
Title	The perfect fit for reproducible interactive research: Galaxy, Docker, IPython
Author	Björn Grüning ¹ , Eric Rasche ² , John Chilton ³ and Dannon Baker ⁴
Affiliation	¹ Bioinformatics Group, Department of Computer Science, University of Freiburg, Germany ² Center for Phage Technology, Texas A&M University, USA ³ Department of Biochemistry and Molecular Biology, PSU, USA ⁴ Department of Biology, Johns Hopkins University, USA
Contact	bjoern.gruening@gmail.com
URL	https://github.com/bgruening/docker-galaxy-stable
License	MIT

For years Galaxy has made advanced bioinformatics software accessible to biologists directly by providing an intuitive webinterface to these applications while fostering reproducibility through the automatic creation of re-runnable protocols of each analysis. With the Tool Shed, Galaxy gained a flexible deployment platform enabling identical software installations across Galaxies.

A major hurdle in using Galaxy today is simply finding an instance with the correct set of tools and with enough computational power and storage necessary for a particular analysis. One interesting solution to this challenge is to move the software (Galaxy and the tools) to the data instead of shipping data to a remote server running the correct software; it's also worth noting that many researchers and institutions simply cannot ship their data elsewhere.

Galaxy is using Docker to solve this problem in a way that is even more reproducible by delivering the entire software stack in a container. Each new release of Galaxy is now available as a production-ready Docker container. Additionally, this Docker image can be extended to build personalized Galaxy flavours, with site-specific sets of tools. For example, a Galaxy Docker flavour containing all necessary tools for RNA-seq analysis, or a genome annotation flavour with the NCBI BLAST suite. These flavours are simple to create and can be easily deployed on Linux, OS-X and Windows.

Likewise, Galaxy now allows running tools securely in Docker containers. The process isolation provided by running Galaxy jobs in Docker containers provides a much higher degree of security than running them as native processes, at least in part due to inability to access other users' data. An exciting new development in this Galaxy/Docker ecosystem is the Galaxy-IPython project. IPython is a popular platform providing a web-based interactive computing and visualization environment. Galaxy-IPython allows Galaxy users to run IPython inside Galaxy and access it via their web browser. Additionally, it extends the default IPython environment by providing easy, secure access to Galaxy, it's API, and the user's data. As Galaxy-IPython is deployed on the Galaxy server, it removes the overhead of big-data downloads and uploads during analysis. All of these features work to enable rapid, iterative, and interactive bioinformatics analysis and software prototyping directly in Galaxy, next to your big data.

Galaxy is a popular tool for teaching bioinformatics applications to biologists - Galaxy IPython is a huge step towards enabling it to be a teaching tool for bioinformatics programming as well.