

Interactive Tools (IT) in Galaxy

Combining synchronous and asynchronous workflows

Tunc Kayikcioglu¹[\[https://orcid.org/0000-0003-2205-8062\]](https://orcid.org/0000-0003-2205-8062), Sanjay Kumar Srikakulam¹[\[https://orcid.org/0000-0002-1752-5060\]](https://orcid.org/0000-0002-1752-5060), Wolfgang Maier¹[\[https://orcid.org/0000-0002-9464-6640\]](https://orcid.org/0000-0002-9464-6640), and Björn Grüning¹[\[https://orcid.org/0000-0002-3079-6586\]](https://orcid.org/0000-0002-3079-6586)

¹ Freiburg Galaxy Team, Bioinformatics Group, Department of Computer Science, University of Freiburg, Germany

Abstract. Galaxy [1] is a GUI-based scientific data management and analysis platform that aims to expand the target audience of scientific computing to scientists without advanced computer literacy or access to computing infrastructure. During the operation of a typical asynchronous workflow, the interaction of the user triggers a job submission to a hosted computing system, the results of which are then reported back to the user upon completion of the task. However, such an asynchronous interaction scheme is not compatible with scientific software that require user interaction during the execution itself (synchronous), e.g. many data visualisation software. Interactive Tools on Galaxy (IT; [2]) leverage tool and job components of Galaxy by taking advantage of the latest containerization solutions to provide access to interactive executable software. With ITs, Galaxy is enabling the more and more common use-case of asynchronous workflows while keeping the possibility to bridge at any time to synchronous workflows and reusing all the other Galaxy data management features.

Dynamic visual analytics applications, such as those using R Shiny, Dash or Ruby on Rails can be incorporated into Galaxy through an extension of the standard Galaxy tool framework. A developer needs to provide a container image (e.g., Docker) of the underlying application and a Galaxy tool specification (a.k.a. 'wrapper'). The existing job management system of Galaxy handles scheduling the execution of the container using its existing job management system. Interactive access through a built-in proxy is provided by presenting the user an auto-generated URL via the usual Galaxy GUI, which is forwarded to an internal port bound by the application of interest. The application will start in a user-specified configuration with access to input datasets. Users can then use the application interactively and commit the results of their work in the form of regular Galaxy output datasets. ITs can be seamlessly integrated into Galaxy: If the IT has been executed as part of a workflow, steps downstream of the IT will be scheduled once the expected outputs from the IT (whether generated through automated execution or through a human in the loop) become available. Irrespective of whether it is part of a Galaxy workflow, the saved efforts of the user will be recorded within their analysis history, and the exported 'notebooks' (the Jupyter 'scripts') can be shared with others as needed just as any other dataset. Such ITs provide a reproducible and resumable computing system which can leverage not only the computation infrastructure, but also the storage system.

As our scientific knowledge deepens, cross-disciplinary collaboration is becoming increasingly important for discovery, because it is rarely possible using only one set of skills. Galaxy currently supports sharing of research assets, such as workflows and histories, with other users and even with non-Galaxy users. However, these sharing capabilities are more limited to sharing a snapshot of the work rather than enabling virtual real-time collaboration. In contrast, the IT framework enables real-time visual analytics collaboration among authorized

users. As such, we expect ITs to have a similar enabling impact on collaborative teams as the multi-access office suite tools that make collaborative editing of research outputs possible. It is important to note that these simultaneous collaborative capabilities are being provided solely by our framework and do not require any modifications to the individual software for support at all.

As a case study, consider a user developing a simple machine learning model for the first time, a task involving a series of commands executed on some test data sets. Galaxy enables those exploratory research scenarios for example with the integration of Jupyter Notebooks [3][4], a popular open source web application that can be used to create and share documents containing source code and its products. A local deployment or usage of such an online resource would potentially present technical and security concerns and prohibit common usage scenarios such as institutional policies or hardware with restrictive configurations. Transfer of training data would also require the user to transfer large amounts of data over the network. The user would also likely need to purchase and/or configure a suitable graphics processor to be able to fully test her code and the model. ITs alleviate all of these issues as the users would immediately gain access to a pre-configured system with access to their other data in the Galaxy storage system.

Up to this date, our public instance, <https://live.usegalaxy.eu>, already had 20,000 IT jobs executed. We think that ITs can provide an opportunity to integrate much more interactive scientific use-cases and visualizations and will increasingly gain popularity in the community. In this talk we will walk through the integration and practical usage of the ITs with the aim of gradually increasing interest in deployment of new ITs and supporting more interesting and diverse use-cases.

Keywords: Interactive Tools, Galaxy, workflows, notebooks, Jupyter, R Studio

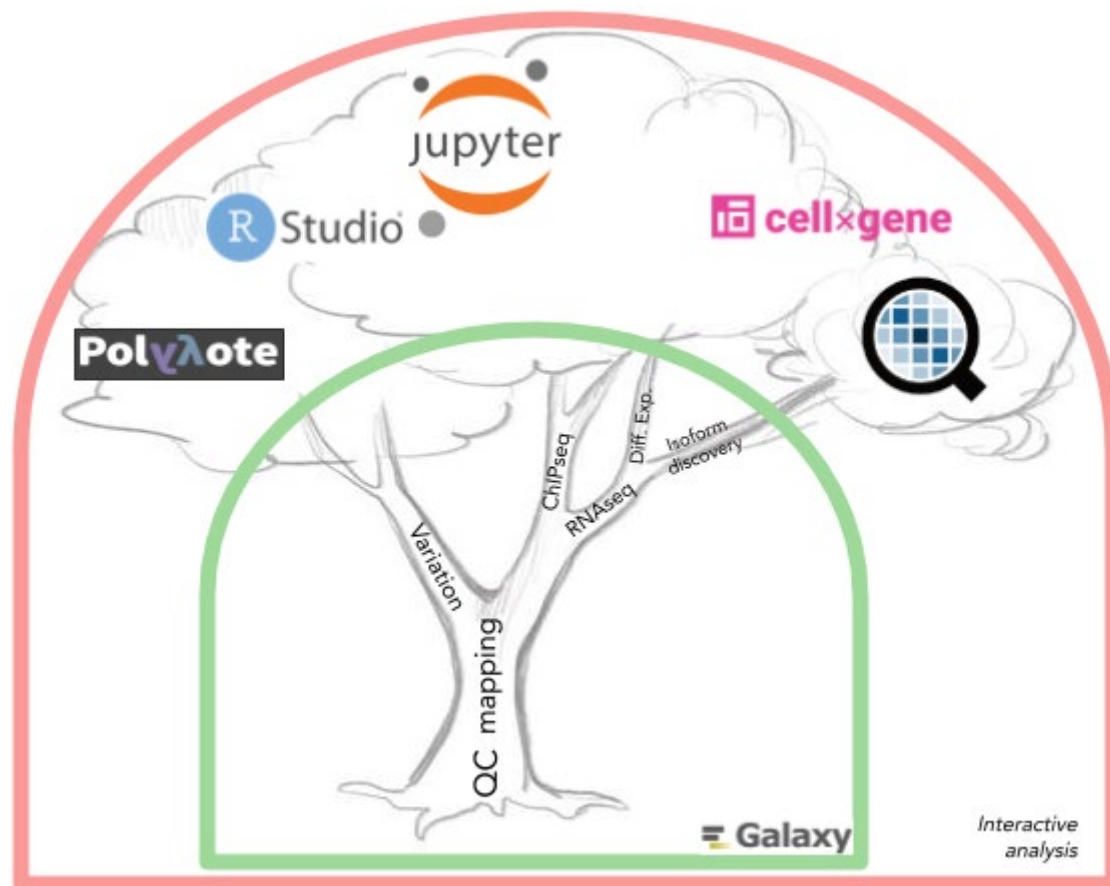


Figure 1. Galaxy and Interactive Tools (IEs). Galaxy’s interface works well for asynchronous processing of thousands of samples with existing tools (green outline). For exploratory data analyses, interactive environments are used. The aim is to enable integration between Galaxy and any IE such as programming frameworks such as Jupyter, RStudio or interactive visualization tools such as HiGlass. The user will be able to move the data between Galaxy and IEs transparently..

Competing interests

The authors declare that they have no competing interests.

References

1. The Galaxy Community - Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update; *Nucleic Acids Research*, Volume 50, Issue W1, 5 July 2022, Pages W345–W351, <https://doi.org/10.1093/nar/gkac247>
2. Gruning, B.A. et al. (2017) Jupyter and Galaxy: Easing entry barriers into complex data analyses for biomedical researchers. *PLoS Comput Biol*, 13, e1005425. <https://doi.org/10.1371/journal.pcbi.1005425>
3. <https://training.galaxyproject.org/training-material/topics/admin/tutorials/interactive-tools/tutorial.html> (accessed Apr. 24, 2023).
4. <https://realpython.com/jupyter-notebook-introduction/> (accessed Apr. 24, 2023).
5. <https://cfp.jupytercon.com/2020/schedule/presentation/184/climate-jupyterlab-as-an-interactive-tool-in-galaxy/> (accessed Apr. 24, 2023).