Project title: European Federation for Cancer Images

Project acronym: EUCAIM

Grant Agreement: 101100633

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

D6.3. EUCAIM Federated Analysis Toolbox

Responsible partner:	BSC
Author(s):	Carles Hernandez-Ferrer (BSC), Josep Lluis Gelpí (UB), Salvador Capella (BSC)
Contributor(s):	Celia Martin (QUIBIM)
Reviewer(s):	Nicolas Georges (CNRS), Francesco Cremonesi (INRIA)
Date of delivery:	30.06.2025
Version:	V1
Due date:	Month 30
Type:	Report
Dissemination level:	Public

Table of contents

1. Introduction

1.1 Document Purpose

The EUCAIM project aims to establish a robust and privacy-respecting infrastructure for cancer image analysis using federated learning across Europe. As part of this effort, WP6 is tasked with developing and maintaining a Federated Analysis Toolbox that provides users and developers with a curated, discoverable collection of tools that are compatible with the federated analysis environment.

This deliverable, D6.3, documents the design, implementation strategy, and current status of the EUCAIM Federated Analysis Toolbox. It outlines the rationale behind choosing bio.tools as the main registry for tools, describes the process and strategy for populating the collection, and presents the alignment of this toolbox with FAIR principles and EUCAIM's broader interoperability goals.

1.2 Document Scope

The purpose of this document is to:

- Describe the EUCAIM Federated Analysis Toolbox as a key component of the project infrastructure.
- Present the strategic decision to adopt bio.tools as the central registry for all tools.
- Report on the current population status of the toolbox.
- Evaluate the FAIRness of registered tools through the OpenEBench Tools Observatory.
- Propose short-term actions (e.g., contentathons) to improve completeness and quality of registration.

The document is self-contained but provides direct links to external resources and references (e.g., the UPV guide for software providers) for more detailed implementation guidelines.

2. bio.tools: Description and Role in EUCAIM

2.1 What is bio.tools?

bio.tools is a curated, community-driven registry of software tools and data services for life sciences. Developed under the ELIXIR infrastructure, it provides a standardized framework for describing and discovering bioinformatics resources. Each entry in bio.tools includes metadata about the tool's function, inputs and outputs, domain of application, licensing, usage, and links to documentation and source code.

Key benefits of bio.tools include:

- Support for the biotoolsSchema, a formalized metadata model.
- Interoperability with external resources like EDAM ontology.
- Integration into tool monitoring and benchmarking systems like OpenEBench.
- Long-term sustainability under the ELIXIR umbrella.

2.2 Why bio.tools was chosen

After discussions across WP5 and WP6, and following milestone M6.2, the project adopted bio.tools as the official registry for the EUCAIM Federated Analysis Toolbox. The rationale includes:

- Discoverability: Public registration makes tools visible to the broader research community.
- Standardization: Ensures consistent metadata and enables tool interoperability.
- Compatibility with FAIR: Tools registered in bio.tools are automatically monitored through OpenEBench for FAIRness metrics.
- Linking to federated infrastructure: Future plans include associating each tool entry with its deployment (e.g., Docker image in Harbor or runtime environment).

The official EUCAIM collection within bio.tools can be accessed at:

https://bio.tools/t?collectionID=%22EUCAIM%22

2.3 Current status of EUCAIM collection

As of June 2025, the EUCAIM collection in bio.tools includes **48 tools**, out of **80 total**[1] tools tracked internally. However, many of the registered tools still contain minimal metadata or lack direct links to execution environments.

The next steps involve:

- Ensuring all tools from the internal list are registered in bio.tools.
- Completing metadata for tools with partial registration.
- Linking each tool with its binary/container location in UPV's Harbor.

3. Overview of the EUCAIM Federated Analysis Toolbox

3.1 Strategy and Infrastructure

The EUCAIM Federated Analysis Toolbox is designed as a dynamic and evolving catalog of software tools that support cancer image analysis in a privacy-preserving federated environment. Rather than being a single, monolithic software system, the toolbox acts as a distributed infrastructure layer — a reference point for tools that are compatible with the data federation strategy and analysis workflows developed across EUCAIM.

Key characteristics of the toolbox include:

- **Decentralization**: Tools may be deployed across multiple nodes or accessed via containers, allowing local execution where data resides.
- **Interoperability**: Tools are described using standard metadata (biotoolsSchema), EDAM ontology, and linked to Docker images, source code, or APIs when available.

---

[1] EUCAIM_Pre_and_Post_processing_tools_and_software_short_list:
EUCAIM_Pre_and_Post_processing_tools_and_software_short_list.xlsx

- **Transparency and discoverability**: The toolbox is publicly visible through bio.tools, increasing the reach and usability of the tools beyond the project consortium.
- **Alignment with FAIR and ELIXIR standards**: The use of bio.tools ensures long-term compatibility with broader European bioinformatics infrastructures.

The toolbox includes tools developed directly within EUCAIM, as well as external or pre-existing tools adopted for federated workflows (e.g., segmentation, annotation, harmonization, model training, or quality control).

Eventually, this toolbox will support deployment through containerization (via UPV's Harbor registry), integration into workflow engines, and harmonization with EUCAIM's future analysis pipelines.

3.2 Tool Registration Workflow

The process of populating the toolbox follows a semi-structured workflow agreed upon by WP6 and under discussion with WP4 and WP5:

1. Tool Identification
    - Tools are proposed by technical partners, use case leaders, or developers.
    - Internal tracking begins using a shared spreadsheet with a few metadata fields.
2. Initial Registration in bio.tools
    - A minimal profile is created on bio.tools with at least: tool name, homepage, and functional description.
    - Tools follow a self risk assessment and provide required metadata using a Google form.
3. Quality Monitoring
    - Once registered, each tool's FAIR metrics are automatically tracked via the OpenEBench Tools Observatory.
4. Curation Cycles (*future work*)
    - Periodic review sessions (e.g., contentathons) are organized to fill gaps in registration and metadata.

This workflow allows the toolbox to be progressively populated and improved, ensuring that tools are discoverable, executable, and aligned with EUCAIM's federated architecture requirements.

4. OpenEBench Tools Observatory and FAIR Metrics

4.1 FAIR Principles Overview

The FAIR principles—Findability, Accessibility, Interoperability, and Reusability—serve as a guideline to enhance the management and stewardship of scientific data and software. Applying these principles ensures that digital assets are:

- **Findable**: Easily located by humans and machines through standardized metadata and persistent identifiers.
- **Accessible**: Retrievable via standardized protocols, potentially with authentication and authorization procedures.

- **Interoperable**: Compatible with other data and tools through the use of shared vocabularies and standards.
- **Reusable**: Suitably described and licensed to allow replication and combination in different settings.

In the context of EUCAIM, adherence to the FAIR principles is crucial for fostering collaboration, ensuring reproducibility, and maximizing the utility of the federated analysis tools across diverse research environments.

4.2 How OpenEBench Supports FAIRification

OpenEBench is the ELIXIR platform dedicated to benchmarking and technical monitoring of bioinformatics tools, web servers, and workflows. It plays a pivotal role in evaluating and promoting the FAIRness of research software. Key functionalities include:

- **Automated Monitoring**: OpenEBench systematically assesses tools registered in bio.tools, evaluating aspects such as metadata completeness, licensing, and documentation.
- **FAIRness Assessment**: It provides metrics that reflect the degree to which tools adhere to FAIR principles, offering insights into areas needing improvement.
- **Integration with bio.tools**: By leveraging the standardized metadata from bio.tools, OpenEBench ensures that assessments are consistent and comprehensive.
- **Community Engagement**: OpenEBench facilitates community-driven benchmarking efforts, encouraging best practices in software development and documentation.

Through these mechanisms, OpenEBench aids EUCAIM in identifying gaps in tool descriptions and guiding enhancements to meet FAIR standards.

4.3 Current FAIR Status of EUCAIM Tools

As of June 2025, the EUCAIM collection in bio.tools includes **48 tools**, representing **60.00%** of the total tools tracked internally (see previous footnote as " EUCAIM Pre and Post processing tools and software short list").

The current FAIRness breakdown is as follows (see annex table A2 for criteria):

- 100.00% of tools provide a description of their functionality.
- 37.50% include a link to detailed documentation.
- 29.17% specify the operations they perform on data.
- 43.75% include a license declaration.

These indicators show a solid foundation for FAIR alignment, particularly regarding discoverability and licensing. However, there are clear areas for improvement, especially in documentation, code availability, and linkage to execution environments. These will be specifically addressed in upcoming contentathon sessions and targeted curation efforts coordinated by WP5 and WP6.

5. Gap Analysis: Registered vs. Non-Registered Tools

To assess the coverage and completeness of the EUCAIM Federated Analysis Toolbox, we conducted a cross-referencing analysis between the internal EUCAIM tool tracking spreadsheet and the EUCAIM collection in bio.tools. This analysis helps quantify both the registration status of tools and the quality of their metadata, forming the basis for ongoing improvement efforts.

5.1 Registration Coverage

As of June 2025:

- The internal tool tracking sheet contains a total of 80 tools.
- 48 tools are registered in bio.tools, representing 60.00% of the total.

This shows that while the registration process has begun, a majority of tools remain either unregistered or pending review.

Status	Count	Percentage
Total tools tracked	80	100%
Tools registered	48	60%
Tools not yet registered	32	40%

Note: *This deliverable will be updated until the end of the EUCAIM project to reflect the dynamically changing number of registered tools.*

5.2 Metadata Completeness of Registered Tools

While 48 tools have been registered in bio.tools, their metadata quality varies significantly, ranging from minimal entries to fully FAIR-aligned profiles. For the purpose of this gap analysis, we categorized the registered tools into two levels based on the presence of essential metadata:

Metadata Completeness Level	Description	Count	% of Registered
Minimal	Only basic metadata (name, brief description, and version).	42	87.50%
Partial	Includes description + at least 1 key element (license, repo, docs, etc.)	33	68.75%

This classification states that we need a high level of curation to make sure that the 12.50% of software not included in the previous table reaches the minimal level of metadata completeness. Moreover, we need to push software contributors to move the 42 tools from a minimal level to partial level fast.

These findings will help prioritize efforts in the upcoming contentathon and partner onboarding activities.

We plan on defining a “complete” entry as one that includes at least (under consideration):

- Description of the tool’s function
- License declaration
- Supported operations and file formats
- Links to documentation and source code

- Link to an execution environment (e.g., Harbor)

The key element of a “complete” entry is the link to the UPV’s Harbor, since it will become the entry point of any execution in the EUCAIM ecosystem.

5.3 Next Steps

To improve the registration and FAIR maturity of the EUCAIM toolbox:

- WP5 and WP6 will organize contentathon sessions to support partners in registering and completing tool entries.
- A validation checklist will be provided to ensure metadata completeness, including links to containers, code, formats, and operations.
- The integration with OpenEBench FAIR metrics will be continuously monitored to track progress and highlight areas for targeted improvement.

6. Contentathon Strategy

In order to accelerate the registration and FAIRification of tools in the EUCAIM Federated Analysis Toolbox, the project will organize a series of collaborative contentathon sessions. These short, focused events will bring together developers, tool providers, and WP5/WP6 curators to improve the completeness and quality of tool metadata in bio.tools.

6.1 Objectives

The contentathons aim to:

- Increase the number of tools registered in bio.tools.
- Improve the metadata quality of existing entries (toward FAIR alignment).
- Ensure consistent linkage to relevant assets: documentation, source code, container images.
- Familiarize partners with the bio.tools and OpenEBench ecosystems.
- Coordinate with containerization and deployment efforts (e.g., Harbor).

6.2 Planning and Format

The first contentathon is scheduled to take place within the next 15–20 days (June 2025), jointly organized by WP5 and WP6. The format will be lightweight and partner-friendly:

- **Target group**: Tool developers and technical leads from EUCAIM partner institutions.
- **Format**: Remote, live session (90–120 minutes) with breakout support.
- **Support materials**: Templates, examples, and direct editing access to the tracking sheet.
- **Resources**:
    - EUCAIM’s Handbook
    - bio.tools registration guide.
    - UPV’s Software Providers Guide.
    - Overview of Harbor and containerization expectations.
    - FAIR metrics dashboard in OpenEBench.

Additional sessions will be planned as needed, with later contentathons focused on packaging and deployment aspects, particularly linking tools to containers and ensuring executable environments are publicly accessible via the UPV Harbor.

7. Next Steps and Conclusions

The creation of the EUCAIM Federated Analysis Toolbox — hosted in bio.tools and monitored via OpenEBench — represents a significant step toward building an interoperable, FAIR-aligned ecosystem of cancer analysis tools across Europe.

The initial population of the toolbox has highlighted both strengths and gaps:

- A growing collection of registered tools (48 to date).
- High compliance with functional descriptions and licensing.
- Gaps in documentation, source code, and containerization linkages.

The next phase will focus on:

- Expanding registration to cover all internally tracked tools.
- Improving metadata completeness through targeted contentathons.
- Linking tools to Docker images in Harbor to enable reproducible deployment.
- Continuously monitoring FAIR maturity using OpenEBench indicators.

Longer term, this toolbox will serve as the gateway to EUCAIM-compatible analysis workflows. It will also support sustainability by aligning with European infrastructures like ELIXIR and EOSC, ensuring visibility and interoperability beyond the lifespan of the project.

Appendix

A1. References

- bio.tools EUCAIM collection https://bio.tools/t?collectionID=%22EUCAIM%22
- OpenEBench Tools Observatory https://openebench.bsc.es/observatory
- UPV Software Providers Guide https://eucaim.gitbook.io/enduserguide/5-userguideswproviders
- UPV Harbor Registry https://harbor.eucaim-node.i3m.upv.es/
- EDAM Ontology (used in bio.tools) http://edamontology.org
- bio.tools main registry https://bio.tools
- biotoolsSchema specification https://github.com/bio-tools/biotoolsSchema
- ELIXIR Tools Platform https://elixir-europe.org/platforms/tools

A2. Short evaluation of tools already submitted into bio.tools

Name	**DES**	**V**	**FN**	**TO**	**DOC**	**LCS**
2D Digital Mammography Harmonization (EUCAIM-SW-046_T-01-03-008)	Y	Y	N	Y	N	N
AITANA	Y	Y	N	Y	Y	Y
Breast Dense Tissue Segmentation	Y	Y	Y	Y	N	N
CT Slice thickness normalization	Y	Y	Y	Y	N	Y
CT-based neuroblastoma tumour detection and segmentation (EUCAIM-SW-022_T-01-02-005)	Y	Y	N	Y	N	N
Cluster based harmonization (EUCAIM-SW-044_T-01-03-006)	Y	Y	N	Y	N	N
DICOM File Integrity Checker (EUCAIM-SW-002_T-01-01-002)	Y	N	N	Y	Y	N
DICOM image similarity-duplicate checker	Y	Y	N	Y	Y	Y
DICOM tags extractor	Y	N	Y	Y	N	Y
DICOM-SEG Annotation	Y	Y	N	Y	Y	Y
Data Ingestion Tool (UPV Reference Node)	Y	Y	N	Y	N	N
Data Integration Quality Check Tool (DIQCT)	Y	N	N	Y	Y	Y
Deep features extraction (EUCAIM-SW-058_T-02-01-007)	Y	Y	N	Y	N	N
Denoising-Inhomogeneity Correction Tool (EUCAIM-SW-015_T-01-01-015)	Y	Y	N	Y	Y	N
Diffusion maps extraction (EUCAIM-SW-056_T-02-01-005)	Y	Y	N	Y	N	N
Digital Research Environment	Y	N	N	N	N	N
EAPP - Extended a Priori Probability tool	Y	Y	Y	Y	N	N
EUCAIM DICOM Anonymizer	Y	Y	Y	Y	N	N

EUCAIM DICOMseg Converter	Y	Y	N	N	N	N
EUCAIM ETL toolset	Y	Y	Y	Y	N	Y
EUCAIM Wizard Tool	Y	Y	Y	N	N	N
FAIR4Health Data Curation Tool	Y	Y	N	Y	Y	Y
Fed-BioMed	Y	Y	Y	Y	Y	Y
MITK	Y	Y	Y	N	Y	Y
ML model for MR series categorisation (EUCAIM-SW-011_T-01-01-011)	Y	Y	N	Y	Y	N
MR-based DIPG tumour detection and segmentation (EUCAIM-SW-020_T-01-02-003)	Y	Y	N	Y	Y	N
MR-based glioblastoma tumour detection and segmentation (EUCAIM-SW-021_T-01-02-004)	Y	Y	N	Y	N	N
MR-based neuroblastoma tumour detection and segmentation (EUCAIM-SW-019_T-01-02-002)	Y	Y	N	Y	Y	N
MRI Image intensity normalization	Y	Y	Y	Y	Y	Y
Mainzelliste	Y	Y	N	Y	Y	Y
NLmCED Filter	Y	N	Y	Y	N	N
NestedComBat	Y	Y	N	Y	Y	Y
Overall survival and treatment response in glioblastoma multiforme (EUCAIM-SW-077_T-02-04-007)	Y	Y	N	Y	N	N
Overall survival in neuroblastoma by GIBI230 (EUCAIM-SW-078_T-02-04-008)	Y	Y	N	Y	N	N
Perfusion maps extraction (EUCAIM-SW-053_T-02-01-002)	Y	Y	N	Y	N	N
QP-Insights Uploader	Y	Y	Y	Y	Y	Y
R1 and T1 maps extraction (EUCAIM-SW-057_T-02-01-006)	Y	Y	N	Y	N	N
R2 maps extraction (EUCAIM-SW-055_T-02-01-004)	Y	Y	N	Y	N	N
Radiatus	Y	Y	N	Y	N	Y
Radiomic features extraction (EUCAIM-SW-054_T-02-01-003)	Y	Y	N	N	N	N
Radiomics based Lung cancer staging	Y	N	N	Y	N	Y
RadiomicsEnabler	Y	Y	N	Y	N	Y
Time Coherence Tool (EUCAIM-SW-001_T-01-01-001)	Y	Y	N	Y	N	N
Trace4Harmonization	Y	Y	N	Y	N	Y
Trace4MedicalImageCleaning	Y	Y	N	N	N	Y
Virtual Imaging Platform	Y	Y	Y	N	Y	Y
Volumetry analysis (EUCAIM-SW-060_T-02-01-009)	Y	Y	N	Y	N	N
openVRE	Y	Y	Y	Y	Y	Y

The following are the meanings of the tags used in the table's header:

- *DESC*. Does the software provide a short description or its purposes?
- *V*. Does the software details its version?
- *FN*. Does the software detail its functions/operations (what the software does as block diagrams)?
- *TO*. Does the software provide tags of topics where it could be used?

- *DOC.* Does the software include a link to detailed documentation?
- *LCS.* Does the software delipcs a license?