / academic institutions
- **3** Large industries
- **2** Other

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Key facts (2) – Areas of Expertise

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Key facts (3) – Areas of Expertise

Cancer clinical research/practice & health data sharing

German Oncology Center

Università di Roma Tor Vergata

IDIBAPS

ΕΛΛΗΝΙΚΗ ΑΝΤΙΚΑΡΚΙΝΙΚΗ ΕΤΑΙΡΕΙΑ HELLENIC CANCER SOCIETY

VISARIS

Patient representation & evaluation

Cyprus University of Technology

PASYKAF

Kingston University London

Legal and ethical issues

TIMELEX

Innovation and business planning

WHITE RESEARCH

Medtronic Further,Together

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Main objectives

AI-based toolbox that supports the decision-making of HCPs in relation to patient prioritization, cancer diagnosis and risk for metastasis

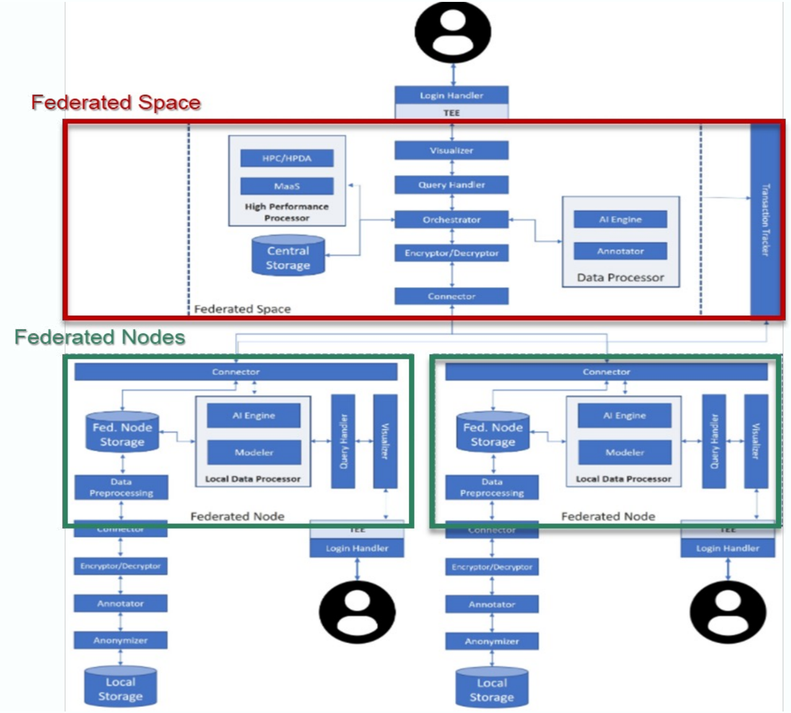
Interoperable pan-European **federated health data repository** (medical images & clinical data) that enables secure data sharing in compliance with ethical, legal and privacy requirements

TARGETED IMPACT ON:

- Healthcare professionals involved in cancer care
- AI developers
  AI experts
- Researchers

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Hybrid (Federated-Centralised) Approach



Hybrid storage:

- Both federated and central nodes available
- Selection based on data provider's needs
- Federated Learning mechanisms

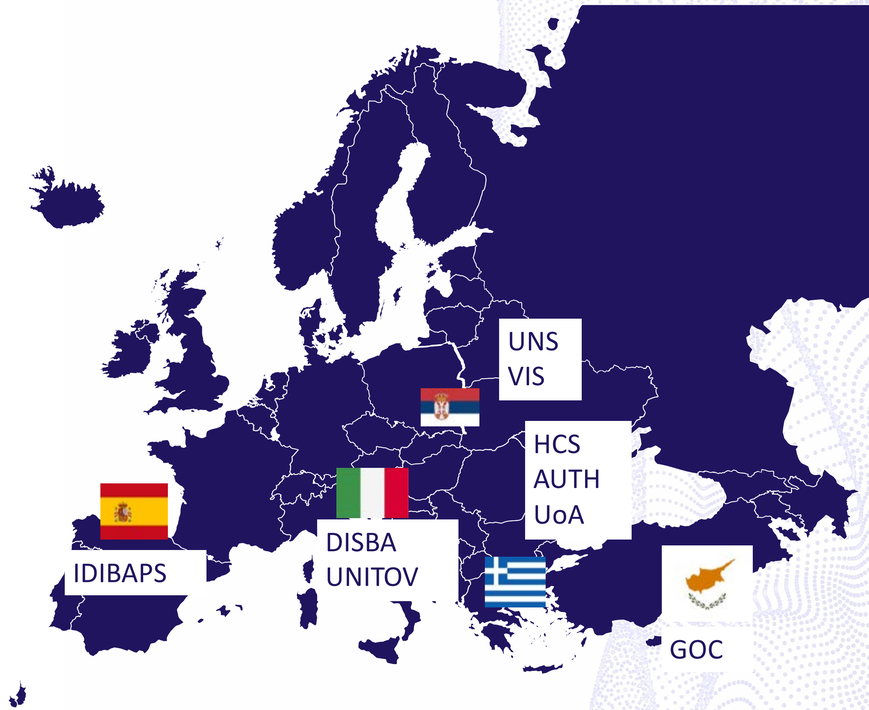
Data federation facilitates:

- aggregation of data from multiple sources
- ethical, legal and privacy compliance

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Data types

- **Imaging & accompanying** clinical & biological data for 4 types of cancer
- **9 Data Providers (12 hospitals) from 5 countries**
- **Imaging Data** (at diagnosis and subsequent timepoints)
  - MRI - CT
  - Ultrasound - PET-CT
  - Mammography - X-ray
  - Histopathological images
  - CT
  - PET-CT
  - X-ray
- **Accompanying Clinical Data (at diagnosis and timepoints)**
  - Demographics, treatment, mutations, lab results
  - Excel Template used for collection



These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Data collection process

Steps followed

ETHICAL APPROVAL ACTIONS

- From M3-M12
- Updates, where required

PATIENT SELECTION

- Medical image type and quality specifications
- Within hospital data collection
- Complete cases –Incomplete cases (less time points available)

IMAGE DE-IDENTIFICATION

- Comomn de-identification protocol defined
- CTP anonymized →
- INCISIVE-specific de-identification tool

CLINICAL DATA EXTRACTION

- Extraction of clinical data from EHR into predefined data templates

DATA QUALITY CHECK

- checks various data quality and structuring parameters

IMAGE ANNOTATION

- Annotation guidelines
- ITK SNAP tool →
- INCISIVE-specific annotation tool

DATA HOMOGENISATION

- Mapping to medical ontologies SNOMED-CT & LOINC (semantic)
- Transformation to FHIR data model (structural)

DATA SHARING through INCISIVE REPOSITORY

- Prepare and monitor data availability

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

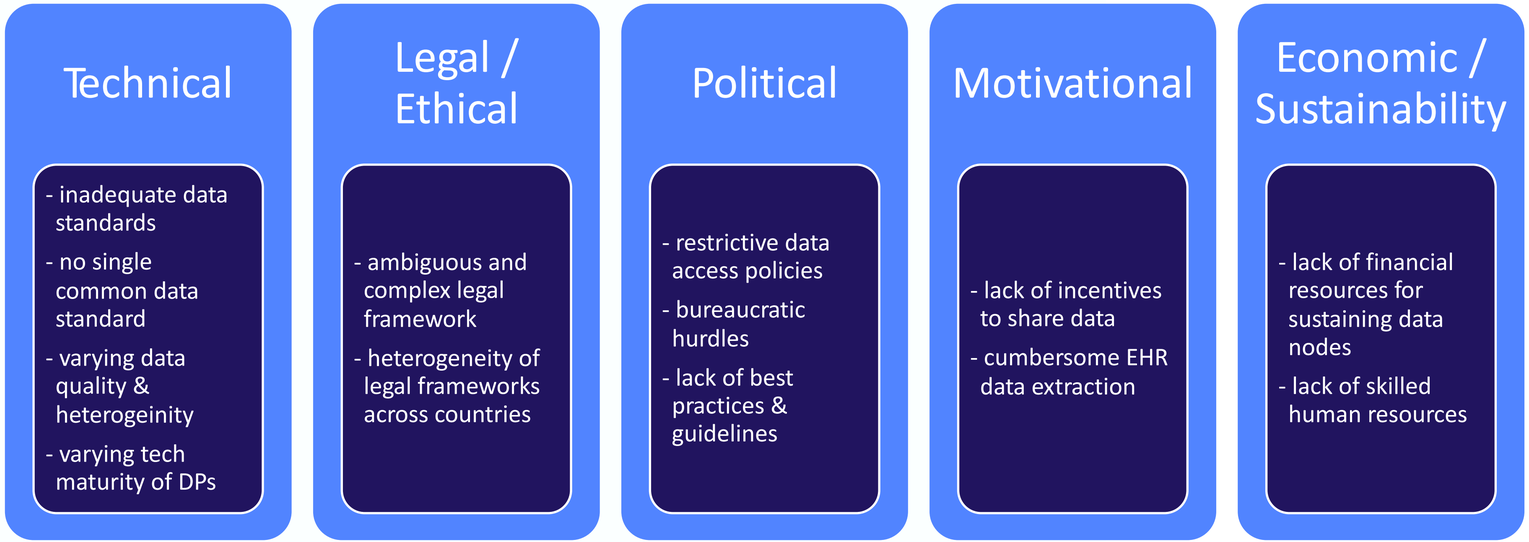
Benefits of data sharing

Transparency	**Maximising the utility and impact of the data collected**	**Making collaboration easier**	**Research acceleration**
Reproducibility	**Data citation & credit**	**Long-term data preservation**	**Meeting requirements of funding and publications**

*based on ODAP The Open Data Assistance Program at Harvard (https://projects.iq.harvard.edu/odap/benefits-sharing-data), https://www.ccdc.cam.ac.uk/Community/depositastructure/cif-deposition-guidelines/benefits-of-data-sharing/

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Challenges of data sharing

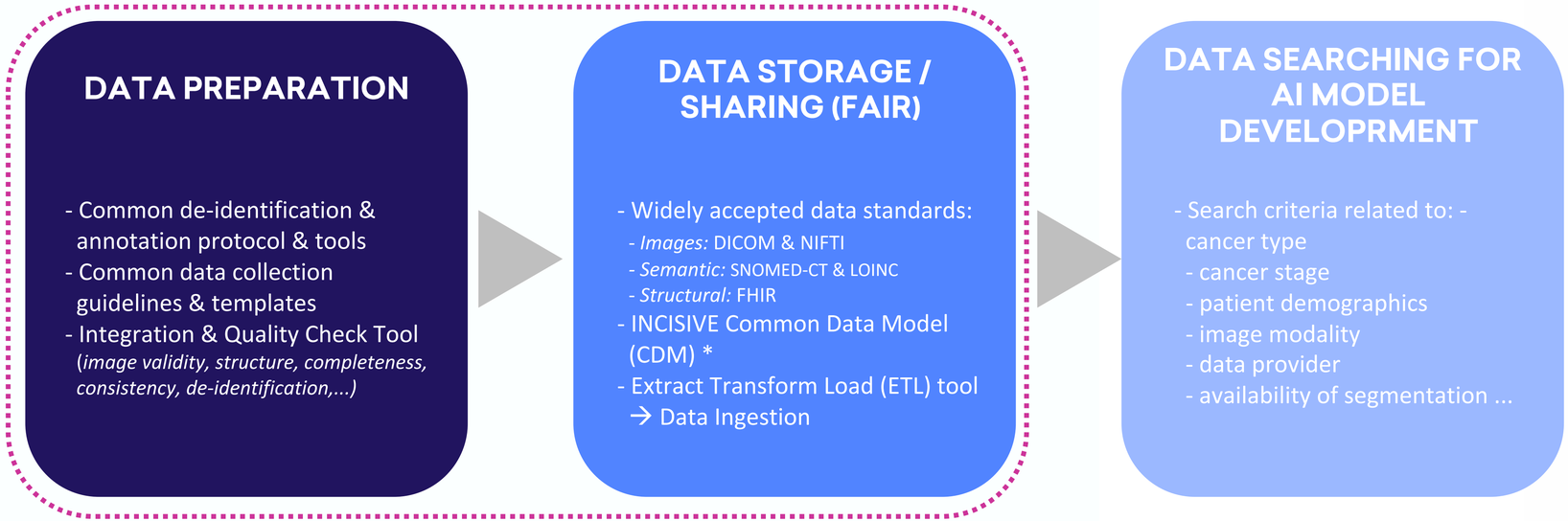


*categorization based on Sane, J. and Edelstein, M. (2015). **Overcoming Barriers to Data Sharing in Public Health: A Global Perspective.**

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Technical challenges

- 3 data management layers for data integration, homogenisation and interoperability
- Lots of technical support to Data Providers!

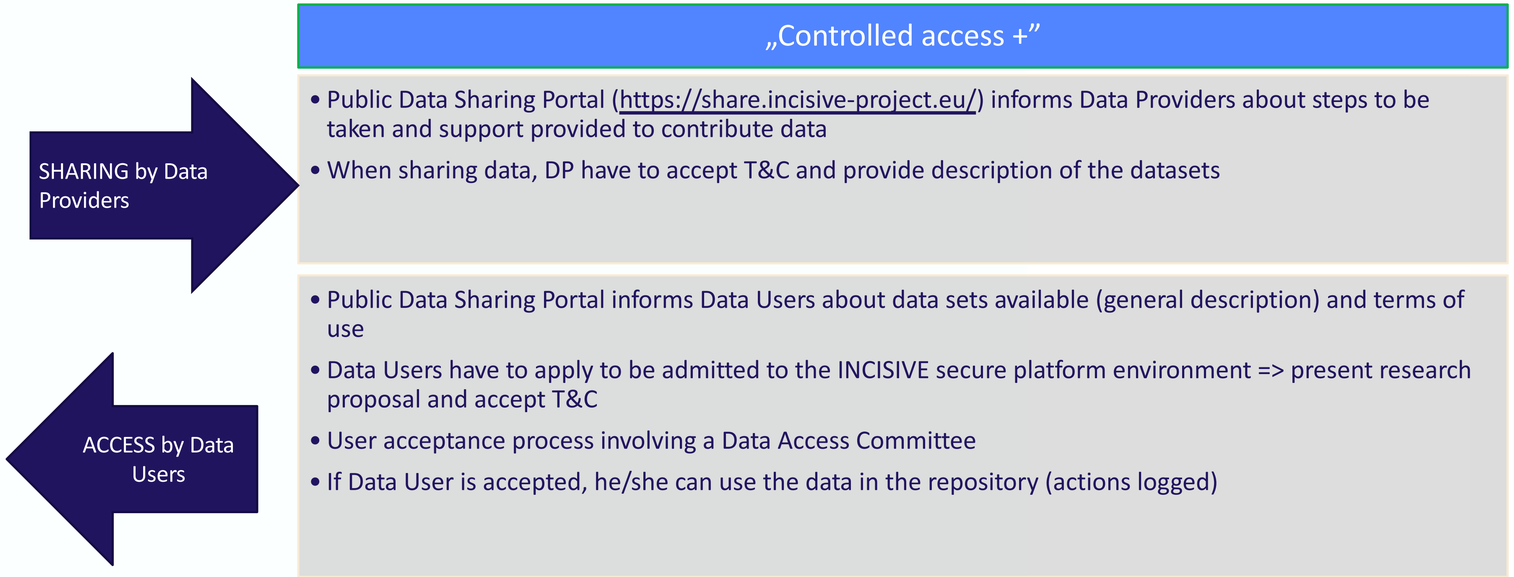


* https://simplifier.net/guide/FHIR-Bundle-Guide/HOME?version=current

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Legal/Ethical/Political Challenges

INCISIVE DATA SHARING FRAMEWORK



These projects has received funding from the European Union’s Horizon 2020 research and innovation programme

INCISIVE Data Sharing Framework

INCISIVE data infrastructure → meets

facilitates data holders to share medical images & health data with minimum effort

LEGAL/ETHICAL REQUIREMENTS

- GDPR-compliant data governance and access management framework (hybrid repository)
  - Data Sharing Agreement, DPIA
  - Data Providers in control of their data
  - Technical and Operational Measures for data protection
    - (Cyber)Security measures
    - User authorisation
    - Logging of data usage actions
    - Federated Learning mechanisms

RESEARCHERS NEEDS

- Data re-usability by researchers, incl. AI developers
- Multi-centre, multi-country data sources
- Standardised, interoperable, quality-checked data

https://share.incisive-project.eu/

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Motivational/Economic Challenges

- Data usage & citation requirement
- Opportunities for networking
- Opportunities for further funding
- Lots of technical support to reduce workload related to data sharing
- Training and capacity-building on various data sharing aspects
- Awareness-raising on benefits of data sharing
- Further motivation required…

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

INCISIVE achievements

AI toolbox: AI models & pipelines, incl. FL mechanism

4 cancer types

Hybrid Data Repository infrastructure

>3.7M cancer images

9.3k de-identified patients

Data interoperability framework (methodology & tools, incl. CDM)

9+ data providers

9/9 data providers integrated / federated repository

Data sharing framework (methodology & tools)

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

INCISIVE Sustainability

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Stay tuned!

Subscribe to our newsletter

incisive-project.eu

Follow us:

@incisiveEu

@incisiveEu

INCISIVE (join our private group)

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

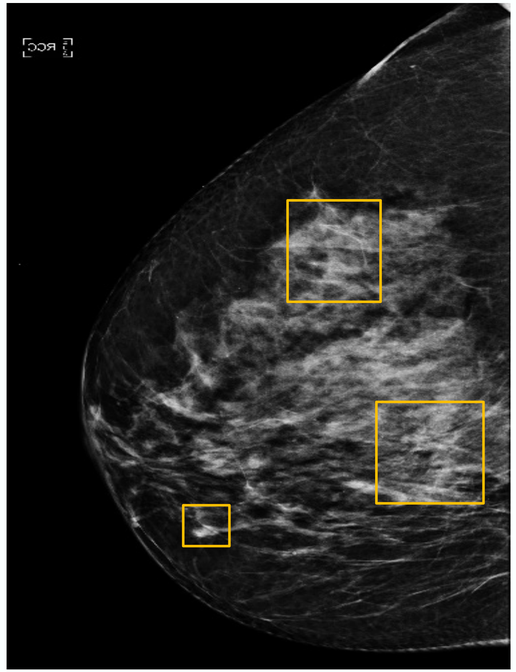
INCISIVE | EUCAIM

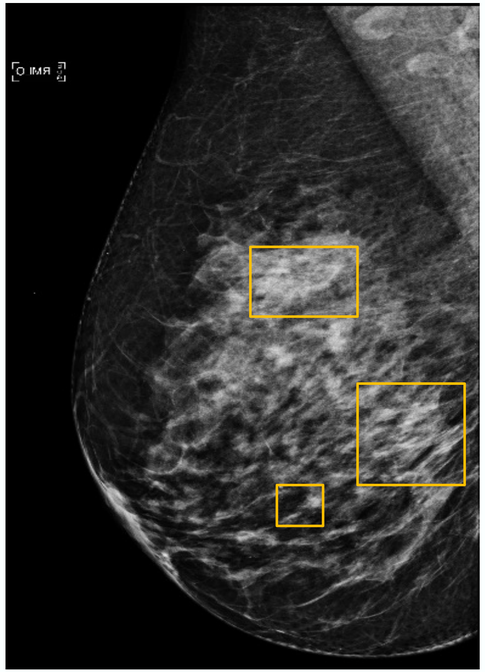
Mr. Oliver Díaz

University of Barcelona, RadioVal project.

“AI development and validation based on FUTURE-AI: the RadioVal case”

Would you TRUST this CAD?



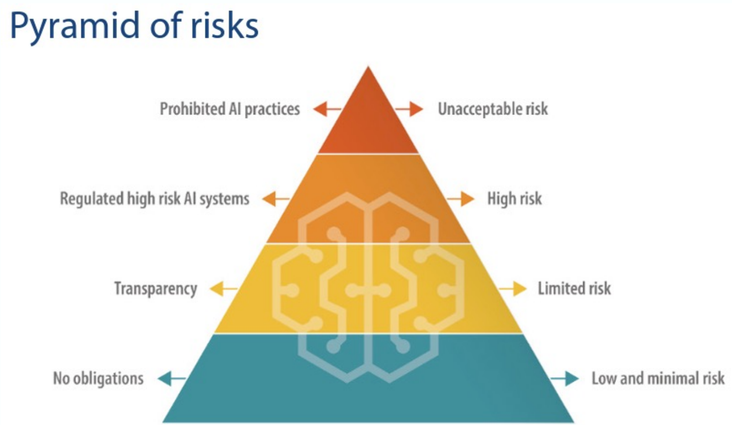


48yo woman. Asymptomatic. Heterogeneously dense breast (ACR C).
Courtesy of Dr. Javier Del Riego – Hospital Universitari Parc Taulí (Sabadell, Barcelona).

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Efforts from the European Commission

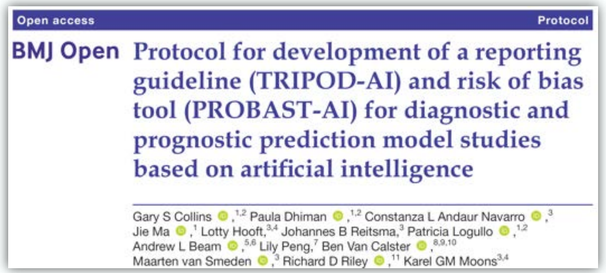
AI Act (under discussion)



(July 2020)

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

International initiatives on AI trustworthiness



Open access Protocol

BMJ Open

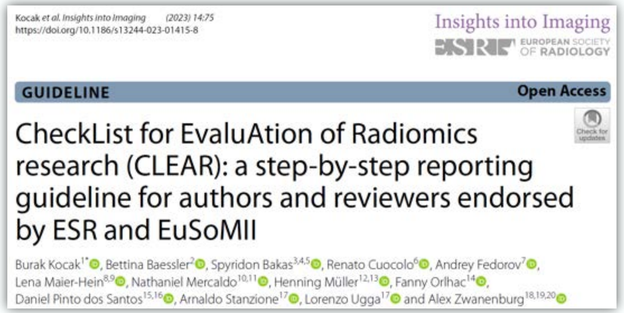
Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence

Gary S Collins, Paula Dhiman, Constanza L Andaur Navarro, Jie Ma, Lotty Hooft, Johannes B Reitsma, Patricia Logullo, Andrew L Beam, Lily Peng, Ben Van Calster, Maarten van Smeden, Richard D Riley, Karel GM Moons

Radiology: Artificial Intelligence EDITORIAL

Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers

John Mongan, MD, PhD • Linda Moy, MD • Charles E. Kahn, Jr, MD, MS



Kocak et al. Insights into Imaging (2023) 14:75
https://doi.org/10.1186/s13244-023-01415-8

Insights into Imaging

ESR EUROPEAN SOCIETY OF RADIOLOGY

GUIDELINE Open Access

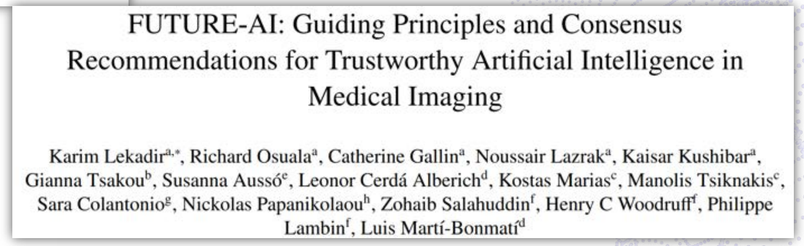
CheckList for EvaluAtion of Radiomics research (CLEAR): a step-by-step reporting guideline for authors and reviewers endorsed by ESR and EuSoMII

Burak Kocak, Bettina Baessler, Spyridon Bakas, Renato Cuocolo, Andrey Fedorov, Lena Maier-Hein, Nathaniel Mercaldo, Henning Müller, Fanny Orlhac, Daniel Pinto dos Santos, Arnaldo Stanzione, Lorenzo Ugga and Alex Zwanenburg

AAPM SCIENTIFIC REPORT MEDICAL PHYSICS

AAPM task group report 273: Recommendations on best practices for AI and machine learning for computer-aided diagnosis in medical imaging

Lubomir Hadjiiski | Kenny Cha | Heang-Ping Chan | Karen Drukker | Lia Morra | Janne J. Näppi | Berkman Sahiner | Hiroyuki Yoshida | Quan Chen | Thomas M. Deserno | Hayit Greenspan | Henkjan Huisman | Zhimin Huo | Richard Mazurchuk | Nicholas Petrick | Daniele Regge | Ravi Samala | Ronald M. Summers | Kenji Suzuki | Georgia Tourassi | Daniel Vergara | Samuel G. Armato III



FUTURE-AI: Guiding Principles and Consensus Recommendations for Trustworthy Artificial Intelligence in Medical Imaging

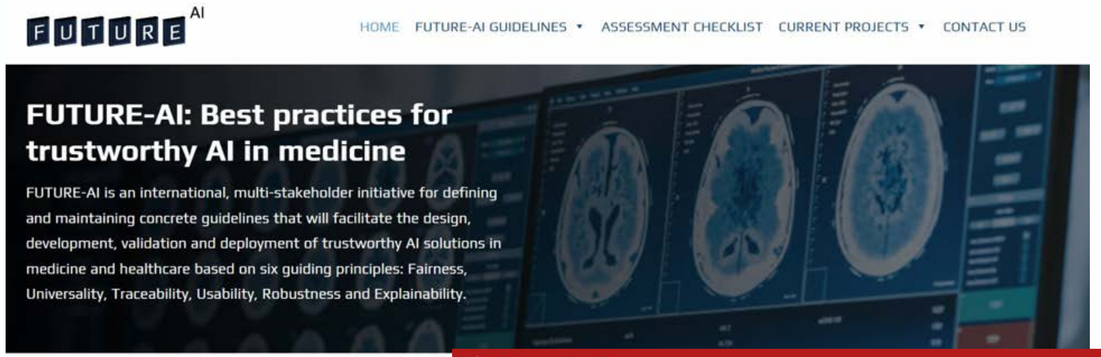
Karim Lekadir, Richard Osuala, Catherine Gallin, Noussair Lazrak, Kaisar Kushibar, Gianna Tsakou, Susanna Aussó, Leonor Cerdá Alberich, Kostas Marias, Manolis Tsiknakis, Sara Colantonio, Nickolas Papanikolaou, Zohaib Salahuddin, Henry C Woodruff, Philippe Lambin, Luis Martí-Bonmatí

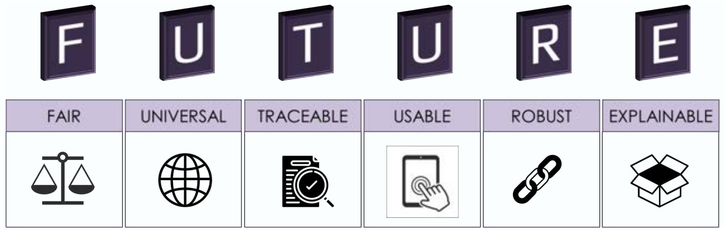
Metrics Reloaded: Recommendations for image analysis validation

LENA MAIER-HEIN, German Cancer Research Center (DKFZ), Germany, Heidelberg University, Germany, and National Center for Tumor Diseases (NCT), Germany
ANNIKA REINKE, German Cancer Research Center (DKFZ), Germany and Heidelberg University, Germany

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

FUTURE-AI (initiative from AI4HI network)





www.future-ai.eu

K. Lekadir *et al.* 2023. arXiv: 2309.12325

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

RadioVal case

radioval

Funded by the European Union

International Clinical Validation of Radiomics Artificial Intelligence for Breast Cancer Treatment Planning

radioval.eu

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

RadioVal case

Current models	Performance Metric	• 0.85 AUC • 0.80 F1-Score

Try our self-assessment checklist: https://future-ai.eu/checklist/

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

INCISIVE | EUCAIM

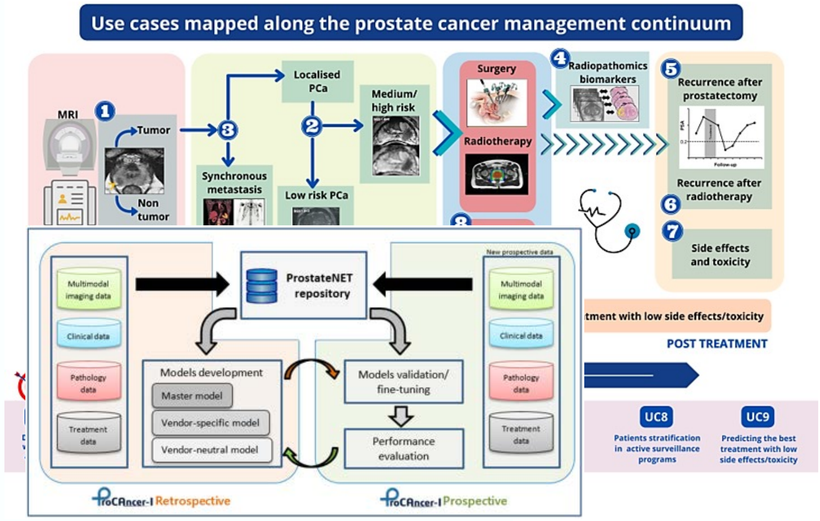
Mr. Kostas Marias
HMU, Head CBML-ICS-FORTH
ProCAncer-I project

ProCAncer-I: Challenges related to multi-center data collection and AI modelling

PROCANCER-I RETROSPECTIVE PROTOCOL:

- To develop AI models in the context of 9 clinical scenarios
- To identify/validate new prognostic and predictive markers in PCa patients

	T2-w	T2-w PI-RADS compliant	DWI	DWI PI-RADS compliant	DCE (optional)	DCE (optional) PI-RADS compliant
acquisition plane	straight axial or oblique axial matching the long axis of the prostate	straight axial or oblique axial matching the long axis of the prostate	same as T2-w	same as T2-w	same as T2-w	same as T2-w
slice thickness	≤ 4 mm	≤ 3 mm and no gap	≤ 4 mm	≤ 4 mm and no gap	≤ 4 mm	≤ 3 mm and no gap
in-plane dimension	≤1.2 mm (phase) x ≤1.2 mm (frequency)	≤0.7mm (phase) x ≤0.4mm (frequency)	≤2.5mm phase and frequency	≤2.5mm phase and frequency	≤2.0 mm x ≤2.0 mm	≤2.0 mm x ≤2.0 mm
b-values (for ADC calculation)	-	-	≤ 100 s/mm² (low) ≤ 1,400 s/mm² (high)	50-100 s/mm² (low) 800-1,000 s/mm² (high)	-	-
temporal resolution	-	-	-		≤ 15 s (preferably)	≤ 15 s (mandatory)
observation time	-	-	-		> 2 min	> 2 min
dose and injection rate	-	-	-		0.1 mmol/kg and 2-3 cc/s	0.1 mmol/kg and 2-3 cc/s



ProCAncer-I

The ProCAncer-I data collection protocols

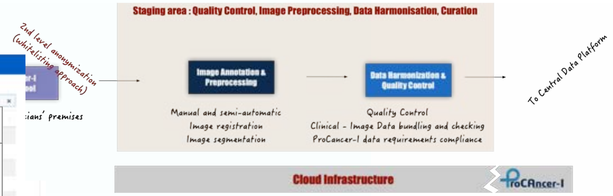
These projects has received funding from the European Union's Horizon 2020 research and innovation programme

ProCAncer-I: Challenges related to multi-center data collection and AI modelling

ProCAncer-I has placed particular emphasis on data curation and especially on quality enhancing mechanisms.

Metadata Repository Quality Check Tool

Type of alert	Description
NOTICE	1. Biopsy is positive but gleason<3+3 2. Index lesion has PI-RADS<=2, although there are other reported lesions with >=3 3. Negative patients with really high PSAs
LOW IMAGES	DICOM series with less than 10 images (per series)
EXCESS DATA	DICOM studies with more than 3500 images
INCOMPLETE	Patients with less than 3 series (since T2, ADC and high b-value are the minimum requirements). Less than 3 means that at least 1 series is missing, possibly failed to upload
ORPHAN	DICOM study with no uploaded metadata. This means that there is a study in the DICOM browser, but there is no uploaded e-crf form accompanying this study.
WARNING	1. Series with suspicious secondary captures (e.g. screenshots) 2. Series has only 1 image (they might contain sensitive patient information)



Data Curation Tool: Guides the user through a set of steps for the curation: motion correction, co-registration, quality check, and approval and storage of the derived images, or rejection, after the manual inspection of the results.

ProCAncer-I quality enhancing technologies

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Major Challenges and Actions Taken

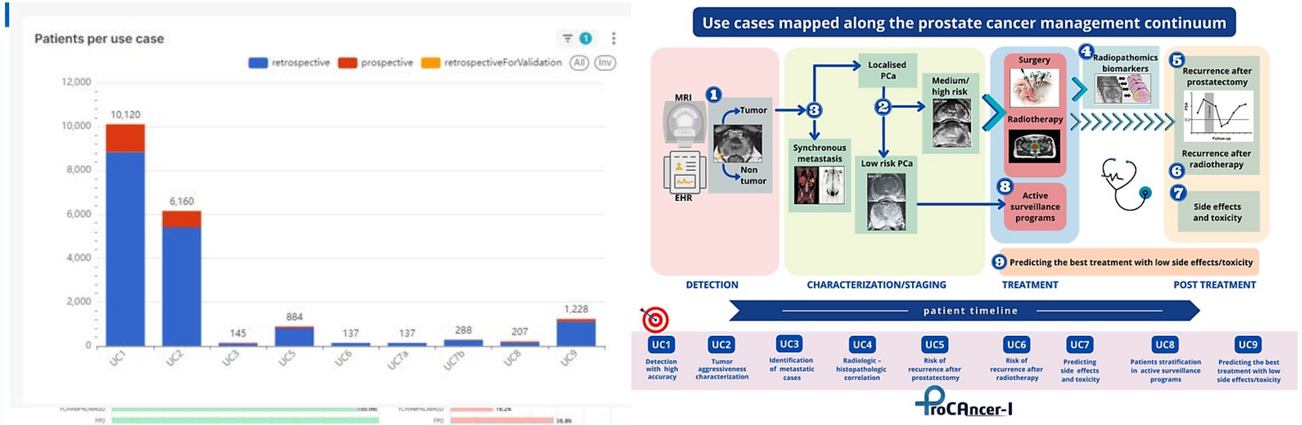
Challenge		Action
Heterogeneity between legal frameworks in different countries/ data providers. **Difficulty** to obtain **Ethics Committee approvals** for clinical data collection with delay to provide requested cases for developing AI models.	➔	Clear and comprehensive **data user agreements** as well as legal experts engagement helping navigate the complexities. **Certain clinical partners were** asked to **increase** their **datasets**, retaining the vendor balance & provide means for **batch uploading** of data for reducing uploading time
No comprehensive imaging model for allowing the registration of imaging metadata, **annotation** and **curation** processes, and used in conjunction with clinical information for cohort creation	➔	**Imaging Extension of OMOP-CDM** to incorporate and standardize image related processes (presented by V. at the European OHDSI Symposium, Rotterdam, Holland, 2023 and accepted for publication at ASCO JCO Clinical Cancer Informatics).
Not standardized clinical and imaging metadata to be used for cohort creation (e.g. sequence type in DICOM tags).	➔	**Developed tools** (e.g. AI model for sequence type identification) and use of standardized vocabularies in the eCRF tool (OMOP vocabularies and RSNA Radlex **standard** terminology)
Delay of **prospective enrollment** due to time limitation collecting the reference standard (especially for negative patients) – not allowed to use MRI examinations before the sign of informed consent.	➔	Data providers were asked to **upload more retrospective cases** to be used for model validation. A **new eCRF upload tool version** was developed to address this group of patients.
Logical discrepancies in **clinical forms** **Secondary image captures** containing sensitive information not removed by the anonymization script	➔	**Developed a metadata quality check tool** using heuristics to identify such issues, and remove secondary captures from the repository.
Low image quality due to **variability** in medical image **acquisition protocols** and presence of **artifacts** (e.g. image blurring, motion artifacts, geometric distortions, signal "graininess")	➔	**Quality assessment:** Prostate Imaging Quality (PI-QUAL) study for developing an automatic tool to assess quality of MRI examinations. **Data cleaning:** Dockerized tools for bias field correction, image denoising, image enhancement, data harmonization.
MRI acquisitions not compliant with the **minimum technical parameters** required to perform reliable quantitative AI modeling.	➔	Acquire a **larger number of prospective** cases using an **MRI compliant** acquisition protocol and upload a subgroup of them for training of the AI models
Inadequate data and **imbalanced datasets** due to missing mandatory information for the reference standard, orphan data.	➔	**Monitoring tool** to identify such issues, **relax constraints** whenever possible, incorporate **external/synthetic datasets**, select proper AI algorithms (DL/radiomics/self-supervising)

ProCAncer-I

The ProCAncer-I major challenges faced

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Data Collection and AI Modeling Challenges



Despite all efforts to a) accelerate data collection, b) ensure/monitor data quality especially in segmentations and c) optimise/adapt AI modeling methodologies, it is not feasible to properly address all the UCs within the lifetime of the project rendering very important the need to extend the project's scope within EUCAIM.

ProCAncer-I AI modeling challenges

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

INCISIVE | EUCAIM

Mr. Ignacio Blanquer
Polytechnic University of Valencia
PRIMAGE and CHAIMELEON projects

“Preserving the legacy and the Virtual Research Environment Concept”

EUCAIM CANCER IMAGE EUROPE

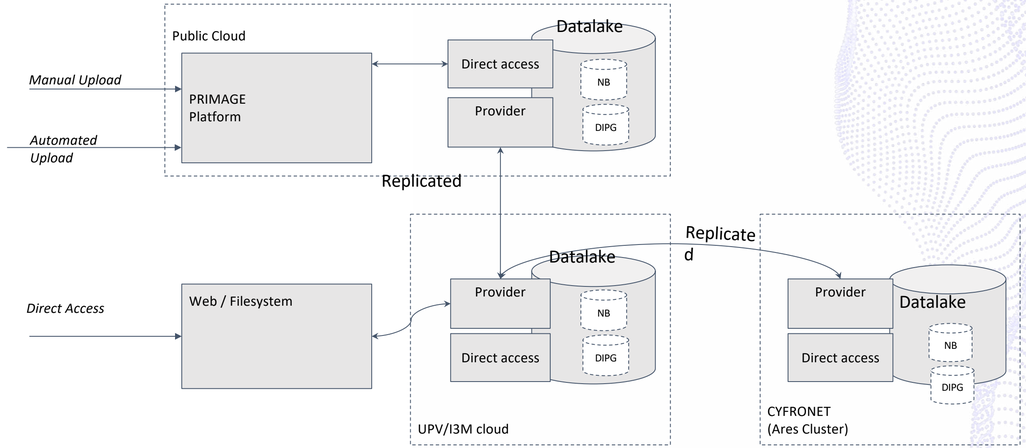
PRIMAGE Medical imaging Artificial intelligence Childhood cancer research

chAIMELEON

PRIMAGE project

The main of PRIMAGE is to build an open cloud-based platform as a Decision Support System (DSS) to assist in the clinical management of two paediatric cancers:

Neuroblastoma (NB), as the most frequent solid cancer of early childhood, and
Diffuse Intrinsic Pontine Glioma (DIPG), being the leading cause of death
related to brain tumours in children.



These projects has received funding from the European Union's Horizon 2020 research and innovation programme

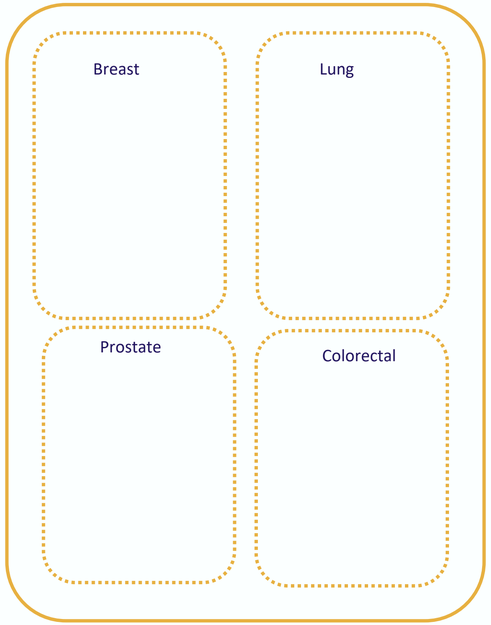
PRIMAGE project

PRIMAGE
Medical imaging
Artificial intelligence
Childhood cancer research

Over 1000 cases of Neuroblastoma and 70 cases of DIPG	A set of tools for segmentation & feature extraction	A distributed processing platform	Two research communities
Need for long-term preservation and compliance to FAIR principles	Need for preservation, annotation, maintenance and execution environments	Need for a sustainable strategy	Need to preserve the cooperation and joint results
Data Transfer & Data Sharing agreements, FAIR-compliant services.	Marketplace and execution environments.	Temporal Federation and further migration into EUCAIM's resources.	Collaboration tools, forum and Joint Research Agreements

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

CHAIMELEON Project



Images + Related clinical data (e-form)

Cloud-based cancer imaging repository as an online resource for the AI community working on the development of cancer management solutions

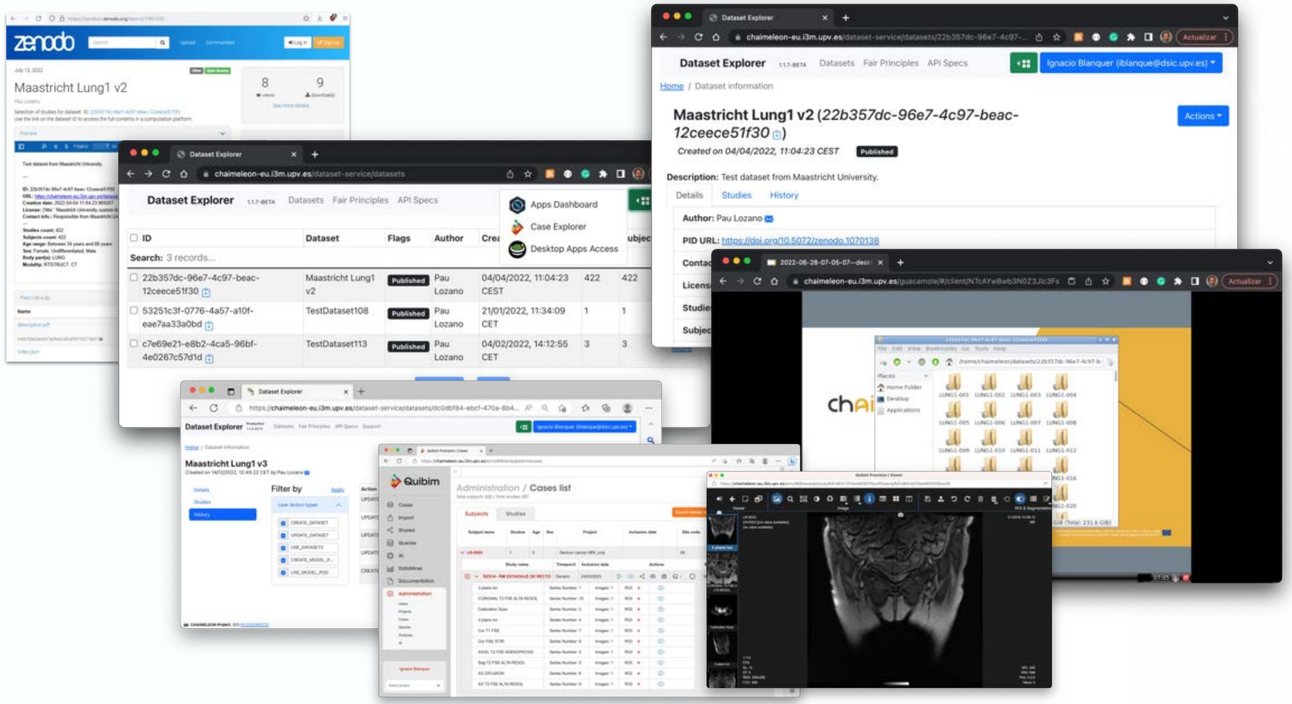
Not just a data warehouse…

- Incorporating all necessary functionalities to allow AI experimentation on the cloud (without downloading the data).
- Powered with automation tools.
- Interoperable with other existing initiatives.

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Virtual Research Environment

https://chaimeleon-eu.i3m.upv.es/dataset-service/datasets



These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Central VRE Model

- It facilitates the traceability of data access.
- Simplified architecture and higher availability.
- Reduced costs due to the centralisation of resources.
- Better user experience.
- Simplifies the development of new applications.
- Benchmarking and validation of tools.

- It requires Data Transfer Agreements and common Data Sharing Agreements.
- May not be compatible with some legal specificities.
- May have issues with pseudonymised data.
- Balance between services used and access delegation.
- Need to homogenise data model.

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Conclusions

- PRIMAGE and CHAIMELEON will end up with a set of joint assets that need to be preserved.
- The federated model is the fastest approach for a smooth transition
  - Progressive adaptation of data from mediators to full compliance.
  - Viable option until the central storage is fully procured.
- The federated model is cumbersome in terms of data exploitation
  - Central storages with isolated partitions can be adequate and efficient.

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

More information

Ignacio Blanquer
Instituto de Instrumentación para Imagen Molecular (I3M)
Universitat Politècnica de València

Twitter: @iblanque

iblanque@dsic.upv.es

github.com/grycap

cancerimaging.eu

These projects has received funding from the European Union's Horizon 2020 research and innovation programme

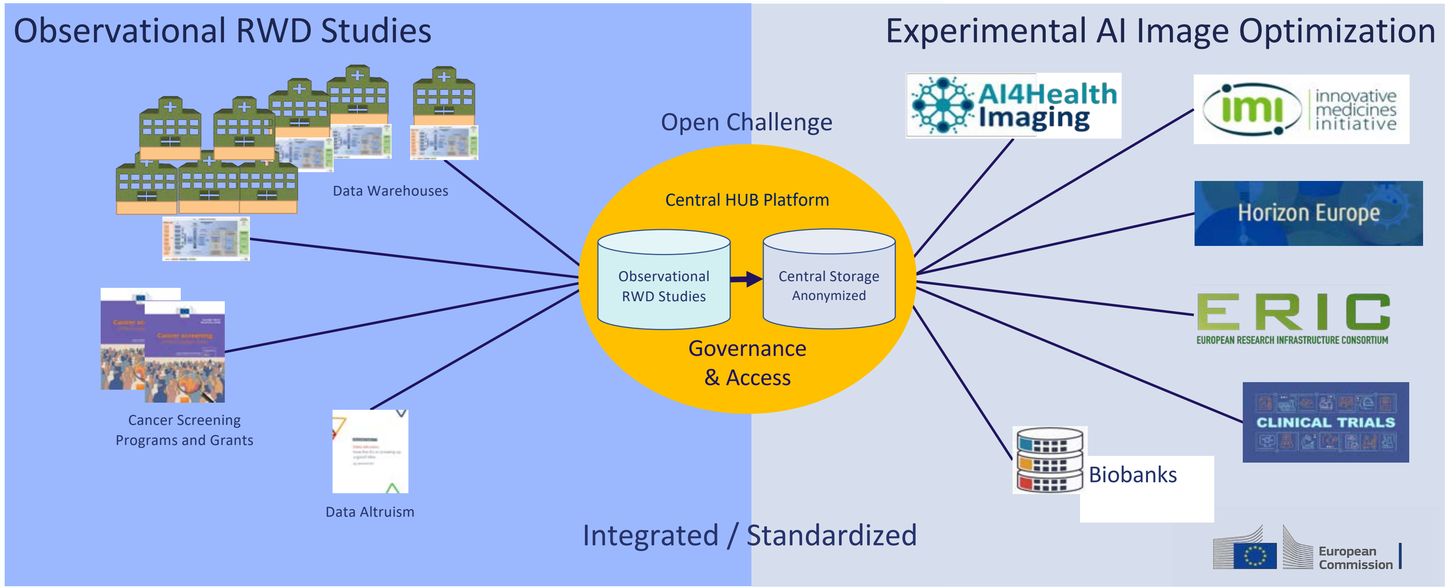
INCISIVE | EUCAIM

Dr. Luis Martí-Bonmatí

La Fe Polytechnic University Hospital

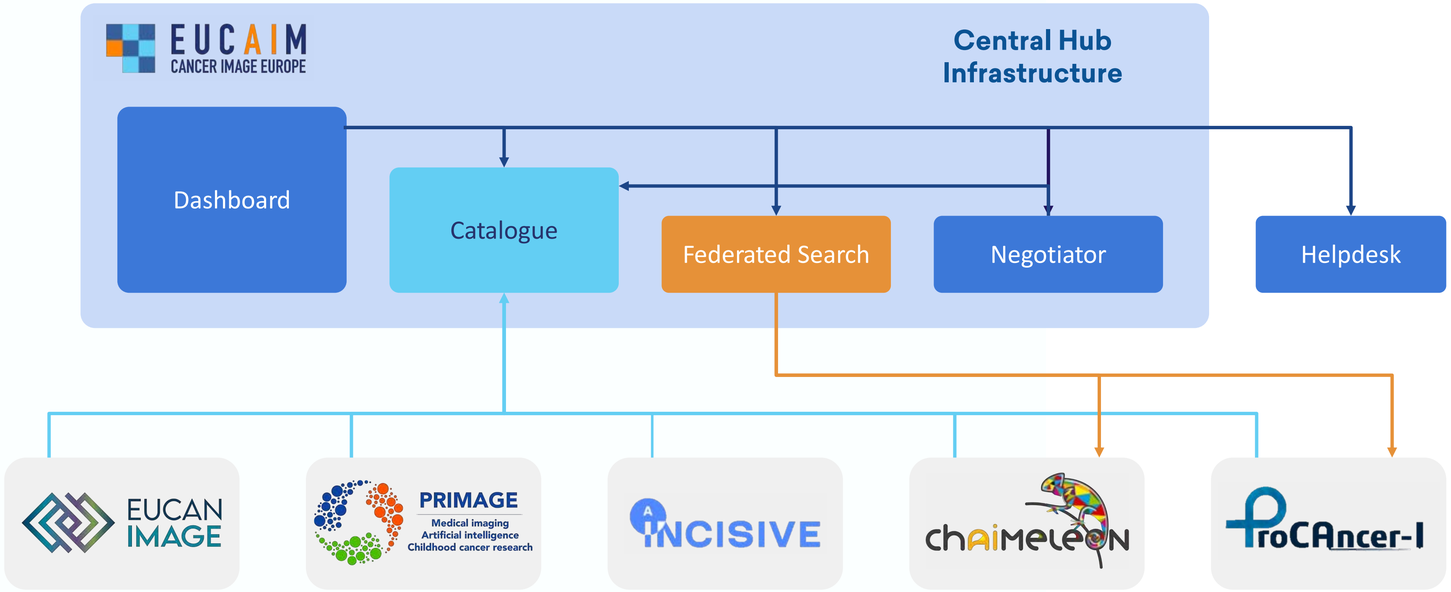
EUCAIM, PRIMAGE and CHAIMELEON projects

Federated Infrastructure with a Central Node
Atlas of Cancer Images



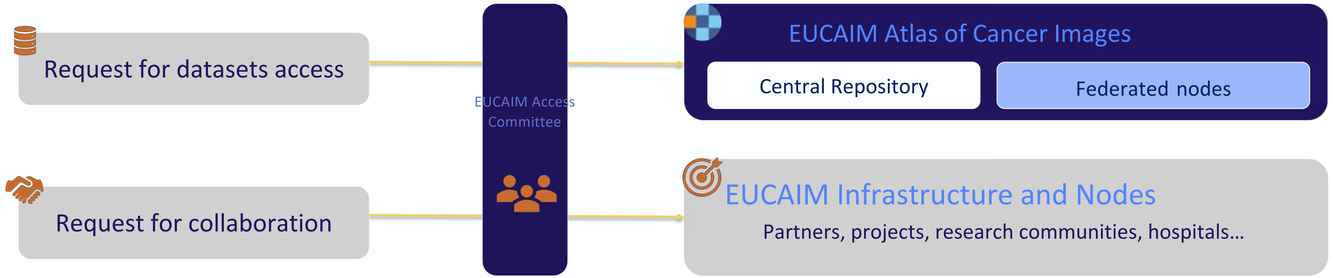
These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Prototype platform

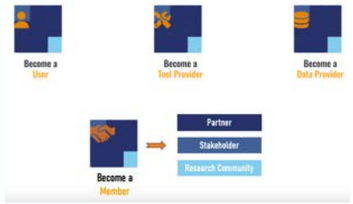


These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Access request process



- Intended users requesting access to datasets
  Researchers and Innovators (development and validation of computational solutions)
- Intended collaborators
  Partners, Stakeholders, Research Communities



These projects has received funding from the European Union's Horizon 2020 research and innovation programme

Sustainability

Main result available beyond the lifetime of EUCAIM

Cancer Image Europe infrastructure

Cancer Image Europe

Introducing the future
European Digital Infrastructure Consortium (EDIC)

Sustainability plan

Currently investigating MCP options

- Primarily focusing on efforts to become an EDIC (already 7 countries EoI).
- Dedicated Working Group has been created (Spanish coordination).
- Your help in getting engaged national representative could be key.
- Communication channel with GDI.

European Genomic Data Infrastructure

These projects has received funding from the European Union's Horizon 2020 research and innovation programme