



POLICY BRIEF

RESEARCH INFRASTRUCTURES

INTRODUCTION: PROJECT TITLE AND PROJECT OBJECTIVES

CALL: HORIZON-INFRA-2021-SERV-01

TOPIC: HORIZON-INFRA-2021-SERV-01-06

PROJECT: Artificial Intelligence for Image Data Analysis in the Life Sciences (AI4Life).

<https://ai4life.eurobioimaging.eu/>

Objectives

- (1) **Democratized FAIR access to AI-based image analysis methods through a common application interface standard – which we call the BioImage Model Zoo (BMZ)** – computationally powered by the European Open Science Cloud (EOSC) infrastructure. The AI4Life service landscape will make even the latest and most advanced methods instantly available to life scientists, particularly to those without in-depth computational expertise. Additionally, our services will be available for local installation (“on-premise”).
- (2) **Establish standards for the submission, storage and FAIR access of reference data, reference annotations (ground-truth), trained AI models, and trainable AI methods.** This will happen in close partnership with existing European infrastructures, e.g., in the BioImage Archive (BIA) at EMBL-EBI and the European Open Science Cloud (EOSC). The BioImage Archive is an integral part of the image data services offered by Euro-BioImaging. The access to our infrastructures will be virtual, open, and free of charge for all interested users in industry and academia. All established standards will be open to foster compatibility and synergy also for AI-driven industry efforts.
- (3) **Simple model deployment, sharing, and dissemination of AI-based methods as a new developer-facing service of the BioImage Model Zoo.** Our standards and services will enable computational scientists from the life sciences, AI, ML and Computer Vision fields to reach a large audience of life science users for their latest AI-based analysis methods. This will allow computational scientists to continuously push the boundary of what AI-based methods can achieve in the context of scientific imaging and image data analyses. While today the deployment and dissemination of a new method or tool is laborious and time consuming, our infrastructures and services will reduce this overhead time and greatly accelerate the rate of computational innovation in academia and in SMEs.
- (4) **Organise Open Calls and Challenges** for outstanding image analysis problems in the EU Mission areas of the Horizon Europe framework programme. Building on the common practice of the AI and Computer Vision fields to cast important analysis problems into well-defined challenge tasks on a prepared body of data, with (a) given reference baseline solution(s), these computational disciplines will engage in a friendly competition to find better and better performing methods for the most important problems. Best performing approaches to each Challenge we organise will immediately become available through our BioImage Model Zoo infrastructure, following the pioneering examples of the Human Protein Atlas (Ouyang, Winsnes, et al. 2019) image classification competitions which had thousands of participating teams world-wide.

- (5) **Empower common image analysis platforms by AI tools.** We will integrate all analysis models in the BioImage Model Zoo into popular end-user software tools. We will start with some of the arguably most common software tools built and maintained by members of our consortium (i.e., Fiji, ilastik, ImJoy, deepImageJ, ZeroCostDL4Mic). This truly makes the shared community models in the Zoo reusable and will enable all model-empowered software tools to offer better and ever-growing analyses. Additionally, we will support developers of other software tools that are not yet part of our consortium to connect to our Model Zoo and become a consumer (and/or producer) of trained models and analysis methods themselves.
- (6) **Organise outreach and training events** for life scientists as targeted image analysis courses and workshops as well as by participation in the largest international conferences. By involving a wide set of life sciences Research Infrastructures (RIs) in the community outreach activities, AI4Life will ensure broad participation, contribution and uptake by imaging scientists rooted in various scientific disciplines, including for example structural biology, medicinal chemistry, marine biodiversity and plant phenotyping. Outreach and teaching events for method developers and potential contributors to Open Calls and Challenges are equally important to create a vibrant community and productive interdisciplinary discourse during the duration of the proposed project.

POLICY IMPLICATIONS AND RECOMMENDATIONS

Implementation of research infrastructures.

AI4Life is committed to enhancing the BioImage Model Zoo (BMZ), an AI model repository that provides life scientists with democratised access to AI-based image analysis methods and FAIR image analysis infrastructure services. In this context, AI4Life has developed the BioEngine, a cloud-based framework designed to facilitate the execution of pre-trained AI models directly from the BMZ website, which allows all users - independent of their level of computational expertise or local compute resources - to test and run models for their image data analysis. However, limited compute resources available to the project consortium still represent a bottleneck on the way to allow users to fully scale up image analysis through the BMZ website. AI4Life is therefore intending to secure computational resources through the European Open Science Cloud (EOSC). Also, see recommendation 7.

Furthermore, to engage developers with large and/or well-known pre-trained model collections as community partners of AI4Life infrastructure, dedicated efforts of bringing their model collections into a FAIR and compatible state would be needed.

RECOMMENDATION 1: Implement Open Calls with cascading grants to support model developers with large and/or well-known pre-trained model collections, enabling them to bring the model collections into a FAIR-compliant state through the BMZ.

RECOMMENDATION 2: Implement Open Calls with cascading grants support to early adopters and software tool developers to contribute to the maturation of the BMZ by achieving tools interoperability. This will help foster development of the BMZ into a community infrastructure visible in the research software development arena as well.

Access to research infrastructures.

AI4Life is supporting users (through direct support) and developers (through contributor services) around the BMZ, improving accessibility of the infrastructure. Together with targeted dissemination and increased usage of the resources through Open Calls that are open to scientists from all life

science domains worldwide, the project will bring novel users through the infrastructure to the EOSC.

RECOMMENDATION 3: Implement Open Calls with cascading grants to support image analysts and tool maintainers and developers to cater the needs and onboard new users.

International co-operation of research infrastructures.

To facilitate the development of community standards for image data and for AI models, the AI4Life project coordinates interactions with various research infrastructures, like EU-Openscreen, Instruct, EMBRC, EMPHASIS and data-heavy RIs like ELIXIR. Specific interaction with the international community is mediated through networks, like Global BioImaging, NEUBIAS, COMULISglobe, volumeEM, etc. Additionally, we are partnering with research teams and organisations beyond Europe, such as Bioimaging North America (BINA), Janelia (e.g. COSEM or Jan Funke) and the BROAD institute (e.g. with Beth Cimini). The aim of such discussions is to align on standards, avoid duplication of efforts, ensure cross-compatibility, and integrate new tools created by extra European groups and agencies. Further dissemination of activities is planned through newly funded projects like foundingGIDE. These interactions underpin the EU policy on Open Science supporting the development of image data as a global resource.

Employment and skills in research infrastructures.

AI4Life has contributed targeted course modules on the BMZ at numerous types of events and occasions (e.g. EMBO courses on image analysis, DL4MIA course at HT, DL for Image Analysis at EMBL or the DL@MBL deep learning courses at the Marine Biology Laboratories in Woods Hole, USA). For the computational community, AI4Life has organised hackathons, where numerous international participants contributed to shaping the AI4Life metadata standards and the infrastructure behind the BioImage Model Zoo. Through the development and dissemination of shared community standards, AI4Life supports and promotes good practices for the submission, storage and access of both image data and AI models. Furthermore, the infrastructure services provided by AI4Life facilitate enhanced image analysis, accelerating the pace of research within the imaging community.

RECOMMENDATION 4: Cultivate community collaboration by continued organisation of training, developer and dissemination events to gather community feedback, foster community engagement and contributions, and thereby enhance the BMZ capabilities.

Greening of research infrastructures.

By promoting the development of interoperable standards for pre-trained AI models AI4Life supports reproducible and efficient science and allows the reuse of scientific outputs, prolonging the utility of existing resources. In addition, by providing access to shared compute infrastructure to run the models in a user-friendly way, AI4Life promotes flexible usage of these resources eliminating the need to duplicate infrastructure into local instances. Furthermore, storing annotated images in the form of AI-ready datasets minimises the need to rerun resource-intensive models to produce outputs.

Interaction of research infrastructures with industry.

AI4Life maintains interactions with key imaging and pharma industry representatives such as Bayer, Zeiss and Leica. Through these contacts, AI4Life aims to promote open standards, compatibility and interoperability at all levels and to engage industry partners to contribute to and freely consume all available analysis models (AI-based tools and methods) in the BMZ service portfolio. A prime example of these collaborations is the integration of BMZ models into Leica's AIVIA microscopy image analysis software.

Technology development, data and digital services, digitalisation.

Through the BMZ infrastructure services, the related standardisation of AI models and methods and the multifaceted training and outreach activities, AI-based methods will become an essential

and easily accessible technology for cutting-edge, imaging-heavy life sciences research. New state-of-the-art methods will immediately propagate to the community and connect microscope and method developers to practitioners in research laboratories and facilities.

RECOMMENDATION 5: Allocate resources to support data reuse to promote improved, accelerated and sustainable technology and service developments.

Level of connection of your RI to EOSC.

AI4Life supports making image data FAIR and AI-ready through a community-driven process. Currently, such datasets are openly available through the [BioImage Archive](#), while the guidelines, together with other project outputs are indexed in the EOSC resources through [OpenAIRE](#). Besides FAIR data, AI4Life is offering digital services (e.g. the BMZ) and will in the future offer web-based human-in-the-loop AI model execution, data labelling, model retraining, etc. The project also interacts with other cross-domain data-driven and EOSC projects, like [iMagine](#), [ANERIS](#), EUCanImage, and EOSC4Cancer, through its extensive network of RI partners.

RECOMMENDATION 6: Recognize the benefits of direct user support for making image data and analysis methods FAIR through data stewardship services. Hence, both funding and policy support to recognise these efforts will be crucial to underpin EU policy on Open Science.

RECOMMENDATION 7: Establishing mechanisms for accessing on-demand compute resources together with technical support and compute opportunities via the EOSC is important to sustain the AI4Life infrastructure that uses resources intensively, but often in bursts. Implementation of access beyond a basic compute contingent on a per-user basis via existing agreements or through compute tokens will need to be enabled.

Sustainability of research infrastructures.

The key outputs of the project that support the sustainability of the infrastructure as a whole include:

- BioImage Model Zoo: This web service makes annotated pre-trained ML/DL models available to researchers and interfaces with the backend compute resources. AI4Life is extending the community partners and usability of this resource, making key contributions to increase its usage and community uptake.
- BioEngine: This backend infrastructure empowers the BMZ with the capability to run the pre-trained models on any image dataset. That way, the AI4Life project provides access to extended compute resources, increasing its capacity to accommodate a growing user load.
- Model, data and annotations standards: To promote the wide use of BMZ models the AI4Life project is developing standards to promote interoperability of the infrastructure. These standards are collaboratively developed within the community, which ensures the long-term sustainability of the outputs.
- Open Calls: While the direct support to the user provided via the Open Calls encourages the use of the BMZ infrastructure, the open call playbook for execution and project selection will be adopted in future projects to run Open calls.