

Development FAIR image analysis workflows and RDM pipelines in Galaxy

Riccardo Massei¹, Beatriz Serrano-Solano², Anne Fouilloux³, Björg Gruening⁴, Yi Sun⁵, Diana Chiang⁴, Matthias Bernt¹, Leonid Kostrykin⁶

¹ Helmholtz, Center for Environmental Research (UFZ), Leipzig

² Euro-BioImaging ERIC Bio-Hub, EMBL Heidelberg, Germany

³ Simula Research Laboratory, Oslo, Norway

⁴ Department of Computer Science, University of Freiburg, Freiburg im Breisgau

⁵ EMBL Heidelberg, Data Science Centre, Heidelberg, Germany

⁶ Biomedical Computer Vision Group, Heidelberg University, BioQuant, IPMB, Heidelberg, Germany

Abstract

Imaging is crucial across various scientific disciplines, particularly in life sciences, where it plays a key role in studies ranging from single molecules to whole organisms. However, the complexity and sheer volume of image data present significant challenges. Managing and analyzing this data efficiently requires well-defined image processing tools and analysis pipelines that align with the FAIR principles—ensuring they are findable, accessible, interoperable, and reusable across different domains. In the frame of NFDI4BIOIMAGE¹ (the National Research Data Infrastructure focusing on bioimaging in Germany), we want to find viable solutions for storing, processing, analyzing, and sharing bioimaging data. In particular, we want to develop solutions to make findable and machine-readable metadata developing analysis pipelines. In scientific research, such pipelines are crucial for maintaining data integrity, supporting reproducibility, and enabling interdisciplinary collaboration. These tools can be used by different users to retrieve images based on specific attributes as well as support quality control by identifying appropriate metadata.

Galaxy, an open-source, web-based platform for data-intensive research, offers a solution by enabling the construction of reproducible pipelines for image analysis². By integrating popular analysis software like CellProfiler and connecting with cloud services such as OMERO and IDR, Galaxy facilitates the seamless access and management of image data. This capability is particularly valuable in bioimaging, where automated pipelines can streamline the handling of complex metadata, ensuring data integrity and fostering interdisciplinary collaboration. This approach not only increases the efficiency of RDM processes in bioimaging but also contributes to the broader scientific community's efforts to embrace FAIR principles, ultimately advancing scientific discovery and innovation.

In the present poster, we showed how to integrate RDM processes and tools in Galaxy. We will showcase how Images can be enriched with metadata (i.e. key-value pairs, tags, raw data, regions of interest) and uploaded to a target OME Remote Objects (OMERO) server using a novel set of OMERO tools developed with Galaxy³. Workflows give the possibility to the user to intuitively fetch images from the local server and perform image analysis (i.e. annotation). Furthermore, we will show the potential integration of electronic lab books such as eLabFTW⁴, cloud storage systems (i.e. OneData)⁵ and interactive notebooks (Jupyter Notebooks)⁶ in the Galaxy pipeline.

Keywords: bioimaging, HCS, FAIR, Workflow Management systems (WMS)

Resources

- **The Galaxy Training Network:** <https://training.galaxyproject.org/> - Resource of Materials for Training in Galaxy

- **Image Analysis Community in Galaxy:** <https://galaxyproject.org/community/sig/image-analysis/> - Overview on the bioimaging community in Galaxy

- **Image Analysis using Galaxy:** <https://www.youtube.com/watch?v=wAHOPT6IRV0> - Online Event on the Topic

Author contributions

RM: Conceptualization, Writing, Resources— original draft; **BSS:** Conceptualization, Resources , Project administration; **AF:** Methodology, Resources; **BG:** Funding acquisition; **DCJ:** Resources; **MB:** Software, Methodology; **LK:** Conceptualization, Writing – review & editing

Competing interests

The authors declare that they have no competing interests.

Funding

This work is part of the NFDI4BIOIMAGE project funded by the Deutsche Forschungsgemeinschaft (DFG grant number NFDI 46/1, project number 501864659). NFDI4BIOIMAGE is a consortium of the National Research Data Infrastructure (Nationale Forschungsdateninfrastruktur, NFDI) and OSCARS, that has received funding from the European Commission's Horizon Europe Research and Innovation programme under grant agreement No. 101129751.

References

- [1] Schmidt, Christian, and Elisa Ferrando-May. "NFDI4BIOIMAGE—An initiative for a national research data infrastructure for microscopy data." *Tage 2021* (2021): 339.
- [2] Kostykin, Leonid, et al. "Development of FAIR image analysis workflows and training in Galaxy." (2025).
- [3] Riccardo Massei, Overview of the Galaxy OMERO-suite - Upload images and metadata in OMERO using Galaxy (Galaxy Training Materials). <https://training.galaxyproject.org/training-material/topics/imaging/tutorials/omero-suite/tutorial.html> Online; accessed Fri Apr 11 2025
- [4] eLabFTW integration in Galaxy - <https://galaxyproject.org/news/2025-04-02-elabftw-integration/>
- [5] Łukasz Opiola, Bartosz Walkowicz, Importing (uploading) data from Onedata (Galaxy Training Materials). <https://training.galaxyproject.org/training-material/topics/galaxy-interface/tutorials/onedata-remote-import/tutorial.html> Online; accessed Fri Apr 11 2025
- [6] Anne Fouilloux, JupyterLab in Galaxy (Galaxy Training Materials). <https://training.galaxyproject.org/training-material/topics/galaxy-interface/tutorials/jupyterlab/tutorial.html> Online; accessed Fri Apr 11 2025