Planemo – A Scientific Workflow SDK

John Chilton[1], Aysam Guerler[2], and The Galaxy Team
[1]Department of Biochemistry and Molecular Biology, Penn State University, University Park, PA 16801, US
[2]Department of Biology, Johns Hopkins University, Baltimore, MD 21218, US
jmchilton@gmail.com

Galaxy is a data analysis platform capable of integrating diverse command-line utilities into a consistent and intuitive web-based interface and API. A salient feature of Galaxy is the ability it provides to compose analysis steps together into workflows. A novel approach to building, refining, and running scientific workflows leveraging Galaxy through the command-line toolkit Planemo will be presented.

Traditionally there have been two methods to build Galaxy workflows - a graphical workflow editor and a workflow extraction interface. Both of these methods are great end-user facing tools that allows users with development experience to build workflows. However, sophisticated bioinformaticians and Galaxy plugin developers (e.g. tool developers) may prefer driving workflow development through their existing tool chains and methodologies such as programming text editors, command-line invocation, test-driven development, and revision control. The approach presented leverages YAML-based workflow descriptions as plain files allowing exactly this.

The approach will be used as a lens to highlight these workflows formats (Format 2 Galaxy workflows and Common Workflow Language (CWL) workflows) as well as important updates from the myriad of recent Galaxy workflow enhancements that have made them dramatically more usable, powerful, and performant.

Format 2 Galaxy workflows map directly to existing Galaxy tool and workflow concepts and are described in a very concise and readable YAML format. These will work without modifications to Galaxy today. CWL specifications for tools and workflows are developed in an open fashion by many organizations with the aim of creating truly portable descriptions. The execution of CWL workflows in Galaxy is being actively worked on and progress will be discussed.

These innovations are enabled in part by many core Galaxy enhancements. The most important of these enhancements will be highlighted. Including:
- The user interface for workflows has been overhauled and improved.
- Workflows now allow nesting, non-data inputs, implicit connections, and many new operations over collections - Galaxy workflows are now vastly more expressive.
- Recent performance enhancements allow Galaxy workflows to scale to thousands of datasets.