



Evolving FAIR Image Analysis in Galaxy for Cross-domain and AI-ready Applications

BioHackathon series:

[BioHackathon Europe 2025](#)

Berlin, Germany, 2025

[Project 9](#)

Submitted: 31 Mar 2026

License:

Authors retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

Published by [BioHackrXiv.org](#)

Diana Chiang ¹, **Pavankumar Videm** ¹, **David Lopez Tabernero** ¹, **Maarten W. Paul** ², **Alireza Heidari** ¹, **Martin Etzrodt** ³, **Beatriz Serrano-Solano** ⁴, and **Leonid Kostrykin** ⁵

¹ Department of Computer Science, University of Freiburg, Freiburg, Germany ² NL-Biolmaging, Leiden University, Leiden, the Netherlands ³ ISCC Foundation ⁴ Euro-Biolmaging ERIC Bio-Hub, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany ⁵ Biomedical Computer Vision Group, Heidelberg University, BioQuant, Heidelberg, ELIXIR Germany

1 | Introduction

The growing integration of image-based technologies across life sciences, environmental research, and other domains has intensified the need for interoperable, reproducible, and FAIR-compliant image analysis infrastructures. The ELIXIR BioHackathon Europe 2025 provided a collaborative environment to advance these goals within the Galaxy ecosystem (Galaxy Community, 2024).

Previous projects at the ELIXIR BioHackathon Europe focused on community development and establishing conventions (Kostrykin et al., 2024) or implementing FAIR workflows and training materials (Kostrykin et al., 2025) for image analysis using Galaxy. Building on the achievements of the previous years, Project 9 at the ELIXIR BioHackathon Europe 2025 was entitled “*Evolving FAIR Image Analysis in Galaxy for Cross-domain and AI-ready Applications*” and aimed to strengthen Galaxy’s capabilities for bioimage analysis while enhancing cross-domain interoperability and reproducibility. The project addressed three major challenges:

1. Improving semantic annotation for image analysis resources.
2. Introducing content-based reproducibility validation mechanisms.
3. Streamlining tools and training materials for accessible bioimage analysis.

To achieve these goals, the project involved participants with heterogeneous backgrounds from different institutions, including ELIXIR, Euro-Biolmaging, and the ISCC Foundation, joining both on-site and online. The goals of the project were also closely aligned with the [OSCARS-FIESTA](#) and [OSCARS-BIICODES](#) projects.

By integrating ontology development with standard-based validation tools, the project contributed to a FAIR-aligned, cross-domain image analysis framework within Galaxy.

2 | Implementation

Our work was structured into three coordinated work packages (WPs) as described below. WP1 focused on cross-domain discovery and interoperability (Section 2.1), WP2 on integrating content-based fingerprints to enhance reproducibility and AI-readiness (Section 2.2), and WP3 on streamlining bioimage analysis in general, by improving tools, training materials, and the usability thereof (Section 2.3).

2.1 | WP1: Cross-domain Support Improvements

WP1 focused on semantic interoperability through improvements to the EDAM Bioimaging ontology. The EDAM Bioimaging ontology extends the core EDAM ontology (Ison et al., 2013)

with terms specific to bioimage data and analysis. However, this ontology extension is still under development. In the future, it will enable fine-grained annotation of tools, workflows, and tutorials within our community for specific analysis tasks. The BioHackathon team proposed extensions to the EDAM Bioimaging ontology to better align with the image analysis tasks supported by Galaxy. Furthermore, we curated a specialized vocabulary that reflects both the latest EDAM Bioimaging version and our specific recommendations (Section 3.1), which was then used to annotate the image analysis tutorials on the Galaxy Training Network (Hiltemann et al., 2023). This enhances the discoverability of the tutorials in the short term while ensuring seamless compatibility with evolving standards.

Through this work package, semantic annotation became a central mechanism to enable cross-domain discovery, interoperability, and alignment with evolving community standards.

2.2 | WP2: Integration of the ISCC ISO Standard in Galaxy

WP2 introduced content-based validation mechanisms by integrating the [International Standard Content Code \(ISCC\)](#) into Galaxy. To support reproducibility, the ISCC-SUM tool suite was integrated into Galaxy. This generates content-based fingerprints to validate final or intermediate results by implementing the International Standard Content Code (ISCC). ISCC-SUM is file-type agnostic and not limited to bioimaging data; it can be extended to various heterogeneous data types, such as omics data. This content-based validation complements Galaxy's workflow provenance tracing, providing format-independent reproducibility assurance aligned with FAIR principles.

Three Galaxy tools were implemented: ([bmcv/galaxy-image-analysis/pull/159](#), [bmcv/galaxy-image-analysis/pull/162](#))

1. A tool to compute ISCC codes for datasets.
2. A tool to compare ISCC codes based on a configurable similarity threshold.
3. A tool to filter datasets according to ISCC similarity constraints.

These tools can be used to incorporate reproducibility-assuring checks as workflow steps directly into analysis pipelines in Galaxy.

The integration of the ISCC-SUM tool suite into Galaxy enhances the AI-readiness of the platform. In this context, ISCC codes have several advantages. One advantage is that ISCC codes enable deduplication and dataset cleaning, which is important for removing redundant training data to prevent bias from over-represented samples. Another advantage is that they allow grouping similar (or dissimilar) data. This can be used, for example, during model training, to select test sets that truly evaluate the generalization of the model. When using pre-trained models, this can also be used to determine whether the image data used in an analysis is too different from the training data (out-of-distribution data), and thus to decide whether a tool or model is appropriate for that image data. Moreover, ISCC codes can also be used to verify data integrity, thus generally enhancing the trustworthiness of analyses.

2.3 | WP3: Streamlining BioImage Analysis in Galaxy

WP3 represented the most extensive implementation effort and focused on improving usability, interoperability, and the overall toolset for bioimage analysis. The entry point for tutorials on image analysis in Galaxy is the *“Imaging”* topic on the Galaxy Training Network (GTN) landing page. This section features a growing set of curated tutorials covering tasks such as image pre-processing, segmentation, and feature extraction. In addition, it introduces advanced analyses such as tissue multiplexing, parameter optimization, and the integration of machine learning models. To streamline the onboarding of new users in the Galaxy ecosystem for image analysis, a major endeavor has been to create the tutorial *“Where to start with bioimage analysis in Galaxy”*, which now provides a fundamental overview of image data representations, data types, tools, and analysis possibilities within Galaxy, including identifying entry points with OMERO. The outcomes are described in Section 3.3.1.

Improving the integration between OMERO and Galaxy has been another building block of this work package. Users with institutional OMERO instances can now transfer image data directly into Galaxy workflows. This reduces friction between image management and analysis, preservation of image metadata, and facilitates institutional adoption. Details on the outcomes are given in Section 3.3.2.

Moreover, we have now developed a Galaxy Lab for the image analysis community. Galaxy Labs provides a community-tailored landing page, offering users a consistent interface with curated tools, workflows, and training materials that can be deployed on any Galaxy server. A production deployment is now available on the [French Galaxy server](#) and will be deployed on other public servers in the future.

Further work included substantial improvements to Galaxy tool integrations, tool implementations, and training materials. These contributions streamlined the interoperability of tools and platforms (e.g., the BioImage Archive, Hartley et al., 2022), and the outcomes are described in Sections 3.3.3 and 3.3.4, respectively.

Collectively, WP3 improved both the technical foundation and the user-facing experience for bioimage analysis in Galaxy.

3 | Results

Most of our work concerned the implementation of tools and tutorials as described in Section 2. In this section, we provide further results and highlight specific developments.

3.1 | A Curated Semantic Framework for Bioimage Analysis

A key result of the project is the development and application of a curated set of annotation terms aligned with the EDAM Bioimaging ontology. These terms were used to annotate tutorials within the Galaxy Training Network and reflect real-world bioimage analysis practice in Galaxy.

The curated term set includes:

- Cell segmentation
- Object feature extraction
- Image thresholding
- High-throughput screening
- Fluorescence microscopy
- Tissue imaging
- Conversion
- Object counting
- Overlay
- Cell tracking
- Multi-channel image
- Data management
- Data handling
- Data sharing
- Parameter optimisation
- Validation
- ROI selection
- Temporal analysis
- Time-lapse data
- Image annotation
- Deep learning
- Data decomposition
- Rasterization

- Tessellation
- Bioimaging

This structured vocabulary improves tutorial discoverability, supports fine-grained categorisation, and prepares the ecosystem for future ontology harmonisation.

3.2 | Format-independent Reproducibility via ISCC

The integration of the ISCC-SUM tool suite into Galaxy establishes content-based reproducibility validation across heterogeneous data types. Unlike checksum-based validation, which is sensitive to file representation, ISCC codes provide similarity-aware fingerprints. This enables:

- Validation of intermediate and final workflow outputs
- Threshold-based similarity comparisons
- Cross-format reproducibility checks
- Extension beyond imaging to omics and other data domains

This complements Galaxy's provenance tracking and advances reproducibility toward a format-independent, FAIR-aligned model. A complementary [tutorial \(galaxyproject/training-material/pull/6460\)](#) was also developed, highlighting the different use cases of the suite described in Section 2.2.

3.3 | Streamlined Integrations, Tools, and Training Materials

In this section, we describe the outcomes of our work on WP3, as described in Section 2.3.

3.3.1 | Onboarding Tutorial for Newcomers

Within the Galaxy Training Network, we developed a new introductory [tutorial \(galaxyproject/training-material/pull/6603\)](#) specifically designed for newcomers to bioimage analysis in Galaxy: The tutorial, "*Where to start with bioimage analysis in Galaxy*", provides a structured and accessible entry point into the conceptual and practical foundations of image analysis. It introduces key concepts such as pixels and voxels, 5D hyperstacks (XYZCT), bit-depth, and spatial calibration, ensuring that users understand the foundations of biological image data rather than prematurely engaging in analyses.

The tutorial also guides users through handling proprietary microscopy formats via Bio-Formats (Linkert et al., 2010), understanding OME-NGFF standards (e.g., Moore et al., 2023), and leveraging the OMERO integration for institutional data access. Sections structured as hands-on walks guide users through practical exercises, including metadata inspection, image filtering, thresholding, and validation. To support informed methodological choices, the tutorial provides a logical decision tree to help users decide between classical computer vision workflows and AI-driven approaches. Finally, it includes a dedicated "pitfall guardrail" section highlighting common sources of error, such as JPEG compression artefacts, unintended RGB channel merging, and photobleaching effects, thereby promoting robust and reproducible image analysis practices from the outset.

3.3.2 | OMERO File Source Plugin

As part of WP3, we significantly enhanced the integration between OMERO and Galaxy by implementing a new OMERO file source plugin ([galaxyproject/galaxy/pull/21367](#)) directly within Galaxy. This plugin enables hierarchical browsing of OMERO's native

Project → Dataset → Image

structure, allowing users to navigate their institutional image repositories in a structured and intuitive way. To ensure scalability, server-side pagination was introduced using HQL queries, enabling efficient browsing even for very large image collections. In addition, server-side search

functionality allows users to filter projects, datasets, and images by name directly on the OMERO server, reducing data transfer overhead and improving responsiveness.

The plugin also introduces a smart image download mechanism tailored to preserve data fidelity and maximise compatibility. Whenever possible, Galaxy attempts to retrieve the original imported image files, thereby preserving the native file format (with certain repository-specific limitations). If original files are not accessible, the system automatically exports multi-page TIFF files containing all Z-planes, ensuring complete volumetric data retrieval. The implementation gracefully handles restricted repositories—such as those that block direct access to original files—thereby enabling robust and FAIR-aligned interoperability between OMERO-based data management and Galaxy-based analysis workflows.

3.3.3 | Improved Integration with Other Platforms

We have updated the BioImage Archive retrieval tool in Galaxy ([bgruening/galaxytools/pull/1713](#)), enhancing its usability. In addition, we have drafted a tool ([bgruening/galaxytools/pull/1716](#)) for uploading image data from Galaxy to BioStudies (Sarkans et al., 2017), which the BioImage Archive builds upon. However, the tool has not yet been deployed to the Galaxy ToolShed (Blankenberg et al., 2014) due to a lack of automated test routines (BioStudies currently does not provide APIs for testing).

In addition, the BiaPy integration in Galaxy has been accommodated by a [tutorial \(galaxyproject/training-material/pull/6621\)](#) that guides the user through the process of performing inference with a deep learning model from BiaPy in Galaxy.

3.3.4 | Enhanced Tool Interoperability

We have updated the integration of Cellpose (Stringer & Pachitariu, 2025) in Galaxy to improve robustness when handling microscopy data ([bgruening/galaxytools/pull/1692](#)). The integration now explicitly requires a single-channel image as input, thereby avoiding common issues arising from multi-channel TIFF files where channel interpretation could previously lead to ambiguous or incorrect segmentation results (or total job failure). By enforcing clear channel selection prior to execution, the updated integration of Cellpose ensures more predictable behaviour and better reproducibility across diverse imaging datasets. In addition, the integration was refined to explicitly document and reference the supported Cellpose version.

We have also updated the integration of ilastik (Berg et al., 2019) as an Interactive Tool in Galaxy ([usegalaxy-eu/galaxy/pull/331](#)). The tool was upgraded to version 1.4.1.post1, ensuring access to recent improvements and bug fixes from the upstream project. Support was added to export processed images as OME-Zarr directly to the Galaxy history, strengthening alignment with modern cloud-compatible bioimaging standards and facilitating downstream reuse. In addition, a bug in the Galaxy integration was resolved that previously prevented images from being properly exported from ilastik sessions into the Galaxy history, restoring seamless transfer of results into reproducible workflows.

Many tools in the Galaxy ecosystem for image analysis use *label maps* (label images) for spatial annotations (e.g., segmentation masks, regions of interest). Other tools, however, use vector-based spatial annotations. To ensure tool interoperability, we already had conversion tools implemented; which, however, only supported vector-based annotations in rigid tabular formats (e.g., coordinates and sizes of rectangles). While such a tabular format is human-readable and easy to edit manually, it is a non-standard format that has limited interoperability with external image analysis tools and platforms, which increasingly rely on GeoJSON for spatial annotations (e.g., QuPath, Bankhead et al., 2017). Earlier attempts to introduce GeoJSON support relied on converting GeoJSON into the existing tabular structure, but this approach imposed structural constraints and restricted the supported geometries, effectively limiting interoperability to a narrow subset of shapes. To address this, the implementation was now refactored to natively process GeoJSON as the primary annotation format ([bmcv/galaxy-image-analysis/pull/160](#)).

Instead of performing intermediate format conversions, GeoJSON geometries are now rasterised directly into label maps, enabling support for a broader range of shapes, including arbitrary polygons, rectangles (not restricted to axis alignment), points, and circles. The previous tabular format remains supported for backward compatibility (internally converted into the more general GeoJSON format). This refactoring not only expanded geometric flexibility and improved compatibility with external tools, but also resolved previously undetected bugs (e.g., rectangle width and height parameters from tabular files were occasionally interchanged during rasterisation).

4 | Conclusion

The project “*Evolving FAIR Image Analysis in Galaxy for Cross-domain and AI-ready Applications*” at ELIXIR BioHackathon Europe 2025 advanced FAIR image analysis in Galaxy along three orthogonal axes: semantic interoperability, reproducibility validation, and streamlining of the overall user experience. By aligning ontology development, ISO-standard integration, and toolset improvements, we have laid the foundation for scalable, cross-domain, and reproducible bioimage analysis within the Galaxy ecosystem.

The work establishes both immediate practical improvements and a forward-looking framework for sustainable, standards-aligned development in the bioimaging community.

Acknowledgements

This work was developed as part of the ELIXIR BioHackathon Europe 2025. We thank *Anup Kumar* for helpful discussions about the goals of the project. We thank *Yi Sun* for help with updating the integration of *ilastik* in Galaxy. We thank *Riccardo Massei* for help with updating the integration of *Cellpose* in Galaxy and writing the *BiaPy* tutorial. BSS acknowledges the OSCARS project, which has received funding from the European Commission's Horizon Europe Research and Innovation programme under grant agreement No. 101129751. The authors utilised the language model ChatGPT developed by OpenAI to assist in structuring and drafting this text.

References

- Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., McQuaid, S., Gray, R. T., Murray, L. J., Coleman, H. G., James, J. A., Salto-Tellez, M., & Hamilton, P. W. (2017). QuPath: Open source software for digital pathology image analysis. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-17204-5> [cito:citation]
- Berg, S., Kutra, D., Kroeger, T., Straehle, C. N., Kausler, B. X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., Eren, K., Cervantes, J. I., Xu, B., Beuttenmueller, F., Wolny, A., Zhang, C., Koethe, U., Hamprecht, F. A., & Kreshuk, A. (2019). Ilastik: Interactive machine learning for (bio)image analysis. *Nature Methods*, 16(12), 1226–1232. <https://doi.org/10.1038/s41592-019-0582-9> [cito:usesMethodIn]
- Blankenberg, D., Von Kuster, G., Bouvier, E., Baker, D., Afgan, E., Stoler, N., Taylor, J., & Nekrutenko, A. (2014). Dissemination of scientific software with galaxy ToolShed. *Genome Biology*, 15(2). <https://doi.org/10.1186/gb4161> [cito:citation]
- Galaxy Community. (2024). The galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Research*, 52(W1), W83–W94. <https://doi.org/10.1093/nar/gkae410> [cito:extends]
- Hartley, M., Kleywegt, G. J., Patwardhan, A., Sarkans, U., Swedlow, J. R., & Brazma, A. (2022). The BioImage archive – building a home for life-sciences microscopy data. *Journal of Molecular Biology*, 434(11), 167505. <https://doi.org/10.1016/j.jmb.2022.167505> [cito:citesAsRelated]

Hiltemann, S., Rasche, H., Gladman, S., Hotz, H.-R., Larivière, D., Blankenberg, D., Jagtap, P. D., Wollmann, T., Bretaudeau, A., Goué, N., Griffin, T. J., Royaux, C., Le Bras, Y., Mehta, S., Syme, A., Coppens, F., Droesbeke, B., Soranzo, N., Bacon, W., ... Batut, B. (2023). Galaxy training: A powerful framework for teaching! *PLOS Computational Biology*, 19(1), e1010752. <https://doi.org/10.1371/journal.pcbi.1010752> [cito:extends]

Ison, J., Kalaš, M., Jonassen, I., Bolser, D., Uludag, M., McWilliam, H., Malone, J., Lopez, R., Pettifer, S., & Rice, P. (2013). EDAM: An ontology of bioinformatics operations, types of data and identifiers, topics and formats. *Bioinformatics*, 29(10), 1325–1332. <https://doi.org/10.1093/bioinformatics/btt113> [cito:citesAsRelated]

Kostyrykin, L., Depoortere, B., Chiang, D., Klumpe, S., Massei, R., Kalaš, M., Fouilloux, A., & Serrano-Solano, B. (2025). *Development of FAIR image analysis workflows and training in galaxy*. https://doi.org/10.37044/osf.io/hbw6m_v1 [cito:citesAsRelated]

Kostyrykin, L., Woller, T., Kalaš, M., Özdemir, B., Andrade Bueno, R., Sun, Y., Kumar, A., Goué, N., Grüning, B., Serrano-Solano, B., & Fouilloux, A. (2024). *Enhancing the image analysis community in galaxy*. <https://doi.org/10.37044/osf.io/w8dsz> [cito:citesAsRelated]

Linkert, M., Rueden, C. T., Allan, C., Burel, J.-M., Moore, W., Patterson, A., Lorange, B., Moore, J., Neves, C., MacDonald, D., Tarkowska, A., Sticco, C., Hill, E., Rossner, M., Eliceiri, K. W., & Swedlow, J. R. (2010). Metadata matters: Access to image data in the real world. *Journal of Cell Biology*, 189(5), 777–782. <https://doi.org/10.1083/jcb.201004104> [cito:usesMethodIn]

Moore, J., Basurto-Lozada, D., Besson, S., Bogovic, J., Bragantini, J., Brown, E. M., Burel, J.-M., Casas Moreno, X., Medeiros, G. de, Diel, E. E., Gault, D., Ghosh, S. S., Gold, I., Halchenko, Y. O., Hartley, M., Horsfall, D., Keller, M. S., Kittisopikul, M., Kovacs, G., ... Swedlow, J. R. (2023). OME-zarr: A cloud-optimized bioimaging file format with international community support. *Histochemistry and Cell Biology*, 160(3), 223–251. <https://doi.org/10.1007/s00418-023-02209-1> [cito:usesMethodIn]

Sarkans, U., Gostev, M., Athar, A., Behrang, E., Melnichuk, O., Ali, A., Minguet, J., Rada, J. C., Snow, C., Tikhonov, A., Brazma, A., & McEntyre, J. (2017). The BioStudies database—one stop shop for all data supporting a life sciences study. *Nucleic Acids Research*, 46(D1), D1266–D1270. <https://doi.org/10.1093/nar/gkx965> [cito:providesDataFor]

Stringer, C., & Pachitariu, M. (2025). Cellpose3: One-click image restoration for improved cellular segmentation. *Nature Methods*, 22(3), 592–599. <https://doi.org/10.1038/s41592-025-02595-5> [cito:usesMethodIn]

Author contributions

Diana Chiang: Investigation, Resources, Writing – review & editing Pavankumar Videm: Software David Lopez Tabernero: Software Maarten W. Paul: Software, Resources Alireza Heidari: Software Martin Etzrodt: Software, Resources Beatriz Serrano-Solano: Conceptualization, Software, Data curation, Writing – review & editing Leonid Kostyrykin: Conceptualization, Software, Data curation, Writing – original draft, Supervision