Deliverable D7.3

Report on FAIR guidelines followed in the consortium

Project Title	Artificial Intelligence For Image Data Analysis In The Life Sciences
Project Acronym	AI4Life
Project Number	101057970
Project Start Date	01.09.2022
Project Duration	36 Months

WP N° & Title	WP7: Communication, Outreach and Training
WP Leaders	EMBL-BioHub; HT
Deliverable Lead Beneficiary	HT
Dissemination Level	PU
Contractual Delivery Date	31.08.2024 (M24)
Actual Delivery Date	19.08.2024
Authors	Florian Jug, Beatriz Serrano-Solano
Contributors	
Reviewers	Teresa Zulueta-Coarasa, Arrate Muñoz-Barrutia, Dorothea Dörr, Estibaliz Gómez-de-Mariscal, Heba Ibrahim

AI4Life has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement number 101057970. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

www.ai4life.eurobioimaging.eu
ai4life@eurobioimaging.eu
@AI4LifeTeam

Change Log

Version	Date	Author	Description of changes
v0.1	09.08.2024	Florian Jug, Beatriz Serrano-Solano	Initial draft
v0.2	15.08.2024	Teresa Zulueta-Coarasa, Arrate Muñoz-Barrutia, Dorothea Dörr, Estibaliz Gómez-de-Mariscal, Heba Ibrahim	Review and suggestions
v0.3	16.08.2024	Florian Jug, Beatriz Serrano-Solano	Final draft including consortium member contributions approved for submission

Acronyms and Abbreviations

AI	Artificial Intelligence
BMZ	BioImage Model Zoo
DL	Deep Learning
DMP	Data Management Plan
FAIR	Findable; Accessible; Interoperable; Reusable

www.ai4life.eurobioimaging.eu
ai4life@eurobioimaging.eu
@AI4LifeTeam

Table of contents

www.ai4life.eurobioimaging.eu
ai4life@eurobioimaging.eu
@AI4LifeTeam

Executive Summary

The AI4Life project aims to facilitate the use and utility of Artificial Intelligence (AI) in the life sciences by ensuring that data and AI models are Findable, Accessible, Interoperable, and Reusable (FAIR). In this way, the project aligns with the FAIR principles to enhance data management and maximise the impact of trained AI methods to facilitate biological research outcomes. This deliverable provides a comprehensive overview of the guidelines and FAIR resources developed and created by the AI4Life consortium to achieve the consortium's objectives.

1. Objectives of AI4Life

AI4Life seeks to integrate advanced AI methodologies with bioimaging technologies to address critical challenges in the life sciences. By leveraging cutting-edge AI techniques, the project aims to improve data analysis, foster collaboration, and drive innovation across various biological and medical research domains. The primary goals of AI4Life are to:

1. Democratise the availability of AI-based image analysis methods.
2. Simplify model deployment, sharing, and dissemination through a new developer-facing service.
3. Empower common image analysis platforms with AI integration.
4. Establish standards for the submission, storage and FAIR access.
5. Organise Open Calls and Challenges for image analysis problems.
6. Organise outreach and training events, i.e., image analysis courses/workshops and participation in international conferences.

2. Alignment with the FAIR Principles

The FAIR principles serve as the cornerstone of the AI4Life project, guiding the development of data management and sharing practices that ensure longevity and broad usability. The consortium has focused on:

- Findability: Creating robust metadata standards and searchable databases to enable researchers to locate relevant datasets and trained AI models easily.
- Accessibility: Ensuring data and models are readily available through open access platforms while maintaining appropriate integration checks.
- Interoperability: Developing compatible data formats and interfaces that directly integrate existing tools and systems.

www.ai4life.eurobioimaging.eu
ai4life@eurobioimaging.eu
@AI4LifeTeam

- Reusability: Establishing comprehensive documentation and licensing protocols to facilitate data reuse in future research.

3. Description of work

Within the AI4Life consortium, we have, from day 1, established an entirely FAIR mentality. This means that virtually everything we are working on is either *ab initio* openly available according to the FAIR guidelines or will be made publicly available when reaching a publishable state. The AI4Life website (https://ai4life.eurobioimaging.eu/) or the BioImage.IO GitHub community (https://github.com/bioimage-io) are the central access points.

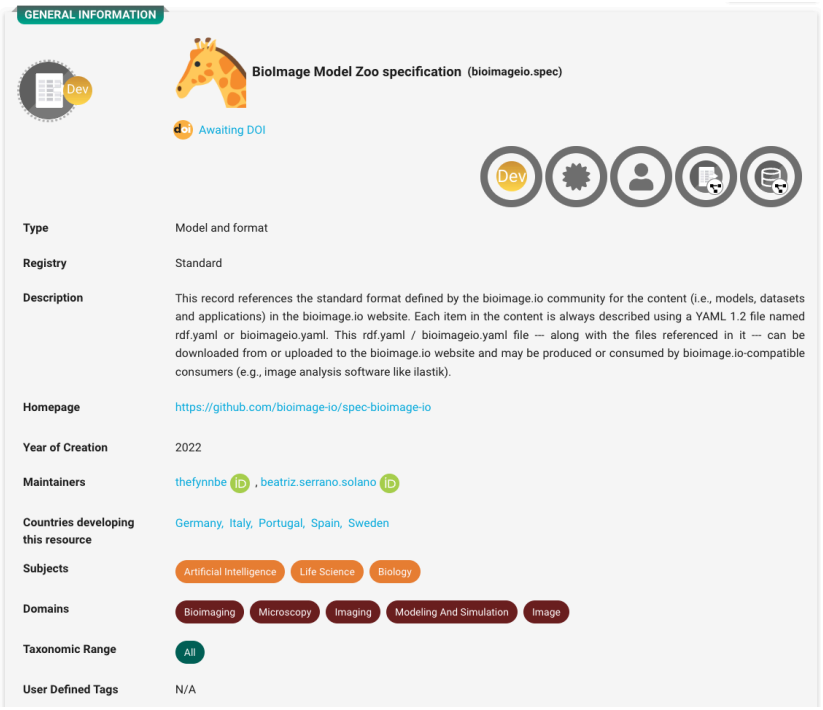
While we strictly adhere to FAIR guidelines in our own work, we are also committed to actively developing and refining these guidelines for the broader community. Our project deliverables and outputs are intentionally crafted to be open and FAIR, ensuring they provide maximum benefit to both our users and the developer communities we support.

In this section, we are, therefore, pointing to both types of outputs, FAIR guidelines and FAIR resources and documents.

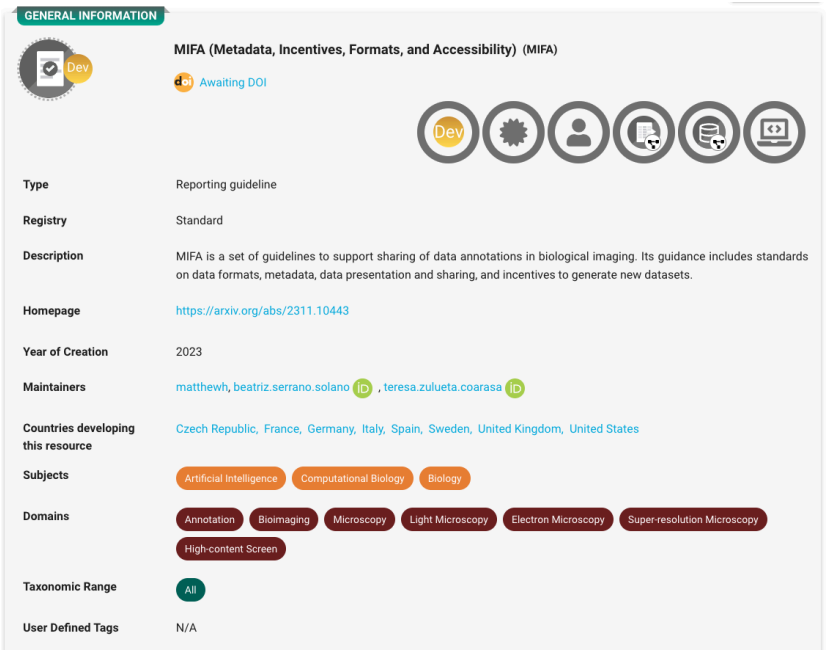
3.1 Guidelines Developed by the Consortium

The AI4Life consortium has collaborated extensively to formulate guidelines that operationalise the FAIR principles within the context of AI and bioimaging. These guidelines, published on FAIRsharing, provide the foundation for creating open and interoperable tools and datasets. Key guidelines include:

- **BioImage Model Zoo (BMZ) model specification**: Available on GitHub and through the FAIRsharing record, published here: https://fairsharing.org/5644. The model specifications define a standard format for trained AI models to be documented, cross-compatible with bioimaging tools (i.e., the community partners), assessed and disseminated through the BioImage Model Zoo.



- **Guidelines for AI-ready Datasets**
  - Manuscript: Metadata, Incentives, Formats, and Accessibility (MIFA), under review in Nature Methods. Preprint: https://arxiv.org/abs/2311.10443
  - FAIRsharing record: https://fairsharing.org/5639
  - Schemas and models to work with the MIFA metadata model: https://github.com/BioImage-Archive/bia-mifa-models



www.ai4life.eurobioimaging.eu
ai4life@eurobioimaging.eu
@AI4LifeTeam

    - **BMZ data specification**: Available on GitHub. These specifications allow for the creation of centralised data collections in the BMZ that associate publicly available datasets and the corresponding models trained on them.
- **Guidelines for community partners**: the community partners of the BioImage Model Zoo are usually, an organization, a company, a research group, or a software team (of one or more) that can consume and/or produce resources of the BioImage.Io model zoo. While their tools are not required to be open source (e.g., commercial microscopy imaging software), their model contributions to the zoo have to be freely available, disseminated primarily through the zoo and interoperable with the rest of the community partners. These guidelines are meant to enable the widespread adoption of AI4Life tools while maintaining the FAIR principles.

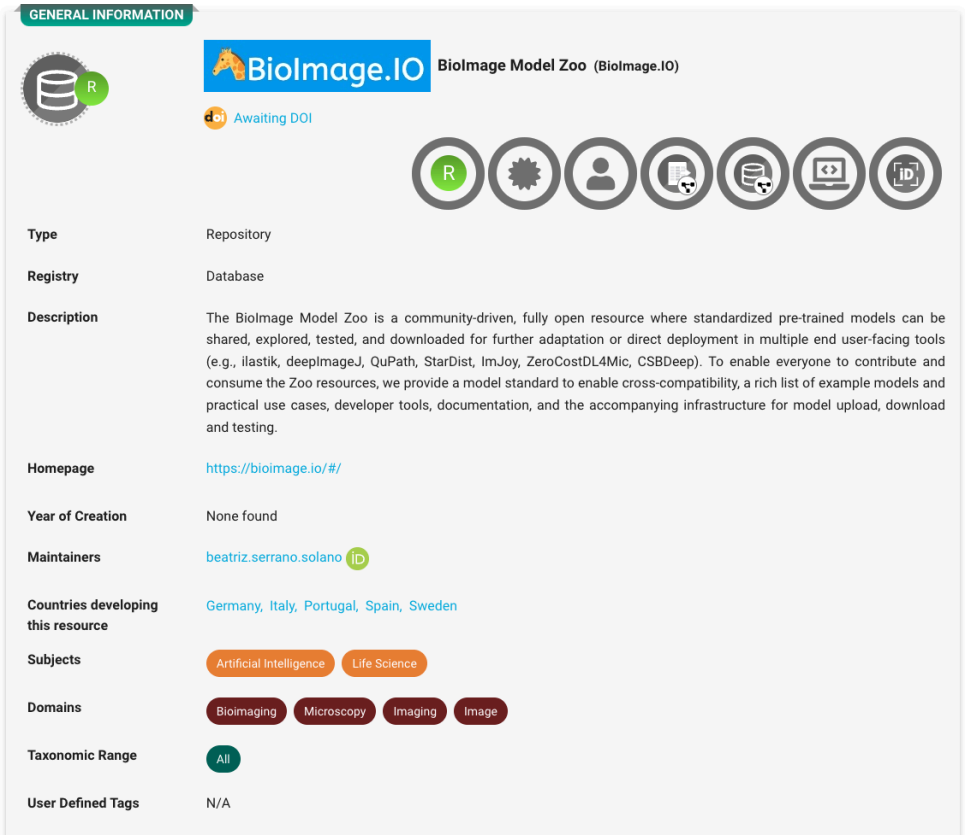
3.2 FAIR Resources by the AI4Life Consortium

Our consortium created and operates on a BioImage.IO Organization on GitHub that serves as the umbrella for the components in the "BioImage.IO ecosystem", where we are hosting the software resources of AI4Life, such as the BioEngine, the BioImage Model Zoo, the BioImage ChatBot and the BioImage.IO Colab, among other accessorial packages. Below we highlight additional FAIR Resources by the AI4Life Consortium:

- **BioImage Model Zoo (BMZ):**
  Virtual Infrastructure to host and serve pre-trained AI models, data and APIs such as Python notebooks,: https://bioimage.io/#/. The BMZ provides a free browser based infrastructure for the interactive dissemination of AI models for bioimaging. Thanks to the specification standards developed in AI4Life (e.g., model specifications), (1) the models are fully documented (Findable for non-expert life-scientists); (2) freely available (Accessible worldwide for their practical use); (3) they are cross-compatible with the community partners' software (Interoperable with the tools in the life sciences community); and (4) deployable through user-friendly tools (Reusable by non-expert life-scientists). Additionally, the entire development of the BioImage Model Zoo is fully open through GitHub.

www.ai4life.eurobioimaging.eu
ai4life@eurobioimaging.eu
@AI4LifeTeam

- **The Bioimage Model Zoo FAIRsharing record:** https://fairsharing.org/5622

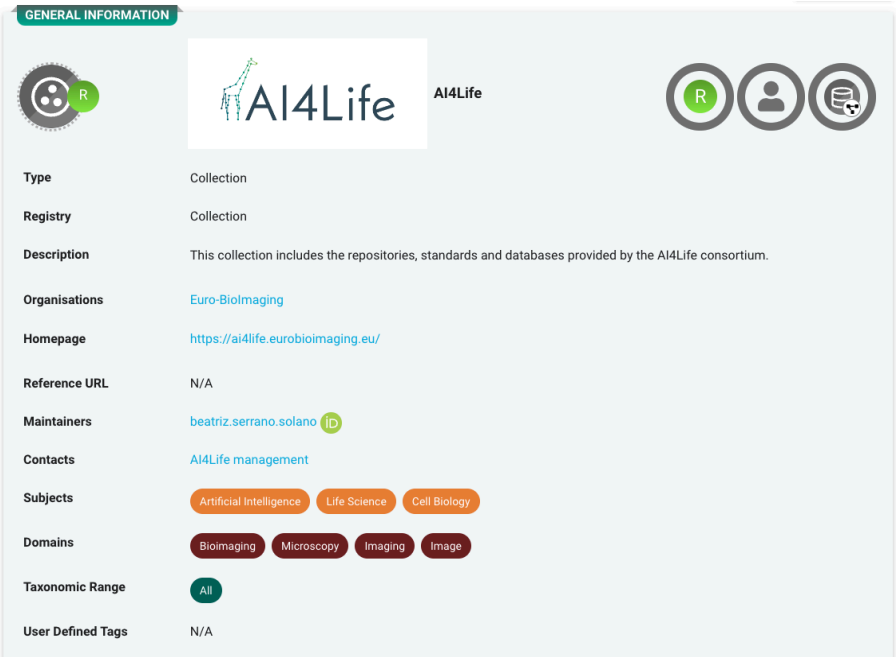


- **Open BMZ Meeting Minutes:** https://github.com/bioimage-io/bioimage.io/issues/28
- **Open Deliverables of the AI4Life Consortium**, including:
    - Data Management Plans (DMPs): Customised DMPs tailored to the specific needs of bioimaging data, emphasising compliance with FAIR standards.
    - Annotation standards and software, libraries, and reference examples.
    - Risk Management Plan
    - Dissemination Strategy of news and information of the consortium.
- **Training and Support Material:** Programs and resources designed to equip researchers with the skills and knowledge necessary to implement FAIR-compliant practices.
    - Training materials on the Galaxy Training Network: https://training.galaxyproject.org/training-material/topics/ai4life/
    - AI4Life Open Call GitHub page: https://github.com/ai4life-opencalls
    - Technical Documentation of the BMZ and other AI4Life infrastructures: https://bioimage.io/docs/#/

www.ai4life.eurobioimaging.eu
ai4life@eurobioimaging.eu
@AI4LifeTeam

- AI4Life YouTube Channel: https://www.youtube.com/@ai4life

- **AI4Life collection on FAIRsharing.org** (gathers all the FAIRsharing records)



- **AI4Life Zenodo Community:** (Open publications, datasets, software, posters, etc.) https://zenodo.org/communities/ai4life
- **Zotero group library**: (Collection of all publications that acknowledge AI4Life.) https://www.zotero.org/groups/5145082/ai4life
- **The BioImage Archive AI Gallery** (containing contributions from AI4Life): https://www.ebi.ac.uk/bioimage-archive/galleries/AI.html

3.3 Impact and Implementation

The AI4Life project catalyses advancements in AI-driven research by making AI-based image analysis methods more accessible to the scientific community. Researchers can leverage our tools and infrastructures to gain deeper insights into complex biological processes and enhance their studies' accuracy and efficiency.

By democratising the availability of user-friendly AI tools, AI4Life empowers researchers from diverse backgrounds to utilise state-of-the-art AI techniques without requiring extensive computational expertise. This is enabled by the resources being freely available through the BMZ, such as ready-to-be-used (pre-trained) Deep Learning (DL) models, model running in the browser, and the model runner BioEngine, all being FAIR. One primary goal was to introduce a level of interoperability between image analysis tools

and workflows that was only possible after an independent model format was created and established in the context of the BMZ work conducted in AI4Life.

Hence, establishing standards for content submission, storage, and FAIR access now promotes interoperability between different research platforms. This standardisation ensures that data and AI models are easily sharable and reusable, reducing redundancy and holding the potential to accelerate scientific progress.

Beyond the technical infrastructure, AI4Life has successfully engaged the research community through Open Calls, Challenges, and training events. These initiatives have facilitated the exchange of ideas, encouraged collaboration, and inspired novel solutions to image analysis problems. To maximise the benefit for the global scientific community, we have made all results that were obtained in such collaborations (e.g., the Open Calls and Challenges) openly available, including a growing collection of training materials and tutorials through online platforms, such as the AI4Life YouTube Channel and the Galaxy Training Network.

4. Conclusion

AI4Life aims to expand its impact by exploring new avenues for AI integration in the life sciences. Building on our achievements in democratising access to AI tools, establishing FAIR guidelines and interoperability standards, and engaging the global research community. Our efforts focus on enhancing the scalability of AI solutions, fostering greater interdisciplinary collaboration, and exploring novel applications of AI in various research areas. We remain committed to driving forward the potential of AI to transform scientific discovery.

www.ai4life.eurobioimaging.eu
ai4life@eurobioimaging.eu
@AI4LifeTeam