Galaxy: Sharing and Automation

Reproducibility and Collaboration within Galaxy

Ann Black-Ziegelbein
annblack@eng.uiowa.edu
Senior Application Development
Molecular Otolaryngology and Renal Research Laboratories
Galaxy for Collaborative Analysis

How can I easily share my data and processes with collaborators? How can my lab work as a collective?
Collaboration concepts within Galaxy

- Galaxy can share via two main ways:
  - Sharing/Publishing the History (and all contained datasets)
  - Moving datasets into a common shared data library
- Galaxy has two main concepts for collaboration:
  - Sharing
    - Via URL Link which can be shared
    - Via adding one or more user’s galaxy user ids (uiowa email address)
  - Publishing for world wide access
- **TIP**: Once a history has been imported, it is a snapshot of the history at that moment. Any future changes to the history will not be reflected unless a new import occurs.
Sharing Histories in Galaxy

Making History Accessible via Link and Publishing It

This history is currently restricted so that only you and the users listed below can access it. You can:

Make History Accessible via Link
Generates a web link that you can share with other people so that they can view and import the history.

Make History Accessible and Publish
Makes the history accessible via link (see above) and publishes the history to Galaxy’s Published Histories section, where it is publicly listed and searchable.

Sharing History with Specific Users

You have not shared this history with any users.

Share with a user

Back to Histories List

History Lists
Saved Histories
Histories Shared with Me
Current History
Create New
Clone
Copy Datasets
Share or Publish
Extract Workflow
Dataset Security
Show Deleted Datasets
Show Hidden Datasets
Purge Deleted Datasets
Show Structure
Export to File
Delete
Delete Permanently
Other Actions
Import from File
Shared Data Libraries

• When would I need to use a shared data library?
  ▪ Common set of datasets used by a lab in analysis
  ▪ Mechanism for batch download of datasets out of Galaxy

• What type of security does a shared data library have?
  ▪ The following privileges can be administered on a per user basis (or left open to everyone)
    • Viewing/accessing data library
    • Adding items to data library (datasets or folders, etc)
    • Modifying data library items (e.g., change name or description of datasets)
    • Managing library permissions
Why would I want to share a History?

- Allows collaborators to see exact ordered steps taken on input data (all configured parameters, notes, annotations, etc)
  - Also great for problem determination!
- Share datasets with a colleague without having to move or copy files

- **Tip:** Cloned or copied datasets and histories do not make copies of the underlying data content (no data bloat). Only metadata which is stored in a database is copied.
Galaxy for Repetitive Analysis

How can I automate a common set of tasks that I need to perform on my data input files?
Repetition of Galaxy Analysis Steps

- Good News! Analysis steps on data do not have to be performed one by one
  - Enter Workflows!
- Galaxy Workflows
  - An ordered series of analysis/computational steps to perform on one or more input files
  - You provide the inputs, Galaxy will automatically execute all the steps on the input files until a final output is reached.
Galaxy Workflows

- Galaxy Workflows can be created:
  - Manually through the Workflow Editor
    - Advanced scenario for those very familiar with Galaxy
  - Exported/Created from a History
    - Galaxy will create the ordered series of steps from the actions outlined in the user history
    - Deleted datasets/steps will NOT be incorporated
    - Initial workflow can be edited/reviewed in the workflow editor.
Getting Fancy with Workflows

• Can control whether intermediate datasets are shown or hidden
• Can have the output of a workflow run be directed to a brand new History
• Can add in post processing actions on each workflow step to do things like:
  ▪ Email you on completion
  ▪ Rename the dataset to your liking
  ▪ Change the datatype
  ▪ Augment column definitions
Sharing Workflows

- Like Histories, Workflows can be shared or published
- Workflows can also be exported/imported
  - To share across multiple Galaxy deployments
  - To archive/version control a workflow
- Penn State Galaxy (usegalaxy.org)
  - Plethora of published workflows to snoop
  - Can export workflows from Galaxy Main and import into Iowa’s Galaxy
  - **Note:** May not work if it requires tools Iowa does not yet have deployed
Running a Workflow

- Workflows can be accessed from:
  - Workflow work area view
  - Tool navigation bar
    - Workflow must be explicitly added to the tool navigation to appear

- Inputs to the workflow must be in current history
- Can view all steps and make some input tweaks before running
- Output can be redirected to a new history

Iowa Initiative in Human Genetics :: http://www.medicine.uiowa.edu/humangenetics
University of Iowa Bioinformatics :: http://bioinform-div.healthcare.uiowa.edu/
Demo!

Demonstration on creating a workflow from a history, making minor edits and running!