

Enhancing the image analysis community in Galaxy

Leonid Kostrykin^{1,2}, Tatiana Woller³, Matúš Kalaš⁴, Bugra Özdemir⁵, Rafael Andrade Bueno³, Yi Sun⁶, Anup Kumar⁷, Nadia Goué⁸, Bjoern Gruening⁷, Beatriz Serrano-Solano⁵, and Anne Fouilloux⁹

1 Biomedical Computer Vision Group, Heidelberg University, BioQuant, IPMB, 69120 Heidelberg, Germany **2** ELIXIR Germany, Institute of Bio- and Geosciences (IBG-5) – Computational Metagenomics, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany **3** Vlaams Instituut voor Biotechnologie, Datacore, Technologiepark-Zwijnaarde 75, 9052 Gent; ELIXIR Belgium **4** ELIXIR Norway, and Department of Informatics, University of Bergen, Norway **5** Euro-Biolmaging ERIC Bio-Hub, European Molecular Biology Laboratory (EMBL) Heidelberg, Meyerhofstraße 1, 69117 Heidelberg, Germany **6** European Molecular Biology Laboratory (EMBL) Heidelberg, Meyerhofstraße 1, 69117 Heidelberg, Germany **7** Department of Computer Science, University of Freiburg, Georges-Köhler-Allee 106, 79110 Freiburg, Germany **8** Université Clermont Auvergne, AuBi platform, F-63000 Clermont-Ferrand, France **9** Simula Research Laboratory, Oslo, Norway

BioHackathon series:

[BioHackathon Europe 2023](#)

Barcelona, Spain, 2023

[Project 16 - Enhancing the image analysis community in Galaxy](#)

Submitted: 08 Apr 2024

License:

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

Published by [BioHackrXiv.org](#)

Introduction

The 2023 edition of [BioHackathon Europe](#) took place in Barcelona in November last year, where a group of people interested in image data analysis gathered around the project nr. 16, ***“Enhancing the image analysis community in Galaxy”***.

Although the [Galaxy platform](#) (The Galaxy Community, 2022) has a strong presence in omics fields like genomics and proteomics, it has yet to establish itself in the imaging community fully. This would be of great value to the scientific communities working with image data, as Galaxy provides free and open, democratised access to extensive computational resources, via common user-friendly interfaces (GUI and API), and with built-in provenance for transparency and reproducibility, fostering FAIR scientific workflows (Serrano-Solano et al., 2022). However, at the moment, **efforts** to integrate tools and create workflows for image analysis in Galaxy **are scattered** across research fields and locations.

This project’s vision is to join those efforts and go beyond, by creating **interdisciplinary image analysis resources**. By fostering collaboration among scientists and enthusiasts from different fields, we aim to create a dynamic community that focuses on image data management and analysis, both within and outside of life sciences. Our goal is to provide both professional and citizen scientists, from all domains of science (e.g. life and Earth sciences, materials science, astronomy, archaeology, or arts), with free access to state-of-the-art, high-performance image analysis.

The initiative started in the summer of 2023 when the group gathered around a newly-created **Expert Group on FAIR Image Data Workflows** within [Euro-Biolmaging](#), the European Research Infrastructure offering open access to (bio-)imaging technologies, training, and data services. Additionally, it aims to tap into the capabilities of [Simula Research Laboratory](#). This Norwegian non-profit research organisation addresses crucial scientific and engineering problems with a societal impact, focusing on five primary ICT (Information and Communication Technologies) research areas: communication systems, cryptography, scientific computing, software engineering, and machine learning. Together, these entities will foster the development of a community dedicated to crafting FAIR computational workflows specifically tailored for image analysis.

Project goals

For the BioHackathon 2023, we set a number of multifaceted goals:

- Gather all the existing resources for image data management and analysis in Galaxy (e.g., tools available in the [Galaxy ToolShed](#) (Blankenberg et al., 2014), training materials published in the [Galaxy Training Network](#) (Hiltemann et al., 2023), and workflows collected in the [WorkflowHub](#) (Goble et al., 2021)).
- Create a list of tools widely used in the image analysis field that are not yet integrated into Galaxy.
- Annotate the existing image analysis tools that are lacking annotations with [EDAM](#) (Black et al., 2022) and especially the [EDAM Bioimaging](#) ontology (Kalaš et al., 2020; Lindblad et al., 2020).
- Create a community-of-practice page in the [Galaxy Community Hub](#).
- Establish community practices and decide on regular interactions.

Results

In this section, we discuss the achievements within the BioHackathon, aligning them with the project goals. We explore the progress made addressing conventions, tool renaming, tool annotation, and the identification of tools not available in Galaxy, together with image data management and image analysis tools.

Landscape analysis of image analysis and image processing tools in Galaxy

The current state of image analysis and image processing tools within Galaxy, as of October 13, 2023, reveals an extensive and diverse landscape. A comprehensive overview provides valuable insights into the existing ecosystem and areas that need to be enhanced.

Tool Statistics

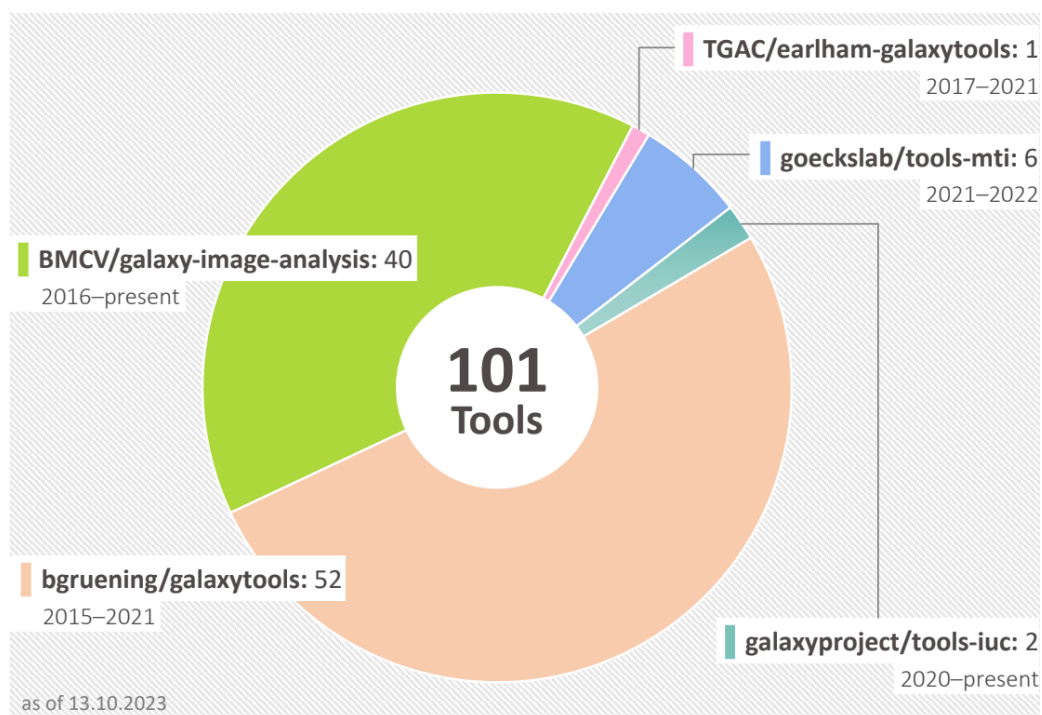


Figure 1: Distribution of tools across repositories

- **Tool count:** The “*Imaging*” category on Galaxy Europe lists 101 tools.

- **Contributors:** Approximately 23 contributors have added and maintained these tools.
- **Repositories:** The tools are distributed across 5 different GitHub repositories, presenting a need for structuration and organization.
- **Categorization challenges:** Some tools with relevance to image analysis are found in the “Interactive” category, indicating a potential need for reclassification and structure clarification.

Big pile of tools — needing structure

Following the gathering and identification of this extensive number of tools, we identified potential obstacles that could be hindering their FAIRness.

Example 1: Inconsistent tool naming schemes

Within the tool repository, there was a noticeable diversity in naming styles, with tools adopting various conventions such as the title case, sometimes capitalized, abbreviated, using camel case, *etc.* Some examples of this diversity are listed below:

- Title Case: *Anisotropic Diffusion*, *Count Objects*, ...
- Capitalized: *Binary To Points*, ...
- Abbreviated: *Binary 2 Label*, ...
- Camel Case: *ColorToGray*, *ConvertObjectsToImage*, ...
- Regular: *Convert image*, *Find edges*, *Filter segmentation*, ...
- Verb-forms vs. Nouns: *Split objects*, *Mahotas-features*, ...

During the BioHackathon, we reached an agreement on a recommended naming scheme for tools, to facilitate finding these tools (see section [Conventions](#)).

Example 2: Low-level vs. high-level tools

The distinction between low-level and high-level tools highlights the diverse needs of the Galaxy user community. Advanced users, with a good understanding of image analysis, benefit from the flexibility and control offered by low-level tools. On the other hand, users with varying levels of expertise, including those less familiar with image analysis, find high-level tools to be more accessible and efficient. This duality in tool design acknowledges the heterogeneous user base within the Galaxy community.

Example 3: Tools with unclear goals

The purpose of some tools can be unclear (*e.g. Gate Finder*, *Concatenate images*, *Compose two raw transformations*, *Add shadow effect*, ...), have typos in their documentation (*labeled* vs. *labled*) or are unrelated to imaging (*GeneSeqToFamily preparation*), emphasizing the importance of defining the purpose of each tool clearly in its metadata.

Enhancing consistency in tool names and descriptions will contribute to a more user-friendly experience.

Example 4: Duplicated or similar tools

Some tools within the Galaxy ecosystem have similar purposes but often originate from different tool suites (*e.g. CellProfiler* or *ImageJ*). Using prefixes or other consistent schemes for indication of the tool suites should facilitate identification and enhance user navigation.

Example 5: Lack of tool interoperability

Incompatibilities between tools were identified, particularly concerning input formats and inconsistent tool IDs. For example, tools like *Count Objects* historically accepted only TIFF inputs, while others, like *Concatenate images* exclusively accepted binary images without

apparent justification. To enhance user experience and interoperability, a suggestion was made for tools to support formats such as TIFF and PNG universally.

Example 6: Different versions of the same tool

Occasionally, the Galaxy tool panel may feature duplicated versions of the same tool, even if their functionalities have undergone alterations. There are different reasons for this, from typos to the way tool names are displayed in the tool panel.

Conventions

We acknowledge the existence of the [IUC recommendations](#) for tool names but decided on a somewhat different pattern, 1. to avoid too long tool names due to the inclusion of tool suites, 2. to comply with the fact that many Galaxy tools in our community are thin tool wrappers around established tools (e.g., tool suites or other Conda packages), while 3. many others implement their main functionality directly within the wrapper.

The conventions described below have a few limitations (see section [Discussion](#)).

Naming

Generally, the name of Galaxy tools in our community should be expressive and concise, while stating the purpose of the tool as precisely as possible. Consistency of the namings of Galaxy tools is important to ensure they can be found easily. To maintain consistency, we consider phrasing names as imperatives a good practice, such as “Analyze particles” or “Perform segmentation using watershed transformation”. An acknowledged exception from this rule is the names of tool wrappers of major tool suites, where the name of a tool wrapper should be chosen identically to the module or function of the tool which is wrapped (e.g., “MaskImage” in CellProfiler).

Tool description

If a Galaxy tool is a thin tool wrapper (e.g, part of a major tool suite), then the name of the wrapped tool (and only the name of the wrapped tool, subsequent to the term “with” as in “with Bioformats”) should be used as the description of the tool (further examples include “with CellProfiler”, “with ImageJ2”, “with ImageMagick”, “with SpyBOAT”, “with SuperDSM”). This ensures that the tool is found by typing the name of the wrapped tool into the “Search” field on the Galaxy interface. The tool description should be empty if a tool is either not part of a major tool suite, or the main functionality of the tool is implemented in the wrapper.

Annotations

We point out that there is a global list of precedential annotations with [Bio.tools](#) identifiers (Ison et al., 2019) in Galaxy (see [mappings](#)), which may outweigh the annotations made in the XML specification of a Galaxy tool (and thus the annotations of a tool reported within the web interface of Galaxy might be divergent). However, since the precedential annotations are subject to possible changes and to avoid redundant work, we do not aim to reflect those in our specifications (those which we make in the XML specifications of Galaxy tools).

Tool annotation

The annotations of tools within the Galaxy Image Community involve the use of the EDAM ontology and Bio.tools. To recognize the role of [BioImage Informatics Index \(BIII\)](#) as the registry of tools, workflows, training, and example data for bioimage analysis (Paul-Gilloteaux et al., 2021), we agreed to use also BIII identifiers in Galaxy, taking into account the BIII’s use of annotations with EDAM Bioimaging.

To fully reflect this decision, we included a new metadata field in the Galaxy tool wrappers that enables the annotation of tools with BIII IDs (see [PR](#)).

We annotated all the tools with Bio.tools IDs, EDAM concepts, and BIII IDs, and adhered to the agreed conventions in the following pull requests:

- <https://github.com/BMCV/galaxy-image-analysis/pull/71>
- <https://github.com/bgruening/galaxytools/pull/1346>
- <https://github.com/goeckslab/tools-mti/pull/53>
- <https://github.com/galaxyproject/tools-iuc/pull/5572>
- <https://github.com/usegalaxy-eu/usegalaxy-eu-tools/pull/641>

Thanks to these annotations and in collaboration with Project 25 "*Increasing the findability, visibility, and impact of Galaxy tools for specialised scientific Communities*" (Zierep et al., 2024), we gathered a list of tools that are of interest for the imaging community.

We also started the discussion to facilitate the Bio.tools annotations when wrapping a tool for Galaxy using the [Galaxy Language Server](#) (see [issue](#)).

Data management tools

Improving data management is crucial when it comes to improving the image analysis experience in Galaxy. During the BioHackathon, we worked on adding support for the [Zarr format](#) ([PR](#)), and a beta version of OME-Zarr ([PR](#)) (Moore et al., 2023), with plans for adding metadata information and checks in the future. This will also facilitate adding support for other community-driven standards, such as [NCZarr](#) and [GeoZarr](#).

Tools were added to convert NetCDF to NCZarr, and various proprietary image file formats to OME-Zarr ([PR](#)), addressing specific image analysis data formats for better data handling in Galaxy.

In addition, a tool to fetch image data directly from the [BioImage Archive \(BIA\)](#) has been implemented ([PR](#)), and tools for data conversion have been included ([PR](#)).

AI-powered image analysis

The integration of AI-powered image analysis in Galaxy involved leveraging models from the [BioImage Model Zoo \(BMZ\)](#) (Ouyang et al., 2022) and conducting an analysis of [Zero-CostDL4Mic](#) dependencies.

We explored the possibility of consuming models from the [BMZ](#) in Galaxy, with specific examples for live cell segmentation and cellular nucleus segmentation boundary models.

Community

We updated the Galaxy subdomain imaging.usegalaxy.eu to contain the list of training materials related to imaging ([PR](#)). We also agreed on renaming our community to "image community", thus also renaming the subdomain to image.usegalaxy.eu.

Discussion

The project has made good progress in gathering existing image analysis tools, creating a list of tools to integrate, and adding annotations to tools that needed them.

Future work

Looking ahead, we plan to continue this work by collaborating closely with the Galaxy, EDAM (Bio-)imaging, and BioImage Model Zoo communities. The long-term goal is to create a

solid community that supports the creation and reuse of reproducible and FAIR image analysis workflows, amongst diverse scientific disciplines.

Community development

We welcome anyone interested, from experienced Galaxy tool developers and image analysts to those who are new to the field, to join our community (info@eurobioimaging.eu) and monthly meetings on the third Wednesday of the month at 4 pm CE(S)T.

Improving conventions

We are aware of the following limitations of our conventions for tool naming, tool descriptions, and annotation (see section [Conventions](#)), which we seek to address in the future:

1. If a Galaxy tool has a Bio.tools identifier included in its XML specification, then its EDAM annotation can be inherited from the Bio.tools record. The EDAM annotations given in the XML specification of a Galaxy tool are precedential over the EDAM annotations in Bio.tools. However, depending on the evolving community guidelines, the precedence could change in the future.
2. Our primary motivation for prominently exposing the name of the wrapped tool in the tool description has been to account for the fact that tools with similar functionality (and hence similar names) exist across different tool suites. However, they can also exist across different plugins within the same tool suite (such as ImageJ2). To account for that, we might want to take the suggested scheme one step further and include the name of the corresponding plugin in the tool description (as in “with ImageJ2, bUnwrapJ”).

Acknowledgements

This work was developed as part of BioHackathon Europe 2023. We thank the contributions from Project 25 (Zierep et al., 2024). This work was supported by [ELIXIR](#), the research infrastructure for life science data. The authors utilised the language model ChatGPT developed by OpenAI to assist in structuring and drafting this text.

References

- Black, M., Lamothe, L., Hager Eldakrouy, Kierkegaard, M., Ankita Priya, Machinda, A., Khanduja, U. S., Drashti Patoliya, Rashika Rathi, Tawah Peggy Che Nico, Umutesi, G., Blankenburg, C., Op, A., Chieke, P., Omodolapo Babatunde, Laurie, S., Neumann, S., Schwämmle, V., Kuzmin, I., ... Kalaš, M. (2022). EDAM: The bioscientific data analysis ontology (update 2021)[version 1; not peer reviewed]. *F1000Research*, 11(ISC Comm J), 1. <https://doi.org/10.7490/f1000research.1118900.1> [cito:citesAsAuthority] [cito:extends]
- Blankenberg, D., Von Kuster, G., Bouvier, E., Baker, D., Afgan, E., Stoler, N., Team, G., Taylor, J., & Nekrutenko, A. (2014). Dissemination of scientific software with Galaxy ToolShed. *Genome Biology*, 15, 1–3. <https://doi.org/10.1186/gb4161> [cito:usesDataFrom] [cito:citesAsAuthority]
- Goble, C., Soiland-Reyes, S., Bacall, F., Owen, S., Williams, A., Eguinoa, I., Driesbeke, B., Leo, S., Pireddu, L., Rodríguez-Navas, L., Fernández, J. M., Capella-Gutiérrez, S., Ménager, H., Grüning, B., Serrano-Solano, B., Ewels, P., & Coppens, F. (2021). *Implementing FAIR Digital Objects in the EOSC-Life Workflow Collaboratory*. <https://doi.org/10.5281/ZENODO.4605654> [cito:usesDataFrom] [cito:citesAsAuthority]
- 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., Bras, Y. L., Mehta, S., Syme, A., Coppens, F., Driesbeke, B., Soranzo, N., Bacon, W., ... and, B. B. (2023). Galaxy

Training: A powerful framework for teaching! *PLoS Computational Biology*, 19(1), e1010752. <https://doi.org/10.1371/journal.pcbi.1010752> [cito:usesDataFrom] [cito:citesAsAuthority]

Ison, J., Ienasescu, H., Chmura, P., Rydza, E., Ménager, H., Kalaš, M., Schwämmle, V., Grüning, B., Beard, N., Lopez, R., Duvaud, S., Stockinger, H., Persson, B., Vařeková, R. S., Raček, T., Vondrášek, J., Peterson, H., Salumets, A., Jonassen, I., ... Brunak, S. (2019). The bio.tools registry of software tools and data resources for the life sciences. *Genome Biology*, 20(1). <https://doi.org/10.1186/s13059-019-1772-6> [cito:citesAsAuthority] [cito:usesDataFrom]

Kalaš, M., Plantard, L., Lindblad, J., Jones, M., Sladoje, N., Kirschmann, M. A., Chessel, A., Scholz, L., Rössler, F., Sáenz, L. N., Mariscal, E. G. de, Bogovic, J., Dufour, A., Heiligenstein, X., Waithe, D., Domart, M.-C., Karreman, M., Plas, R. V. de, Haase, R., ... We are welcoming new contributors! (2020). EDAM-bioimaging: the ontology of bioimage informatics operations, topics, data, and formats (update 2020) [version 1; not peer reviewed]. *F1000Research*, 9(ELIXIR), 162. <https://doi.org/10.7490/f1000research.1117826.1> [cito:citesAsAuthority] [cito:usesDataFrom] [cito:extends]

Lindblad, J., Plantard, L., Jones, M., Sladoje, N., Domart, M.-C., Karreman, M., Mariscal, E. G. de, Sáenz, L. N., Kyprianou, C., Muñoz-Barrutia, M. A., Plas, R. V. de, Heiligenstein, X., Madunić, I. V., Karaica, D., Sage, D., Haase, R., and all contributors to the previous versions, & Kalaš, M. (2020). *EDAM Bioimaging: The ontology of (bio-)image informatics and machine learning* (Version alpha06) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.3695725> [cito:usesDataFrom]

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:citesAsAuthority] [cito:usesMethodIn]

Ouyang, W., Beuttenmueller, F., Gómez-de-Mariscal, E., Pape, C., Burke, T., García-López-de-Haro, C., Russell, C., Moya-Sans, L., de-la-Torre-Gutiérrez, C., Schmidt, D., Kutra, D., Novikov, M., Weigert, M., Schmidt, U., Bankhead, P., Jacquemet, G., Sage, D., Henriques, R., Muñoz-Barrutia, A., ... Kreshuk, A. (2022). BioImage Model Zoo: A Community-Driven Resource for Accessible Deep Learning in BioImage Analysis. *bioRxiv*. <https://doi.org/10.1101/2022.06.07.495102> [cito:citesAsAuthority] [cito:usesDataFrom]

Paul-Gilloteaux, P., Tosi, S., Hériché, J.-K., Gaignard, A., Ménager, H., Marée, R., Baecker, V., Klemm, A., Kalaš, M., Zhang, C., Miura, K., & Colombelli, J. (2021). Bioimage analysis workflows: Community resources to navigate through a complex ecosystem [version 1; peer review: 2 approved]. *F1000Research*, 10, 320. <https://doi.org/10.12688/f1000research.52569.1> [cito:citesAsAuthority] [cito:usesDataFrom]

Serrano-Solano, B., Fouilloux, A., Eguinoa, I., Kalaš, M., Grüning, B., & Coppens, F. (2022). Galaxy: A Decade of Realising CWFR Concepts. *Data Intelligence*, 4(2), 358–371. https://doi.org/10.1162/dint_a_00136 [cito:citesAsSourceDocument]

The Galaxy Community. (2022). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Research*, 50(W1), W345–W351. <https://doi.org/10.1093/nar/gkac247> [cito:citesAsAuthority] [cito:extends]

Zierep, P., Batut, B., Kalaš, M., Kayikcioglu, T., Nasr, E., Soranzo, N., Thang, W. C., Wang, J., & Gustafsson, O. J. R. (2024). How to increase the findability, visibility, and impact of Galaxy tools for your scientific community. *BioHackrXiv*. <https://doi.org/10.37044/osf.io/qjbxk> [cito:usesMethodIn] [cito:obtainsSupportFrom]