



HAL
open science

Bioimage analysis workflows: community resources to navigate through a complex ecosystem

Perrine Paul-Gilloteaux, Sébastien Tosi, Jean-Karim Hériché, Alban Gaignard, Hervé Ménager, Raphaël Marée, Volker Baecker, Anna Klemm, Matúš Kalaš, Chong Zhang, et al.

► To cite this version:

Perrine Paul-Gilloteaux, Sébastien Tosi, Jean-Karim Hériché, Alban Gaignard, Hervé Ménager, et al.. Bioimage analysis workflows: community resources to navigate through a complex ecosystem. F1000Research, 2021, 10, pp.320. 10.12688/f1000research.52569.1 . hal-03208600

HAL Id: hal-03208600

<https://nantes-universite.hal.science/hal-03208600>

Submitted on 21 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



OPINION ARTICLE

Bioimage analysis workflows: community resources to navigate through a complex ecosystem [version 1; peer review: 2 approved]

Perrine Paul-Gilloteaux ^{1,2}, Sébastien Tosi ³, Jean-Karim Hériché⁴, Alban Gaignard¹, Hervé Ménager^{5,6}, Raphaël Marée⁷, Volker Baecker ⁸, Anna Klemm ⁹, Matúš Kalaš ¹⁰, Chong Zhang¹¹, Kota Miura¹², Julien Colombelli ³

¹Université de Nantes, CNRS, INSERM, l'institut du thorax, Nantes, F-44000, France

²Université de Nantes, CHU Nantes, Inserm, CNRS, SFR Santé, Inserm UMS 016, CNRS UMS 3556, Nantes, F-44000, France

³Institute for Research in Biomedicine, IRB Barcelona, Barcelona Institute of Science and Technology, BIST, Barcelona, Spain

⁴Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Heidelberg, 69117, Germany

⁵Hub de Bioinformatique et Biostatistique, Département Biologie Computationnelle, Institut Pasteur, USR 3756, CNRS, Paris, 75015, France

⁶CNRS, UMS 3601, Institut Français de Bioinformatique, IFB-core, Evry, 91000, France

⁷Montefiore Institute, University of Liège, Liège, Belgium

⁸Montpellier Ressources Imagerie, BioCampus Montpellier, CNRS, INSERM, University of Montpellier, Montpellier, F-34000, France

⁹BioImage Informatics Facility, SciLifeLab, Stockholm, Sweden

¹⁰Computational Biology Unit, Department of Informatics, University of Bergen, Bergen, Norway

¹¹Department of Information and Communication Technologies, University Pompeu Fabra, Barcelona, Spain

¹²Nikon Imaging Center, University of Heidelberg, Heidelberg, Germany

V1 First published: 26 Apr 2021, 10:320
<https://doi.org/10.12688/f1000research.52569.1>

Latest published: 26 Apr 2021, 10:320
<https://doi.org/10.12688/f1000research.52569.1>

Abstract

Workflows are the keystone of bioimage analysis, and the NEUBIAS (Network of European BioImage Analysts) community is trying to gather the actors of this field and organize the information around them. One of its most recent outputs is the opening of the F1000Research NEUBIAS gateway, whose main objective is to offer a channel of publication for bioimage analysis workflows and associated resources. In this paper we want to express some personal opinions and recommendations related to finding, handling and developing bioimage analysis workflows.

The emergence of "big data" in bioimaging and resource-intensive analysis algorithms make local data storage and computing solutions a limiting factor. At the same time, the need for data sharing with collaborators and a general shift towards remote work, have created new challenges and avenues for the execution and sharing of bioimage analysis workflows.

These challenges are to reproducibly run workflows in remote

Open Peer Review

Approval Status

	1	2
version 1		
26 Apr 2021	view	view

1. **Katy Wolstencroft** , Leiden University, Leiden, The Netherlands

2. **Assaf Zaritsky** , Ben-Gurion University of the Negev, Beer Sheva, Israel

Any reports and responses or comments on the article can be found at the end of the article.

environments, in particular when their components come from different software packages, but also to document them and link their parameters and results by following the FAIR principles (Findable, Accessible, Interoperable, Reusable) to foster open and reproducible science.

In this opinion paper, we focus on giving some directions to the reader to tackle these challenges and navigate through this complex ecosystem, in order to find and use workflows, and to compare workflows addressing the same problem. We also discuss tools to run workflows in the cloud and on High Performance Computing resources, and suggest ways to make these workflows FAIR.

Keywords

Bioimage analysis, workflows, components, collections, FAIR principles, NEUBIAS, remote computing, scientific workflow management systems, knowledge database



This article is included in the **NEUBIAS - the Bioimage Analysts Network gateway**.

Corresponding author: Perrine Paul-Gilloteaux (perrine.paul-gilloteaux@univ-nantes.fr)

Author roles: **Paul-Gilloteaux P:** Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Tosi S:** Conceptualization, Funding Acquisition, Project Administration, Writing – Review & Editing; **Hériché JK:** Conceptualization, Writing – Review & Editing; **Gaignard A:** Conceptualization, Writing – Review & Editing; **Ménager H:** Conceptualization, Writing – Review & Editing; **Marée R:** Writing – Review & Editing; **Baecker V:** Writing – Review & Editing; **Klemm A:** Writing – Review & Editing; **Kalaš M:** Conceptualization, Writing – Review & Editing; **Zhang C:** Conceptualization; **Miura K:** Conceptualization, Funding Acquisition; **Colombelli J:** Conceptualization, Funding Acquisition

Competing interests: No competing interests were disclosed.

Grant information: PPG and VB acknowledge France-BioImaging infrastructure supported by the French National Research Agency (ANR-10-INBS-04). AG and HM acknowledge l'Institut Français de Bioinformatique supported by the French National Research Agency (ANR-11-INBS-0013).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Paul-Gilloteaux P *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Paul-Gilloteaux P, Tosi S, Hériché JK *et al.* **Bioimage analysis workflows: community resources to navigate through a complex ecosystem [version 1; peer review: 2 approved]** F1000Research 2021, 10:320 <https://doi.org/10.12688/f1000research.52569.1>

First published: 26 Apr 2021, 10:320 <https://doi.org/10.12688/f1000research.52569.1>

Introduction

Workflows are the keystone of bioimage analysis,^{1,2} and the NEUBIAS community is trying to gather the actors of this field and organize the information around them. One of its most recent outputs is the opening of the F1000Research NEUBIAS gateway, whose main objective is to offer a channel of publication for bioimage analysis workflows³ and associated resources.

In this paper, we aim to express some personal opinions and recommendations related to finding, handling and developing bioimage analysis workflows.

A bioimage analysis workflow is defined as a set of computational tools assembled in a specific order to process bioimages and estimate some parameters relevant to the biological system under study. To classify these computational tools, in the NEUBIAS community, we have defined these terms: *workflows*, *components* and *collections*^{1,4} as follows. A *workflow* is built as a sequence of *components* coming from one or multiple software packages. It takes an image as input and outputs processed images, numerical values and/or annotations (e.g. biological objects outlines). A *component* is the software implementation of an image processing or analysis algorithm. We call *collection* the software package that gathers *components* and can be in the form of a generalist software platform such as ImageJ and Fiji,⁵ Icy,⁶ CellProfiler,⁷ specialized platforms, such as analyzing a specific modality of microscopy e.g. super resolution image data;⁸ or computationally optimized libraries such as ImgLib2⁹ or ITK.¹⁰ Most of the time, standalone components cannot solve complex bioimage analysis problems on their own – that is why they need to be carefully assembled.

The emergence of resource-intensive analysis algorithms, e.g. supervised machine learning with convolutional neural networks, and of "big data" in bioimaging make local data storage and computing solutions a limiting factor. At the same time, the need for data sharing with collaborators and a general shift towards remote work, have created new challenges and avenues for the execution and sharing of bioimage analysis workflows.

These challenges are to reproducibly run workflows in remote environments, in particular when their components come from different *collections*, but also to document them and link their parameters and results by following the FAIR principles (Findable, Accessible, Interoperable, Reusable)¹¹ to foster open and reproducible science.

In this opinion paper we focus on giving some directions to the reader to tackle these challenges and navigate through this complex ecosystem, in order to find and use workflows (and *components*), and to compare workflows addressing the same problem. We also discuss tools to run workflows in the cloud and on High Performance Computing (HPC) resources, and suggest ways to make these workflows FAIR.

Finding workflows or components for a specific biological problem or task

The first challenge in the creation of a workflow is to avoid duplicating the effort and being able to easily find and customize a workflow that has been used for a similar biological problem. Today, browsing the documentation of bioimage analysis tools, or asking a specific question in a generic forum, such as the newly created [Image.sc forum](#),¹² will help guide the biologist or microscopist to existing tools. We believe that while this can be a good starting point it may not be sufficient. The NEUBIAS training courses^{13,14} and the NEUBIAS Academy (see¹⁵ in this Gateway) are two of the educational resources that can also help finding and adapting existing workflows. Exposing tools and workflows in a knowledge database has also been identified as very useful by the community. [Table 1](#) illustrates some examples of such databases where bioimage analysts can reference their workflows using the proposed standardized framework and vocabulary in order to make them findable.

BIII, the BioImage Informatics Index, has been created in the context of the NEUBIAS network and with the effort of tens of volunteers. Software tools (>1343), image databases for benchmarking (>24) and training materials (>71) for bioimage analysis are referenced and curated following standards constructed by the community. The range of software tools available includes workflows (>172), specific components (>898), and collections (>302). All entries are exposed following FAIR principles and accessible for other usage. They are described using Edam Bio Imaging,¹⁶ a dedicated extension of the generalist EDAM ontology¹⁷ for bioimage analysis, bioimage informatics, and bioimaging. It is developed in a community spirit, in collaboration between numerous bioimaging experts and ontology developers. It is used in BIII to describe the applications of these tools, by describing the operations performed (such as segmentation, visualization, or lower level operation) and the field of applications of these tools such as the imaging modalities to which it can be applied (i.e. [EDAM Bioimaging Ontology](#)). EDAM Bioimaging has now a solid basis. This basis is incrementally defined at specific meetings (i.e. taggathons) where suggestions for new terms, crowd-sourced from free tags by BIII users, are inspected and moderated for inclusion, or contrasted by bioimage analysis experts when no term is found adequate.

Table 1. Workflow finder. Some examples of such databases where bioimage analysts can reference their workflows.

Workflow finder	Target audience	Link
BIII	Bio Image Analyst, Biologist, Software developer	https://biii.eu
bio.tools	Bioinformatics/Computational Biology	https://bio.tools
Quantitative-plant	Plant Biologist	https://www.quantitative-plant.org/
Bio Image Model Zoo	Bio Image Analyst, Biologist, Focused on AI pretrained models	https://bioimage.io

Table 2. Example of websites to compare existing workflows on reference datasets

Benchmarking site	Link	Purpose
BIAFLOWS	https://biaflows.neubias.org/#/projects	Allows live testing of workflows
Grand-Challenges	https://grand-challenge.org/challenges/	Lists open challenge and results
Kaggle	https://www.kaggle.com/c/data-science-bowl-2018	One shot challenge for nuclei. Very generalist challenge platform

Similar initiatives exist, either for a broader range of applications, for example bio.tools,^{18,19} which has gathered more than 20000 software tools in the full range of life science applications, or for more specific application topics, for example [Quantitative Plant](https://www.quantitative-plant.org/), which focusses on tools for the analysis of image data of plants²⁰ or [BioImage.io](https://bioimage.io) for pre trained AI deep learning models.

By feeding the description of a workflow in the knowledge database BIII (following the [recommendation provided](#)), and thanks to workflow/tools interoperability standards, these workflows can be found by other bioimage analysts or automatically discovered and consumed by other registries, such as bio.tools, to reach a broader community.

Comparing workflows

Once a candidate workflow has been found, the natural question is then if it is the best solution for the particular task one wants to solve. [Table 2](#) shows three examples of resources comparing workflows.

[BIAFLOWS](https://biaflows.neubias.org/)²¹ is an open-source web platform to reproducibly deploy and publicly benchmark image analysis workflows with a strong focus on microscopy bioimages. The database stores scientific datasets, metadata, and versioned image analysis workflows with parameters optimized for the corresponding datasets. The workflows can be run remotely. The results (e.g. object annotations) from different workflows (or from runs with different parameter values) can be visualized remotely as an overlay on the original images. When the images hold reference annotations, the results are automatically benchmarked by commonly adopted benchmark metrics targeting one of the nine currently supported [problem classes](#). The benchmark metrics of each workflow run can be browsed per image or as overall statistics over whole datasets. BIAFLOWS brings an automated mechanism leveraging DockerHub to encapsulate, version and make the workflows and their complete execution environment available upon every new release. Overall, BIAFLOWS enables integration and web-based evaluation of heterogeneous workflows originally written for diverse languages and libraries.

The [Grand Challenge](#) is a website cataloguing a set of challenges, focusing mostly on medical imaging. These challenges are usually hosted by a conference such as IEEE ISBI and run as an annual edition with specific reporting^{22,23} and they gather and evaluate competing workflows to solve a common bioimage analysis task. In the microscopy imaging communities, a particular effort has gone towards nuclei segmentation with the goal of developing a universal nuclei segmenter that works across different imaging modalities, as for instance with the Kaggle Data Science Bowl of 2018, providing a considerable amount of annotated data.²⁴

Towards reproducibility and interoperability in bioimage analysis

The current paradigm for bioimage analysts is to create workflows using a single platform or application, aka *collection*, for example Fiji,²⁵ CellProfiler,⁷ or Icy.⁶

By allowing the possibility to script a workflow calling their components with simplified programming language, these platforms offer ways to share and document the workflows for other users. Besides script creation, there are also options to create sharable elements with no programming skills, as detailed in.²⁶ This only requires the deployment of the software package to be run.

This reliance on graphical user interfaces favors the development of components built for a single *collection*. While this has stimulated the gathering of active communities around these *collections*, the coexistence of many multifunctional collections that are developed independently is not ideal for cloud deployment and FAIR principles. The graphical user interfaces are often not compatible with the type of remote computing offered by cloud technologies and the large collections contain largely overlapping components that are however not interoperable with each other. These *collections* therefore do not offer a unified and granular way of describing an image processing workflow. This situation also often requires users to learn multiple platforms to be able to complete their workflows. Code notebooks such as [CodeOcean capsules](#) or [Jupyter notebooks](#) offers also an easy access to cloud computing or HPC, but several aspect of workflow management are also left to the user, in particular data provenance.

At the same time that the field is shifting to running workflows in the cloud or high performance computing environments, there also comes the need to run more complex workflows integrating tools and data coming from different life science fields, such as genomics or proteomics data, or spatial transcriptomics. In addition to the integration of component from different communities, one can face the challenge to run again a previously created workflow and encounter versioning problems, with time and evolution of software packages and *component* versions. Specific configuration issues also make tedious the portability for the execution of a workflow from an environment to another, such as moving between HPCs or cloud computing platforms. While the use of virtual machines accessible from a web browser to emulate a personal desktop experience may be seductive, the bioimage analysis community should not isolate itself from other communities, and in particular not from bioinformatics community. Several bioinformatics communities have already started to tackle these issues through the use of scientific workflow management systems (SWMS)^{27,28} and standardized software packaging practices.²⁹ These SWMS have also the advantage to tackle standardized workflow description, machine-readable as well as human readable, for FAIR principles. In comparison, the usual documentation, provided when documenting a workflow in bio image analysis current practices, is usually addressed to humans (which is already laudable and not yet common practice).

One of the key elements to enable reproducibility and portability is containerization and software packaging that facilitates the reliable creation of containers. Containerization consists in embedding a piece of software, and all its dependencies and specific configuration in one file called a container image, so that the software can run consistently across different computing environments. [Table 3](#) shows examples of workflow management systems with usage in bioimage analysis. This containerization can be performed at the level of each individual workflow component (such as in Galaxy^{30,31}), or for complete workflows (such as in BIAFLOWS²¹ or [coming to grand-challenges](#)). Biocontainers³² is proposing a standard and recipes for these containerizations, as well as a marketplace for the containers, today mostly for -omics data processing.

Table 3. Some example of scientific workflow management systems.

Name (SWMS)	Example of use in bioimage analysis	Reference or link
Galaxy	²	³¹
NextFlow	^{37,38}	³⁹
SnakeMake	⁴⁰	⁴¹
BIAFLOWS	https://biaflows.neubias.org/#/projects (click Try online)	²¹
BioImageIT	https://bioimageit.github.io/bioimageit_gui/tutorial_pipeline.html	https://bioimageit.github.io/#/
Knime	²	⁴²

As a community we need to join this effort for a better exploitation and reproducibility by other communities of the imaging data produced by our workflows. One of the particularities of workflows in bioimage analysis is the need for visual and accurate feedback at critical workflow steps. This human-in-the-loop requirement has so far prevented the community from using SWMS more widely. But this is now changing as image processing tools and visual feedback are now getting incorporated in SWMS.^{21,31,33}

Towards findability and accessibility of image analysis workflows

At a general level in life-science and not specifically for the bioimage analysis community, coordination efforts are ongoing in the direction of the “FAIRification” of workflows, but also the ease to access HPC resources to run them. They are led by European Research Infrastructures, such as ELIXIR.³⁴ ELIXIR is an intergovernmental organization that aims to coordinate the resources offered nationally for databases, software tools and access to cloud storage and HPCs, and associated training material. BIII, the finder tool mentioned above, is now for example part of the recommended interoperability resources. EOSC-Life is an ESFRI cluster project involving the 13 biomedical research infrastructures whose goal is to create an open, digital and collaborative space for biological and medical research in the European Open Science Cloud. This includes making image data and image processing and analysis workflows compliant with the FAIR principles, while enabling interoperability with tools and data from other life science domains as mandated by the European Commission. Galaxy³⁰ has been identified and selected as an aggregator of communities, and selected by EOSC-Life as an exemplary workflow management system that promotes cross-communities interoperability in the cloud. This does not mean that the bioimage analysis community needs to restrict to this particular choice, but it means that the workflows have to be compatible with this choice and to prepare for a future where local compute resources will not anymore be used to run a workflow.

To ease this interoperability, a common description needs to be defined, in order to be able to make workflows interoperable and compatible with different infrastructure environments. The description of a workflow is different from the workflow itself: it is a human- and machine-readable description following standard syntax or vocabularies that will allow this workflow to be FAIR.³⁵ A workflow should be associated with a standardized description (such as a unique identifier for the workflow itself, their component, but also their creator) and a description of its constitutive components and their configuration. The researchers who created the workflow can be identified by their ORCID. The **Common Workflow Language**³⁶ could be used as a standard to describe workflows in an interoperable way since it has reached a sufficient level of maturation and flexibility. To further facilitate their findability by web search and indexing engines, lightweight metadata can be provided through the *Schema.org* controlled vocabularies or *Bioschemas*, a specific extension for Life Science resources.

Galaxy is one of many SWMS, a more exhaustive list curated by the Common Workflow Language organization can be found [here](#). **Table 3** focuses on SWMS used in the bioimage analysis field, and it details their specificities. These specificities tend to support the message that the effort should not be in trying to push the implementation of workflows in only one solution, but rather to allow and ease the portability of workflows in multiple frameworks and execution environments, an approach supported by initiatives CWL. We then argue that these standards are key to facilitate the workflow ecosystems and further promote open and reproducible sciences.

Conclusion

The field of bioimage analysis, partly thanks to the NEUBIAS community, has been recently consolidated. Its community has contributed to the emergence of new tools to find, launch, compare and learn how to use and customize image analysis workflows. We believe that today the field has become mature enough to contribute to the general open science effort in life science and to enable better access to data and computational resources. This effort should help promote workflow sharing and reuse and a wider data integration and interoperability. We deeply encourage the bioimage analyst community, and by extension the associated software developer community, to sustain this effort and to rely on these tools. In particular, we encourage bioimage analysts to describe their workflows thoroughly by following the CWL standard, index them in BIII, and share them in SWMS such as BIAFLOWS compatible with Galaxy.

Data availability

No data are associated with this article.

Acknowledgements

This publication was supported by COST (European Cooperation in Science and Technology) under COST Action NEUBIAS (CA15124). The authors wish to thank all contributors for their continuous and invaluable input to the aforementioned online community resources developed under the framework of NEUBIAS (BIII, BIAFLOWS) and whose contributions are described and acknowledged in their respective publications.

References

1. Miura K, Paul-Gilloteaux P, Tosi S, *et al.*: **Workflows and Components of Bioimage Analysis**. In: *Bioimage Data Analysis Workflows*. Miura K, Sladoje N Éd. Cham: Springer International Publishing; 2020, p. 1–7.
2. Wollmann T, Erfle H, Eils R, *et al.*: **Workflows for microscopy image analysis and cellular phenotyping**. *J. Biotechnol.* nov. 2017; vol. **261**, p. 70–75.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Cimini BA, *et al.*: **The NEUBIAS Gateway: a hub for bioimage analysis methods and materials**. *F1000Res.* juin 2020; vol. **9**, p. 613.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Tosi KM, Tosi S: **Epilogue: A Framework for Bioimage Analysis**. In: *Standard and Super-Resolution Bioimaging Data Analysis*. Wheeler A, Henriques R, Éd. Chichester, UK: John Wiley & Sons, Ltd; 2017, p. 269–284.
5. Schneider CA, Rasband WS, Eliceiri KW: **NIH Image to ImageJ: 25 years of image analysis**. *Nat. Methods.* juill. 2012; vol. **9**, no 7, p. 671–675.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. de Chaumont F, *et al.*: **Icy: an open bioimage informatics platform for extended reproducible research**. *Nat. Methods.* juill. 2012; vol. **9**, no 7, p. 690–696.
[PubMed Abstract](#) | [Publisher Full Text](#)
7. Carpenter AE, *et al.*: **CellProfiler: image analysis software for identifying and quantifying cell phenotypes**. *Genome Biol.* 2006; vol. **7**, no 10, p. R100.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Levet F, *et al.*: **SR-Tesseler: a method to segment and quantify localization-based super-resolution microscopy data**. *Nat. Methods.* nov. 2015; **12**(11): 1065–1071.
[PubMed Abstract](#) | [Publisher Full Text](#)
9. Pietzsch T, Preibisch S, Tomančák P, *et al.*: **ImgLib2—generic image processing in Java**. *Bioinformatics.* nov. 2012. vol. **28**, no 22, p. 3009–3011.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. McCormick M, Liu X, Jomier J, *et al.*: **ITK: enabling reproducible research and open science**. *Front. Neuroinformatics.* 2014; vol. **8**.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
11. Wilkinson MD, *et al.*: **The FAIR Guiding Principles for scientific data management and stewardship**. *Sci. Data.* déc., 2016; vol. **3**, no 1, p. 160018.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Rueden CT, *et al.*: **Scientific Community Image Forum: A discussion forum for scientific image software**. *PLOS Biol.* juin 2019; vol. **17**, no 6, p. e3000340.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. Miura K, Sladoje N: **Bioimage Data Analysis Workflows**. *Learn. Mater. Biosci.*
14. Kota M: *Bioimage data analysis*. Wiley-VCH; 2016.
15. Highlights from the 2016-2020 NEUBIAS Training Schools for Bioimage Analysts - A Success Story.. In: Martins GG1*, Cordelières FP2*, Colombelli J3, *et al.*: **Highlights from the 2016-2020 NEUBIAS Training Schools for Bioimage Analysts - A Success Story**.
[Publisher Full Text](#)
16. Kalaš M, *et al.*: **EDAM-bioimaging: the ontology of bioimage informatics operations, topics, data, and formats (2019 update)**. *F1000Res.* 2019; vol. **8**.
[Publisher Full Text](#)
17. Ison J, *et al.*: **EDAM: an ontology of bioinformatics operations, types of data and identifiers, topics and formats**. *Bioinformatics.* mai 2013; vol. **29**, no 10, p. 1325–1332.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
18. Ison J, *et al.*: **Tools and data services registry: a community effort to document bioinformatics resources**. *Nucleic Acids Res.* janv. 2016; vol. **44**, no D1, p. D38–D47.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Ison J, *et al.*: **The bio.tools registry of software tools and data resources for the life sciences**. *Genome Biol.* déc. 2019; vol. **20**, no 1, p. 164.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Lobet G, Draye X, Périlleux C: **An online database for plant image analysis software tools**. *Plant Methods.* 2013; vol. **9**, no 1, p. 38.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. Rubens U, *et al.*: **BIAFLOWS: A Collaborative Framework to Reproducibly Deploy and Benchmark Bioimage Analysis Workflows**. *Patterns.* juin 2020, vol. **1**, no 3, p. 100040.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
22. Chenouard N, *et al.*: **Objective comparison of particle tracking methods**. *Nat. Methods* mars 2014; vol. **11**, no 3, p. 281–289.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
23. Ulman V, *et al.*: **An objective comparison of cell-tracking algorithms**. *Nat. Methods.* déc. 2017; vol. **14**, no 12, p. 1141–1152.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
24. Caicedo JC, *et al.*: **Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl**. *Nat. Methods.* déc. 2019; vol. **16**, no 12, p. 1247–1253.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. Schindelin J, *et al.*: **Fiji: an open-source platform for biological-image analysis**. *Nat. Methods.* juill. 2012; vol. **9**, no 7, p. 676–682.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
26. Miura K, Nørrelykke SF: **Reproducible image handling and analysis**. *EMBO J.* févr. 2021; vol. **40**, no 3.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. Leipzig J: **A review of bioinformatic pipeline frameworks**. *Brief. Bioinform.* mars 2016; p. bbw020.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
28. Gaignard A, Skaf-Molli H, Belhajjame K: **Findable and reusable workflow data products: A genomic workflow case study**. *Semantic Web.* août 2020; vol. **11**, no 5, p. 751–763.
[Publisher Full Text](#)
29. Gruening B, *et al.*: **Recommendations for the packaging and containerizing of bioinformatics software**. *F1000Res.* mars 2019; vol. **7**, p. 742.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. Afgan E, *et al.*: **The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update**. *Nucleic Acids Res.* juill. 2016; vol. **44**, no W1, p. W3–W10.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
31. Jalili V, *et al.*: **The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update**. *Nucleic Acids Res.* juill. 2020; vol. **48**, no W1, p. W395–W402.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. da Veiga Leprevost F, *et al.*: **BioContainers: an open-source and community-driven framework for software standardization**. *Bioinformatics.* août 2017; vol. **33**, no 16, p. 2580–2582.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
33. Dietz C, *et al.*: **Integration of the ImageJ Ecosystem in KNIME Analytics Platform**. *Front. Comput. Sci.* mars 2020; vol. **2**, p. 8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Harrow J, *et al.*: **ELIXIR-EXCELERATE: establishing Europe's data infrastructure for the life science research of the future**. *EMBO J.* mars 2021; vol. **40**, no 6.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. Goble C, *et al.*: **FAIR Computational Workflows**. *Data Intell.* janv. 2020; vol. **2**, no 1-2, p. 108–121.
[Publisher Full Text](#)
36. Amstutz P, *et al.*: **Common Workflow Language, v1.0**. *figshare.* 2016; p. 5921760 Bytes.
[Publisher Full Text](#)
37. Harshil P, Karishma M, Angelova M, *et al.*: **nf-core/imcyto: nf-core/imcyto v1.0.0 - Platinum Panda**. *Zenodo.* 2020.
38. van Maldegem F, *et al.*: **Characterisation of tumour immune microenvironment remodelling following oncogene inhibition in preclinical studies using an optimised imaging mass cytometry workflow**. *Cancer Biology, preprint*, févr. 2021.
[Publisher Full Text](#)
39. Ewels PA, *et al.*: **The nf-core framework for community-curated bioinformatics pipelines**. *Nat. Biotechnol.* mars 2020; vol. **38**, no 3, p. 276–278.
[PubMed Abstract](#) | [Publisher Full Text](#)
40. Schmie C, Steinbach P, Pietzsch T, *et al.*: **An automated workflow for parallel processing of large multiview SPIM recordings**. *Bioinformatics.* avr. 2016; vol. **32**, no 7, p. 1112–1114.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
41. Mölder F, *et al.*: **Sustainable data analysis with Snakemake**. *F1000Res.* janv. 2021; vol. **10**, p. 33.
[Publisher Full Text](#)
42. Berthold MR, *et al.*: **KNIME - the Konstanz information miner: version 2.0 and beyond**. *ACM SIGKDD Explor. Newsl.* nov. 2009; vol. **11**, no 1, p. 26–31.
[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 04 June 2021

<https://doi.org/10.5256/f1000research.55867.r84664>

© 2021 Zaritsky A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Assaf Zaritsky 

Department of Software and Information Systems Engineering, Ben-Gurion University of the Negev, Beer Sheva, Israel

In this Opinion Article Paul-Gilloteaux and representatives of the Network of European BioImage Analysts (NEUBIAS) present their views on the current challenges and solution in using bioimage analysis workflows – modular image analysis pipelines that are used to process bioimaging data. Specifically, the authors survey current approaches for finding the most appropriate workflows for a given task, evaluating and comparing workflows, reproducing results, sharing code and executing workflows remotely. The opinion is dealing with a timely and important topic, and is mostly well written and easy to follow and an enjoyable read.

The authors present several alternative solutions for each challenge, highlighting the contribution made by NEUBIAS. This is totally legitimate since this is an opinion piece written from NEUBIAS perspective. However, I do think this point should be emphasized by including the Network of European BioImage Analysts (NEUBIAS) in the title, by providing a brief background description about this community and by providing explicit information regarding NEUBIAS members contribution (e.g., mentioning that BIAFLOWS was developed by a NEUBIAS member).

In my opinion, the ideas presented in the third section (“Towards findability and accessibility of image analysis workflows”) could be integrated in the previous sections. This will improve the flow of the text without losing any content. I do not see any conceptual advantage of having a separate section as in the current form.

Perhaps the authors would consider including some of the recent platforms that make machine/deep learning applications accessible to users such as ImJoy and ZeroCostDL4Mic? It is perhaps worth mentioning the uniqueness of machine learning-based components where training is dependent on large amounts of data, but the resulting model can be lightly disseminated (however re-training with new data is adding another layer of complexity related to parameter settings in traditional workflows).

Another related idea that NEUBIAS is heavily involved in is training (for components development, workflow reconstruction, usage and disseminations). Perhaps the authors would be interested in including some ideas in their opinion? I recently discussed related topics¹.

More specific comments, opinions and suggestions:

The Abstract and the beginning of the Introduction are identical. I recommend making them separate entities, where the Abstract summarizes the main ideas, while the Introduction provides more extended background and context for the rest of this piece. In the Abstract I suggest to start with the context, move to presenting NEUBIAS and workflows (do not forget to briefly explain what is a workflow) and finish with the content of this paper (I would remove the sentence on F1000 gateway, why is it relevant in the Abstract or at all?).

Page #3 (Introduction): Since component is the building block of a workflow, I recommend defining component before workflow.

Page #3: "We believe that while this can be a good starting point it may not be sufficient", can you briefly mention why this is not sufficient?

Page #4: "the natural question is then if it is the best solution for the particular task one wants to solve". I think that most users will not find this a "natural" question, rather, their goal is to find a "good-enough" workflow to answer the question they are interested in.

Page #5 ("Toward reproducibility"): perhaps it is worth mentioning and citing the recent paper on integrating ImageJ and CellProfiler² and/or a recent opinion reflecting on some of the aspects discussed in this section³?

Minor formatting issue: the reference comes after the period (".") or comma (",") in several locations (e.g., refs #5-7, #10, #16, #18, #24,#26).

References

1. Driscoll MK, Zaritsky A: Data science in cell imaging. *J Cell Sci.* 2021; **134** (7). [PubMed Abstract](#) | [Publisher Full Text](#)
2. Dobson ETA, Cimini B, Klemm AH, Wählby C, et al.: ImageJ and CellProfiler: Complements in Open-Source Bioimage Analysis. *Curr Protoc.* 2021; **1** (5): e89 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Levet F, Carpenter A, Eliceiri K, Kreshuk A, et al.: Developing open-source software for bioimage analysis: opportunities and challenges. *F1000Research.* 2021; **10**. [Publisher Full Text](#)

Is the topic of the opinion article discussed accurately in the context of the current literature?

Yes

Are all factual statements correct and adequately supported by citations?

Yes

Are arguments sufficiently supported by evidence from the published literature?

Yes

Are the conclusions drawn balanced and justified on the basis of the presented arguments?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Computational cell dynamics, data science in cell imaging, bioimage analysis

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 20 May 2021

<https://doi.org/10.5256/f1000research.55867.r83885>

© 2021 Wolstencroft K. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Katy Wolstencroft 

Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands

This article describes current progress in the area of creating and sharing scientific workflows for bioimage analysis. The article describes community activities that allow the collection and comparison of workflows. It also discusses the necessity for describing workflows and their components in standardised ways, using common vocabularies.

The article serves as a gateway to bioimaging analysis workflow resources. As such, it is a good introduction and provides a clear overview. However, the reality of creating bioimaging analysis workflows, and particularly of reusing them, has some large bottlenecks. It would be beneficial to discuss, even if only briefly, what the bottlenecks are and describe current major challenges. The authors already mention interoperability and provenance as large challenges, but the connection to the data is another large challenge. Are datasets described in common formats? Are there existing ways to work with data generated with different proprietary formats? For this point, I miss a reference to the work of the Open Microscopy Environment community (e.g. ¹).

Are there common representations for bioimaging data? How practical is it to move such large datasets around for the purposes of analysis? For widespread use of bioimaging workflows, advances in data management and data standards must also contribute.

Describing and sharing analysis workflows is a great step forward for the bioimage analysis community, and this paper shows that this field is maturing and gaining good traction.

References

1. Linkert M, Rueden CT, Allan C, Burel JM, et al.: Metadata matters: access to image data in the real world. *J Cell Biol.* 2010; **189** (5): 777-82 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the topic of the opinion article discussed accurately in the context of the current literature?

Partly

Are all factual statements correct and adequately supported by citations?

Yes

Are arguments sufficiently supported by evidence from the published literature?

Yes

Are the conclusions drawn balanced and justified on the basis of the presented arguments?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Scientific workflows, FAIR data, semantic data integration.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research