# Galaxy: A Decade of Realising CWFR Concepts

**Beatriz Serrano-Solano[1]\*[†], Anne Fouilloux[2]\*[†], Ignacio Eguinoa[3,4], Matúš Kalaš[5], Björn Grüning[1] & Frederik Coppens[3,4]**

[1]Bioinformatics Group, University of Freiburg, Baden-Württemberg 79098, Germany

[2]Department of Geosciences, University of Oslo, Oslo 0316, Norway

[3]Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent 9052, Belgium

[4]VIB, Gent, Oost-Vlaanderen 9052, Belgium

[5]Department of Informatics, University of Bergen Ringgold standard institution, University of Bergen, Bergen, Hordaland 5008, Norway

## ABSTRACT

Despite recent encouragement to follow the FAIR principles, the day-to-day research practices have not changed substantially. Due to new developments and the increasing pressure to apply best practices, initiatives to improve the efficiency and reproducibility of scientific workflows are becoming more prevalent. In this article, we discuss the importance of well-annotated tools and the specific requirements to ensure reproducible research with FAIR outputs. We detail how Galaxy, an open-source workflow management system with a web-based interface, has implemented the concepts that are put forward by the Canonical Workflow Framework for Research (CWFR), whilst minimising changes to the practices of scientific communities. Although we showcase concrete applications from two different domains, this approach is generalisable to any domain and particularly useful in interdisciplinary research and science-based applications.

\*    These authors contributed equally to this work.

†    Corresponding author: Beatriz Serrano-Solano (Email: beatrizserrano.galaxy@gmail.com; ORCID: 0000-0002-5862-6132) and Anne Fouilloux (Email: annefou@geo.uio.no; ORCID: 0000-0002-1784-2920).

## 1. INTRODUCTION

Workflows[①] are essential for reproducible research, automation, and re-use of analyses [1]. The gaps between available workflow technology and scientific practices across the diversity of data-intensive research areas are still huge, even though there are clear indications of recurring operations [2]. Despite the recent encouragement to follow the FAIR principles, the daily research processes have not changed substantially. Due to new developments and the increasing pressure to apply best practices, initiatives to improve the efficiency and reproducibility of scientific workflows are becoming more prevalent, focusing on standardisation and integration within the community best practices.

Galaxy is an open-source workflow management system with a web-based interface that allows accessible, reproducible, and transparent computational research [3, 4]. Galaxy encompasses all core components to implement the concepts that are put forward by the Canonical Workflow Framework for Research (CWFR) [5]. The *FAIRification* of workflows relies heavily on the metadata associated with tools that compose these workflows. Well-described tools are key, not only to ensure interoperability, but also to improve their findability and accessibility.

Most tools used by scientists lack associated metadata. To address this issue, each tool in Galaxy has a *wrapper* describing the tool itself along with the input and output parameters, annotations with ontologies, and a Persistent IDentifier (PID), among others. Together, a tool plus its wrapper constitute a "*Galaxy Tool*". The integration of such tools in Galaxy is paramount, since only Galaxy Tools can be combined into workflows to compose "*Galaxy Workflows*" that can be automated and run efficiently. In this article, we discuss the importance of well-annotated tools and the specific requirements to ensure reproducible research with FAIR outputs. We describe how Galaxy and its ecosystem provide essential features that enable researchers to seamlessly publish FAIR workflows reusable by the community.

## 2. GALAXY TOOL DEVELOPMENT PROCESS

The usage of standards and linked data would be more widespread if these were automatically handled in the frameworks where research is done. This is an endeavour of Galaxy and the process to create *Galaxy Tools* has been formalised so that it can be largely automated.

Figure 1 shows an example where an open-source code is packaged using Conda[②] and a container automatically created out of it. The packaging of open-source codes with Conda can be done by anyone, not necessarily by code maintainers or the Galaxy community. The *Galaxy Tool wrapper* is an XML file containing tool information about the requirements, inputs, outputs, and can be annotated with Bio.tools PID and EDAM ontology terms to capture metadata corresponding to its functions, data types, formats, etc.

---

[①] In this article, a workflow is defined as a series of tasks taking inputs and generating outputs. Each task can be another workflow or a basic unit referred to as a "tool".

[②] https://docs.conda.io/en/latest/

The more metadata is added upstream, the better downstream well-annotated *Galaxy Tools*. To improve the findability of Galaxy Tools and make them accessible to the whole Galaxy community, all these are gathered in the Galaxy Tools repository, termed the *Galaxy Tool Shed*[③].
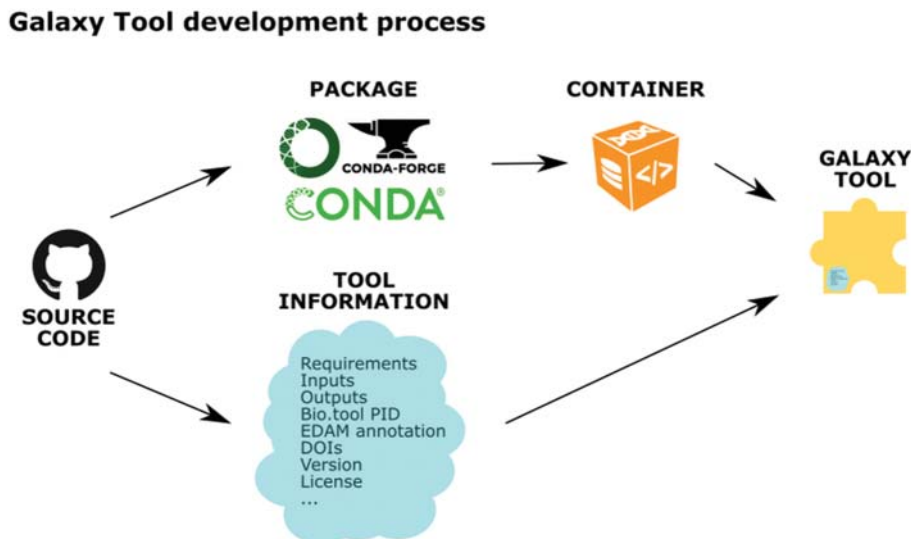
## Galaxy Tool development process



**Figure 1.** Example of development of a *Galaxy Tool* and wrapper describing it.

### 2.1 Packages and Containers

*Galaxy Tools* require runtime environments, underlying libraries, and may depend on tools that are developed and maintained outside the Galaxy ecosystem. The dependencies required for executing all the *Galaxy Tools* involved in a *Galaxy Workflow* can yield an extensive list, including incompatibilities across the *Galaxy Tools* of the workflow. For instance, a *Galaxy Tool* can require the usage of the GDAL[④] library 3.4.0 while another *Galaxy Tool* would only work with GDAL 2.4.2. To prevent incompatibilities between the workflow tasks, reduce the complexity of runtime environments and with that increase the maintainability and reusability of software environments, each task in a workflow is isolated from the other ones. Software packages, especially in conjunction with a package manager, are very common in the open source community, with RPM[⑤] and DPKG[⑥] as prominent examples. Conda is one of the latest generation of package managers and has been selected for *Galaxy Tools* because it is widely used by the scientific community, operating-system agnostic and programming-language independent. However, using package managers does not solve all reproducibility and accessibility issues.

---

[③]  https://toolshed.g2.bx.psu.edu/

[④]  https://gdal.org/

[⑤]  https://rpm.org/

[⑥]  https://www.debian.org/doc/manuals/debian-faq/pkgtools.en.html

Containers offer a higher level of abstraction, by isolating the software environment completely from the host system. This increases reproducibility, for the price of containers being more complicated to design, build and use. Also, not all container technologies are supported on all compute platforms (for instance, HPC typically does not support Docker [9, 10]). As there is usually no one-size-fits-all solution, this means that multiple ways to resolve the dependencies need to be supported. Each *Galaxy Tool* is annotated with tool dependencies using the Conda package manager, which increases tool modularity and portability. For all Conda packages [11] used by *Galaxy Tools*, containers are generated automatically by the BioContainers infrastructure, ensuring that for all *Galaxy Tools* both Docker and Singularity containers exist. This enables Galaxy to choose between Conda, Docker or Singularity for every single task in the workflow. The choice is often driven by the administrators of the computing resources on which a Galaxy Tool is run: for instance, on HPC, Singularity is often required while on cloud computing, a Conda package is often sufficient.

### 2.2 EDAM

*Galaxy Tools* need to be described consistently to allow findability, comparison, and to guide users in their choice. EDAM⑦ is an ontology of data analysis and data management [6], designed for semantic annotation of tools, workflows, and other resources. The EDAM ontology contains over 3,500 concepts with preferred terms, synonyms, definitions, related terms, relations between concepts, and links to other resources. EDAM comprises four sections: topics, operations, data types, and data formats. Although the bulk of concepts in EDAM is specific to life sciences, EDAM also contains numerous higher-level concepts that are not specific to a particular scientific or application domain. In addition, there are mechanisms to extend EDAM to other domains, related or unrelated to biosciences. Examples are EDAM Bioimaging (which contains concepts related to imaging, image analysis, and machine learning, mostly unspecific to a scientific domain [7]) and the work on EDAM concepts for geoscientific, environmental, and humanitarian applications⑧.

EDAM is a shared ontology used across diverse resources that addresses the description of tools across domains. In addition to Galaxy, EDAM is used, for example, also in Debian and the Common Workflow Language (CWL; both described in [8]), FAIRsharing⑨, and especially Bio.tools (described below). Using EDAM as the common ontology enables interchange and integration of semantic annotations across the diverse resources.

### 2.3 Bio.tools

Bio.tools⑩ is an open registry of computational tools for research in life sciences [12]. Bio.tools collates over 20,000 tools encompassing software with command-line, graphical, or programmatic interfaces, web APIs, web applications, and database portals. The records are created and maintained openly by the scientific community [13], supplemented by partial automation and centralised curation.

---

⑦ https://bioportal.bioontology.org/ontologies/EDAM?p=classes
⑧ https://github.com/edamontology/edam-geo
⑨ https://fairsharing.org
⑩ https://bio.tools

A substantial portion of these tools fulfils the requirements of reliable, FAIR CWFR components. Such tools are free, open source software (FOSS), well-documented, easy to set up, and usable in reproducible, interoperable workflows. A tool record in Bio.tools is identified by a PID and contains extensive information about the registered tool, including semantic annotation with the EDAM ontology, and numerous links to e.g., documentation, source code, packages, containers, user support, etc.

The Bio.tools registry focusses on tools applicable to biosciences. However, the open-source software<sup>℗</sup> and the data model running it [14] are generic, usable for setting up registries of computational tools in other domains. Whenever available, Galaxy Tools are annotated with Bio.tools PIDs.

## 3. FAIR GALAXY WORKFLOWS

### 3.1 Galaxy Workflow Assembly

All *Galaxy Workflows* created by assembling *Galaxy Tools* (as shown on Figure 2) are not FAIR by default. To become *FAIR Galaxy Workflows* more annotations need to be added such as license, authors and institutes, following the best practices of the *Intergalactic Workflow Commission (IWC)*®. These Galaxy Workflows are then reviewed, tested and then packaged using the RO-Crate packaging format for publication as FAIR Digital Objects (FDOs) on the WorkflowHub®, a FAIR and open registry for workflows (Figure 2).
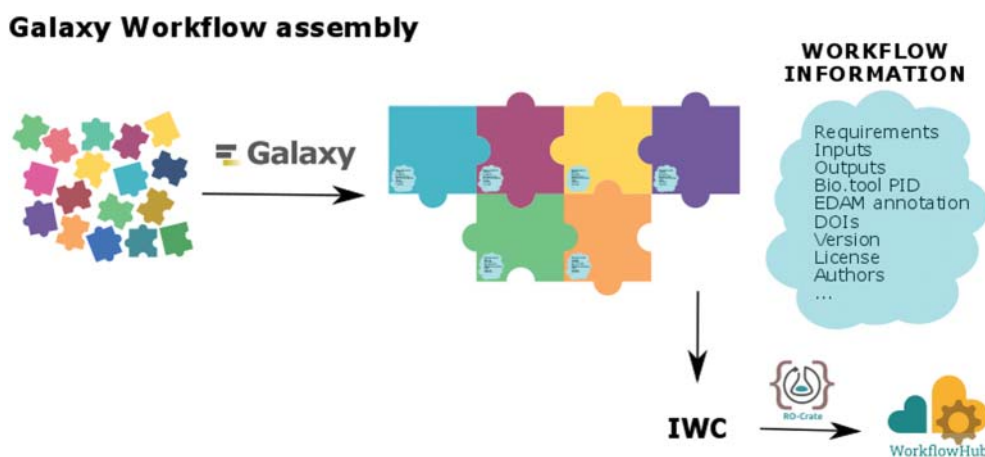


**Figure 2.** Galaxy allows the combination of interoperable *Galaxy Tools* into *Galaxy Workflows* that inherit the metadata from the *Galaxy Tools* composing it. The resulting *Galaxy Workflows* can be deposited in the repository of the IWC, and exported to the WorkflowHub as RO-Crate objects.

---

<sup>℗</sup>  https://github.com/bio-tools
<sup>®</sup>  https://github.com/galaxyproject/iwc
<sup>®</sup>  https://workflowhub.eu

### 3.2 FAIR Digital Objects (FDOs) through the WorkflowHub

The WorkflowHub is a domain-independent registry for computational workflows and it is designed to be agnostic to the workflow management system used to describe the workflow. Workflows are exported or imported to the WorkflowHub using the Workflow RO-Crate® format.

RO-Crate is a general-purpose lightweight packaging format for research data, which defines a workflow-specific profile® e.g., a minimum set of conventions, types and properties to be present. The Workflow RO-Crate profile requires at least one computational workflow® and it is also recommended accompanying the native workflow definition with an abstract Common Workflow Language [15] (CWL) description and a diagram to visualise the workflow. This facilitates finding and comparing workflows across platforms, thereby extending the interoperability. Within the RO-Crate, workflow entities metadata is annotated using Bioschemas® markup to further increase the findability. RO-Crate aligns with the principles of FAIR Digital Objects [16] and is being adopted by services across scientific domains.

When a *Galaxy Workflow* is submitted to the WorkflowHub, an abstract representation using the Common Workflow Language Abstract Operation® is generated®. The resulting abstract *Galaxy Workflow* contains a high-level description of all the *Galaxy Tools* used in the *Galaxy Workflow*, e.g., inputs and outputs (formats, types) as well as the type of operations, but without any reference to a concrete implementation of the *Galaxy Tools*.

## 4. CONNECTING COMPONENTS TO UNDERPIN CWFR THROUGH GALAXY

### 4.1 Galaxy Workflow Execution

*Galaxy Workflows* can only be executed on *Galaxy instances,* i.e. on deployments of the Galaxy software with a set of available compute and storage resources. Depending on the resources needed, a Galaxy instance can be deployed in various environments, from a personal computer to a cloud setting. In Galaxy, the result of the execution of a *Galaxy Workflow* is stored as a *Galaxy History*. A *Galaxy History* keeps track of the data provenance combined with other metadata, such as the *Galaxy Tool* version and any parameter used to run it. Depending on the user's needs, the *Galaxy History* can be shared with particular users, with a group, or publicly with all the users of the given *Galaxy instance*. Conceptually, a *Galaxy History* contains all information required to build a FAIR Digital Object that scientists can re-run to reproduce the analysis (same *Galaxy Workflow*, same inputs and parameters), or reuse the *Galaxy Workflow* for a different purpose, potentially on another *Galaxy instance*. This feature has proven to be very useful also for training purposes [17] where instructors can, for example, follow the progress of trainees.

---

® https://www.researchobject.org/ro-crate
® https://www.researchobject.org/ro-crate/profiles.html
® https://www.researchobject.org/ro-crate/1.1/workflows.html
® https://bioschemas.org/
® https://www.commonwl.org/v1.2/Workflow.html#Operation
® https://github.com/workflowhub-eu/galaxy2cwl

### 4.2 Creation of Galaxy Workflows

The research process requires flexibility and researchers need to be able to easily create their own *FAIR Galaxy Workflows* by assembling and executing *Galaxy Tools* one after the other, checking intermediate results, choosing the next *Galaxy Tool* depending on the result, etc. Similar to *Galaxy Tools*, datasets in Galaxy (workflow inputs, intermediate, and final results) are also annotated with, at least, the name of the dataset, the data type, permissions, and user-defined tags. Dealing with data types is not straightforward, and the ongoing creation of new data types in almost all research disciplines is challenging. For many standard types, there are known software libraries that can detect the file type. Galaxy can infer it using a so-called "sniffer", simplifying the assignment of data types by researchers and minimising errors.

The *Galaxy History* contains every *Galaxy Tool* that has been run in a given analysis. For complex analyses, a multistep set of *Galaxy Tools* will be executed. This interlinked set of *Galaxy Tools* constitutes a *Galaxy Workflow* that can be extracted directly from the *Galaxy History*.

*Galaxy Workflows* can also be created via the *Galaxy Workflow Editor*, a graphical user interface in which users can select tools and connect them with each other. Users are guided, e.g., connections are constrained by data types, which significantly limits potential errors. *Galaxy Workflows* can be imported to be executed and inherits metadata from the *Galaxy Tools* and dataset annotations to track the provenance. Additional metadata can be added, such as the name of the *Galaxy Workflow*, version, license, author, tags, and labels. Galaxy provides a validation wizard to check if a *Galaxy Workflow* follows the best practices, and guides users through the process by highlighting missing annotations.

*Galaxy Workflows* are meant to execute *Galaxy Tools* in batch mode without further human intervention, although for some applications, it may be useful to explore alternative pathways (e.g., using Interactive Tools in Galaxy like Jupyter Notebooks[ⓩ] or visualisations). Metadata and provenance of Interactive Tools in Galaxy also have to be captured to be FAIR.

*Galaxy Workflows* can also be composed of other *Galaxy Workflows*, which can be seen as sub-workflows. This flexibility and the different degrees of granularity provide a framework tailored to the needs of different communities. For instance, many bioinformatics tools have a very fine granularity, focused on one specific operation, while some climate tools can be coarse-grained: a tool can be as complex as a climate model that is composed of many components.

The *Galaxy Workflows* can be shared with an arbitrary set of users or publicly. A Galaxy user can make one of its *Galaxy Workflow* accessible via a weblink: anyone with this weblink can then view, and import or download it. However, to make a *Galaxy Workflow* public, the user needs to explicitly make it accessible via a link and publish it to the Galaxy's 'Published workflows' section of a Galaxy instance. Anyone will be able to search, find, view, import, and download it. Although the corresponding published workflow is only available in the given Galaxy instance, it can be imported and executed in a different Galaxy instance.

---

ⓩ  https://jupyter.org

Based on the exported file, all necessary tools (and the exact versions) can be automatically installed by an administrator[⊛].

Making workflows available through a specific *Galaxy instance* has limitations, e.g., for findability. To overcome this, *Galaxy Workflows* can also be registered in the WorkflowHub (see section 3.2): workflows listed in the WorkflowHub are easier to find, since the collection is independent of a *Galaxy instance*. In addition to the *Galaxy Workflow* file, additional metadata and information can be added to enrich the information and increase the FAIRness.

## 5. EXAMPLE APPLICATIONS

The proposed approach has been put into practice by different scientific communities, as can be seen in the different training materials deposited in the Galaxy Training Network repository[⊛]. In this section, we highlight two demonstrators from European projects (EOSC-Nordic and EOSC-Life).

As part of the European project EOSC-Life, the demonstrator "Image Repository and Scalable Mining"[⊛] had, as its main goal, the re-mining of large-scale FAIR image resources to extract information that was not within the scope of the original study. This exemplary workflow (Figure 3A) consists of the first automated part using modules of the popular image analysis suite CellProfiler[⊛], available in Galaxy, to perform cell segmentation and feature extraction. Once the data is reduced, the downstream analysis can be customised using RStudio interactively. This way, the analysis can benefit from the HPC infrastructure, keeping the reproducibility and transparency of the results.

The climate science demonstrator[⊛] of the European project EOSC-Nordic (Figure 3B) utilises Galaxy to offer a flexible computational environment to collaborate, understand, co-develop, implement, and test new scientific developments to better forecast climate change and develop sound responses. This effort is complemented by the European project RELIANCE[⊛], which focuses on the management of the research lifecycle among Earth-science communities and Copernicus[⊘] users. A typical climate modelling workflow usually starts with the retrieval of relevant data from various providers, what can be re-used to design new Earth System Model[⊛] simulations. During this step, scientists from different disciplines often need to work together in a co-design effort: this is a very interactive task where changes need to be immediately validated

---

[⊛]  https://ephemeris.readthedocs.org
[⊘]  https://training.galaxyproject.org
[⊛]  https://www.eosc-life.eu/d6
[⊛]  https://cellprofiler.org
[⊛]  https://www.eosc-nordic.eu/demonstrating-eosc-nordic
[⊛]  https://www.reliance-project.eu
[⊘]  Copernicus is the European Union's Earth observation programme, "looking at our planet and its environment for the benefit of Europe's citizens" (https://www.copernicus.eu/en).
[⊛]  https://www.climateurope.eu/earth-system-modeling-a-definition,   https://soccom.princeton.edu/content/what-earth-system-model-esm

and results visualised with RStudio, or Python in Jupyter Notebooks. Once all the involved scientists agree on the scenario and all the input datasets are ready, the Earth System Model can be run in operational mode: this long simulation is run as an automated workflow. Finally, analysing the results of the simulation is usually done in interactive environments such as Panoply to visualise data, and/or the Pangeo Jupyter ecosystem, both usable directly from within Galaxy.
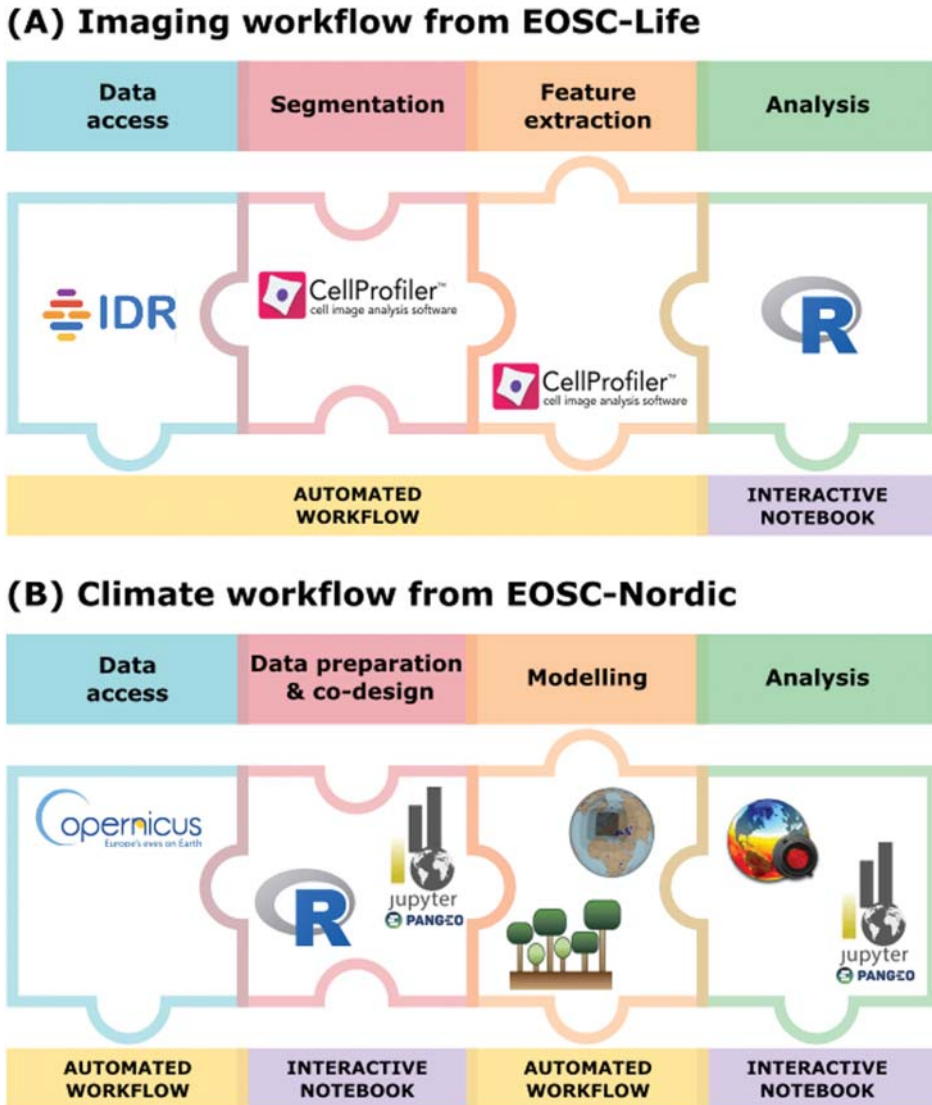
## (A) Imaging workflow from EOSC-Life

| Data access | Segmentation | Feature extraction | Analysis |
|---|---|---|---|

IDR — CellProfiler cell image analysis software — CellProfiler cell image analysis software — R

| | AUTOMATED WORKFLOW | | INTERACTIVE NOTEBOOK |

## (B) Climate workflow from EOSC-Nordic

| Data access | Data preparation & co-design | Modelling | Analysis |
|---|---|---|---|

Copernicus — R, jupyter PANGEO — jupyter PANGEO

| AUTOMATED WORKFLOW | INTERACTIVE NOTEBOOK | AUTOMATED WORKFLOW | INTERACTIVE NOTEBOOK |

**Figure 3.** **Exemplary workflows.** Each puzzle piece corresponds to a possible tool (or a set of tools) that can be used in a particular step of the workflow. **(A) Image analysis workflow from EOSC-Life**, describing all the different stages from data access, via the automated image analysis workflow, to the final analysis of the extracted features. **(B) Climate modelling workflow from EOSC-Nordic**, for improving climate predictions.

## 6. CONCLUSION

To realise the Canonical Workflow Frameworks for Research, we need to facilitate the usage of standards in the platforms researchers are already familiar with. This will allow the production of FAIR data without major changes in the practices of scientific communities, which will yield faster results towards open science.

The Galaxy Project has been a protagonist of open science for over a decade. Galaxy provides a workflow analysis platform, part of a broader ecosystem, that is production-ready and addresses the needs of diverse user patterns across scientific disciplines. The adoption of the EDAM ontology allows describing tools and data in a controlled way, while the integration with Bio.tools provides unique, persistent identifiers that are platform-agnostic. The wide adoption of RO-Crate to support FAIR Digital Objects and its full integration in Galaxy will be a major milestone towards publishing FAIR data for all aspects of the computational scientific workflow, through services such as the WorkflowHub.

Galaxy captures relevant metadata to reproduce an analysis in its environment. Being able to export FDOs, from within Galaxy, will allow exposing details about the data without the need to create them externally, as is the case in the current integration with the WorkflowHub. Galaxy started this journey a while ago with the adoption of standards such as BagIt[⑧] and lately BioCompute Objects [18] as well as RO-Crate.

Galaxy keeps a detailed record of each workflow invocation, but mapping it into an interoperable format is a big challenge. The end goal is being able to fit all relevant execution details from the Galaxy data model into an RO-Crate package, following the W3C PROV data model[⑨]. The features described in this article are required but not sufficient to achieve this. For example, PIDs for input data are necessary, while often these are not yet available at the time of analysis. However, through integration with data management platforms, such information could be tracked.

Especially for interdisciplinary research, availability of a common technical framework, such as Galaxy and its ecosystem, is crucial for enabling analyses combining different community practices. The framework needs to build on current practices and have sufficient support within the communities, in order to be sustainable. Among all the efforts to reach out to new scientific communities, training is undoubtedly the most critical one [19], together with the integration of the community tools and data sources. A production-ready, flexible analysis environment that supports and stimulates FAIR data, combined with adequate training, will allow closing the gap between technical possibilities and community practices, and realising the goals of transparent and accessible open science.

---

[⑧]  https://tools.ietf.org/id/draft-kunze-bagit-16.html

[⑨]  https://www.w3.org/TR/prov-overview/

## ACKNOWLEDGEMENTS

## AUTHOR CONTRIBUTIONS

All authors contributed to the writing of the manuscript. In addition, B. Grüning (bjoern.gruening@gmail.com) and F. Coppens (frederik.coppens@psb.vib-ugent.be) conceived the idea of writing the manuscript, and B. Serrano-Solano (beatrizserrano.galaxy@gmail.com) led the writing process. B. Serrano-Solano and A. Fouilloux (annefou@geo.uio.no) led the work on the two concrete research examples.

## REFERENCES

[1]    Atkinson, M., et al.: Scientific workflows: Past, present and future. Future Generation Computer Systems 75, 216–227 (2017)

[2]    Jeffery, K., et al.: Not ready for convergence in data infrastructures. Data Intelligence 3(1), 116–135 (2021)

[3]    Goecks, J., et al.: Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biology 11, Article No. R86 (2010)

[4]    Afgan, E., et al.: The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Research 46(W1), W537–W544 (2018)

[5]    Hardisty, A., Wittenburg, P. (eds.): Canonical Workflow Framework for Research (CWFR)—position paper—version 2, December 2020. Working paper. Available at: https://osf.io/9e3vc/. Accessed 29 November 2021

[6]    Black, M., et al.: EDAM: The bioscientific data analysis ontology (update 2021) [version 1; not peer reviewed]. F1000Research 2022, 11(ISCB Comm J): poster (2022)

[7]    Kalaš, M., et al.: EDAM-bioimaging: The ontology of bioimage informatics operations, topics, data, and formats (update 2020). F1000Research 9(162), poster (2020).

[8]    Möller, S., et al.: Robust cross-platform workflows: How technical and scientific communities collaborate to develop, test and share best practices for data analysis. Data Science and Engineering 2, 232–244 (2017)

[9]    Younge, A.J., et al.: A tale of two systems: Using containers to deploy HPC applications on supercomputers and clouds. In: IEEE International Conference on Cloud Computing Technology and Science (CloudCom), pp. 74–81 (2017)

[10]   Kurtzer, G.M., Sochat, V., Bauer, M.W.: Singularity: Scientific containers for mobility of compute. PLOS ONE 12(5), e0177459 (2017)

[11]   da Veiga Leprevost, F., et al.: BioContainers: An open-source and community-driven framework for software standardization. Bioinformatics 33(16), 2580–2582 (2017)

[12]   Ison, J., et al.: The bio.tools registry of software tools and data resources for the life sciences. Genome Biology 20, Article No. 164 (2019)

[13]  Ison, J., et al.: Community curation of bioinformatics software and data resources. Briefings in Bioinformatics 21(5), 1697–1705 (2020)

[14]  Ison, J., et al.: biotoolsSchema: A formalized schema for bioinformatics software description. GigaScience 10(1), giaa157 (2021)

[15]  Crusoe, M.R., et al.: The CWL community: Methods included: Standardizing computational reuse and portability with the common workflow language.  arXiv preprint arXiv: 2105.07028 (2021)

[16]  Soiland-Reyes, S., et al.: Packaging research artefacts with RO-Crate. Available at: https://doi.org/10.3233/DS-210053. Accessed 29 November 2021

[17]  Serrano-Solano, B., et al.: Fostering accessible online education using Galaxy as an e-learning platform. PLOS Computational Biology 17(5), e1008923 (2021)

[18]  Simonyan, V., Goecks, J., Mazumder, R.: Biocompute objects—A step towards evaluation and validation of biomedical scientific computations. PDA Journal of Pharmaceutical Science and Technology 71(2), 136–146 (2017)

[19]  Batut, B., et al.: Community-driven data analysis training for biology. Cell Systems 6(6), 752–758 (2018)

## AUTHOR BIOGRAPHY

**Beatriz Serrano-Solano** coordinates the European ELIXIR-Galaxy Community and has previously worked on a large-scale image analysis project using Galaxy within EOSC-Life. She is involved in outreach, training, and community engagement activities for the global Galaxy community.
ORCID: 0000-0002-5862-6132

**Anne Fouilloux** is a research software engineer at the University of Oslo, Norway, working on the development of climate tools in Galaxy for fostering open science practices.
ORCID: 0000-0002-1784-2920
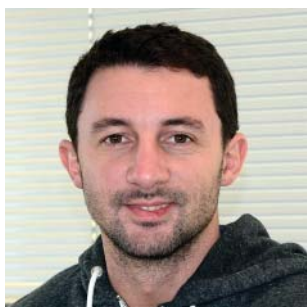
**Ignacio Eguinoa** is a research software engineer at VIB, home of the ELIXIR Belgium Node, working on implementing FAIR data infrastructures. He is also in charge of *usegalaxy.be*, the Belgian Galaxy instance.
ORCID: 0000-0002-6190-122X

**Matúš Kalaš** is a senior researcher at the University of Bergen, Norway, working on the EDAM ontology, and contributing to the open science movement with descriptions and data models of computational tools and other resources, and with education.
ORCID: 0000-0002-1509-4981

**Björn Grüning** is Galaxy, Bioconda and BioContainers core developer. He works at the University of Freiburg, Germany, and is responsible for the European Galaxy server and its ecosystem. He is an ambassador of open science as a co-lead of the ELIXIR Tools Platform, EOSC-Life, and the ELIXIR Galaxy Community.
ORCID: 0000-0002-0173-3009

**Frederik Coppens** is the Head of Node of ELIXIR Belgium and IT manager at the VIB-UGent Center for Plant Systems Biology. He co-leads the ELIXIR Galaxy Community and WorkflowHub, co-authored the EOSC Interoperability Framework and is involved in the EOSC Association and EOSC-Life project.
ORCID: 0000-0001-6565-5145