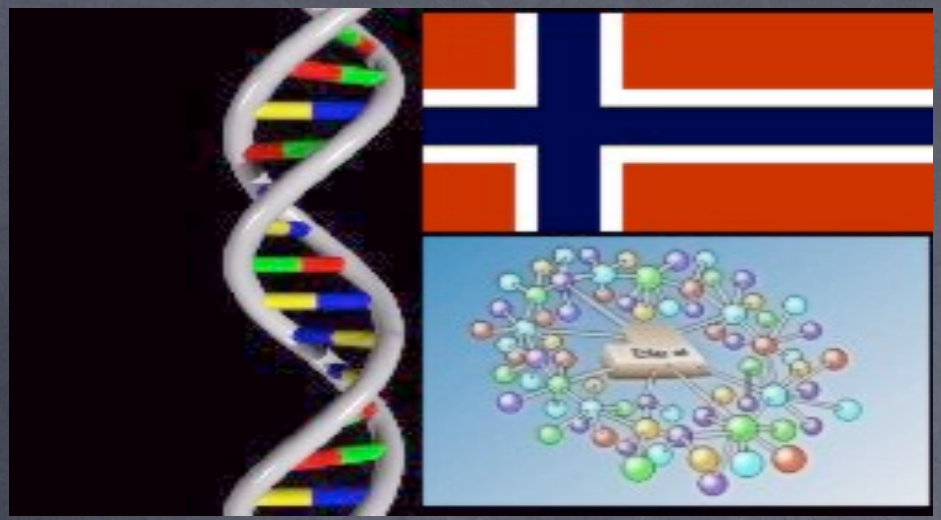
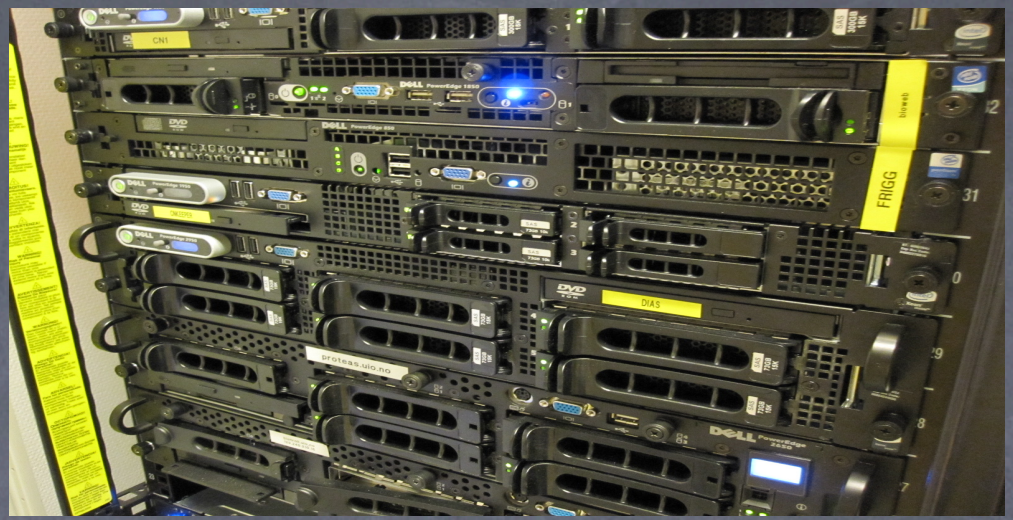




The Galaxy workflow

George Magklaras PhD RHCE

Biotechnology Center of Oslo &

The Norwegian Center of Molecular Medicine

University of Oslo, Norway

http://www.biotek.uio.no

http://www.ncmm.uio.no

http://www.no.embnet.org

Talleres Internacionales de Bioinformática - UNAM- Enero 2012

Agenda

- Workflows and credible research
- Galaxy: quick overview of the framework
- Sign up to a server and upload/get data
- Add steps to the history and make workflows
- Publish your histories and workflows
- Demo of workflows on our Galaxy server

What is a workflow?

- "A workflow consists of a sequence of concatenated (connected) steps."
- "It is a depiction of a sequence of operations..."

Source: Wikipedia

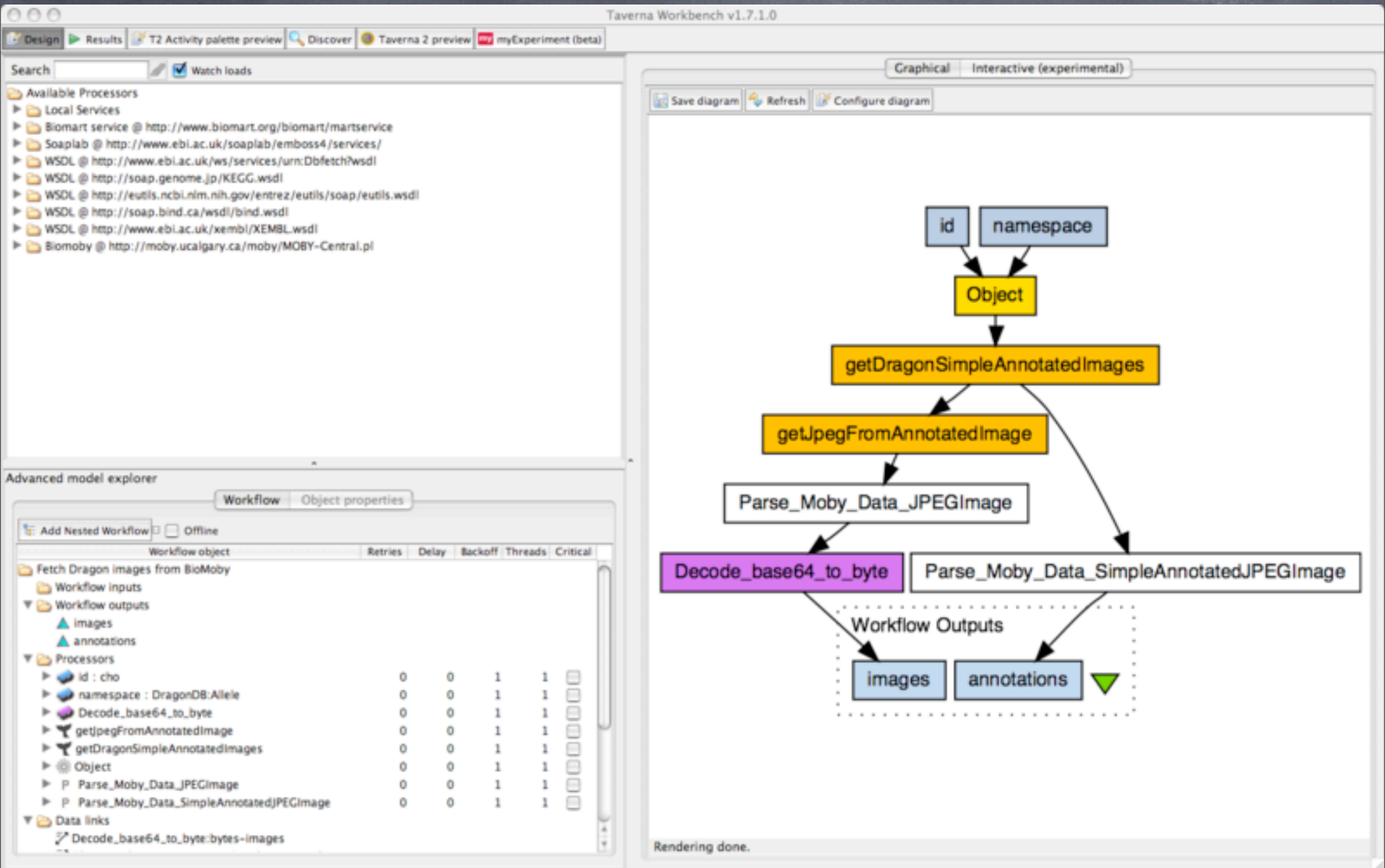
Reproducible research

- How do you judge the quality of research work?
- You need: Paper + Data + computing environment + workflow (Jon Claerbout, Stanford)
- Paper + Data = No longer good enough

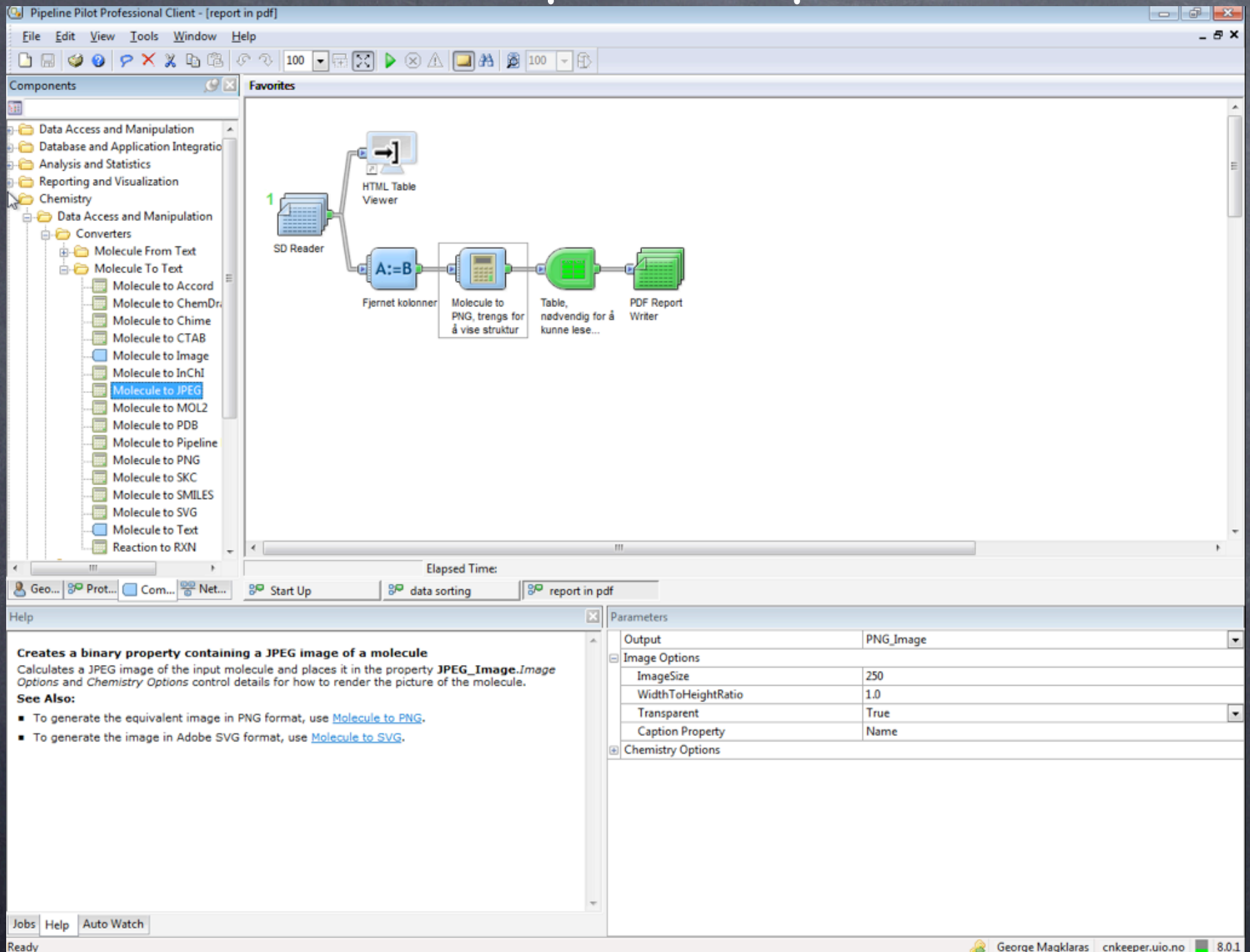
Bioinformatics Workflow Management System

- A BWMS gives you the data and the toolset.
- The workflow as a series of well-defined computational steps.
- Helps you design your processing pipeline and get your results.
- Helps you convince yourself and others about the credibility of your research.

BWMS Example 1: Taverna



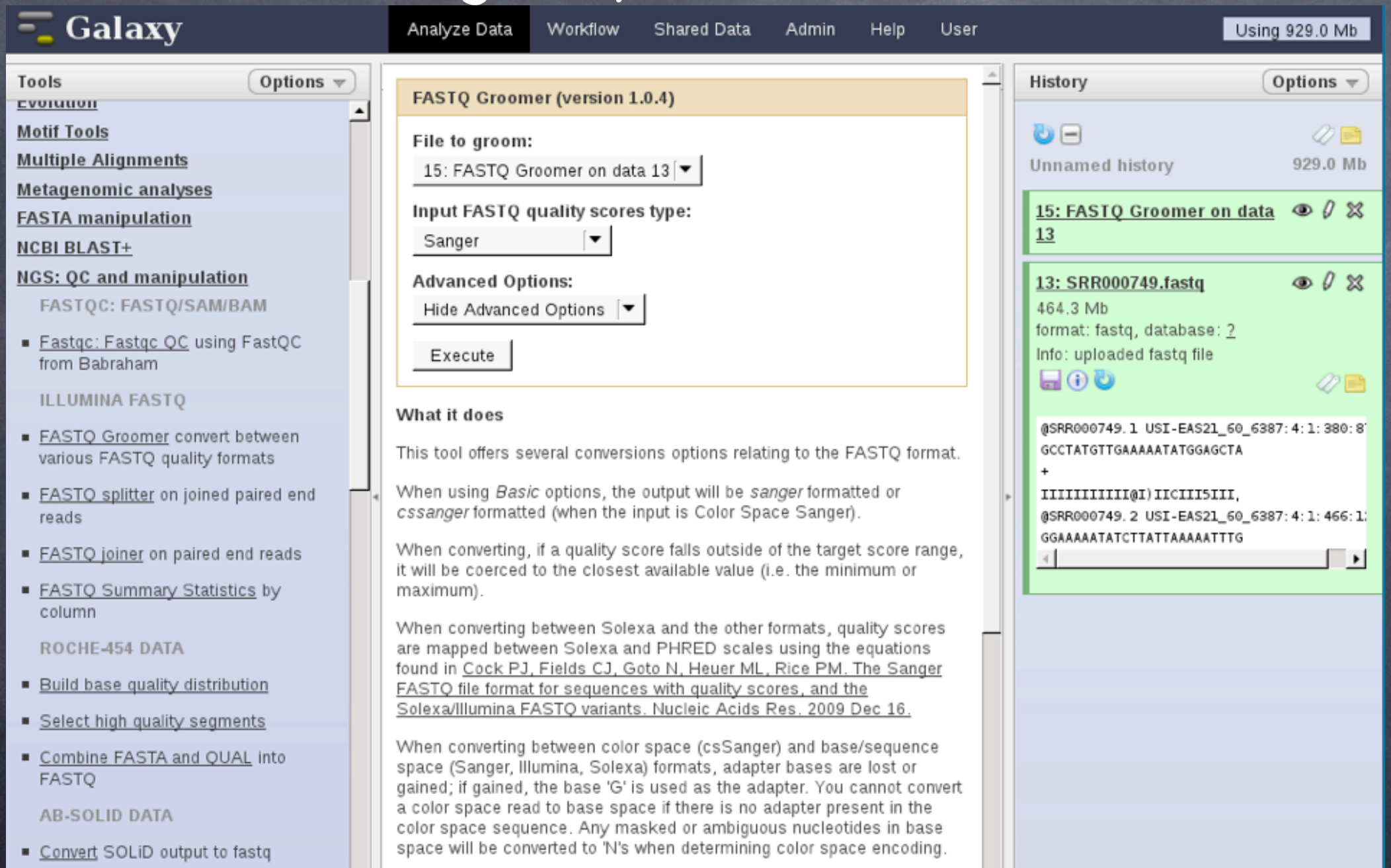
BWMS Example 2: Pipeline Pilot



The Galaxy BWMS

- User friendly: Only requires an up-to-date web browser and an internet connection.
- Contains already a large number of integrated tools for NGS
  https://bitbucket.org/galaxy/galaxy-central/wiki/NGSLocalSetup
- Framework for integrating other tools
  https://community.g2.bx.psu.edu
- Has an active community that develops the base code plus modules

The galaxy web interface



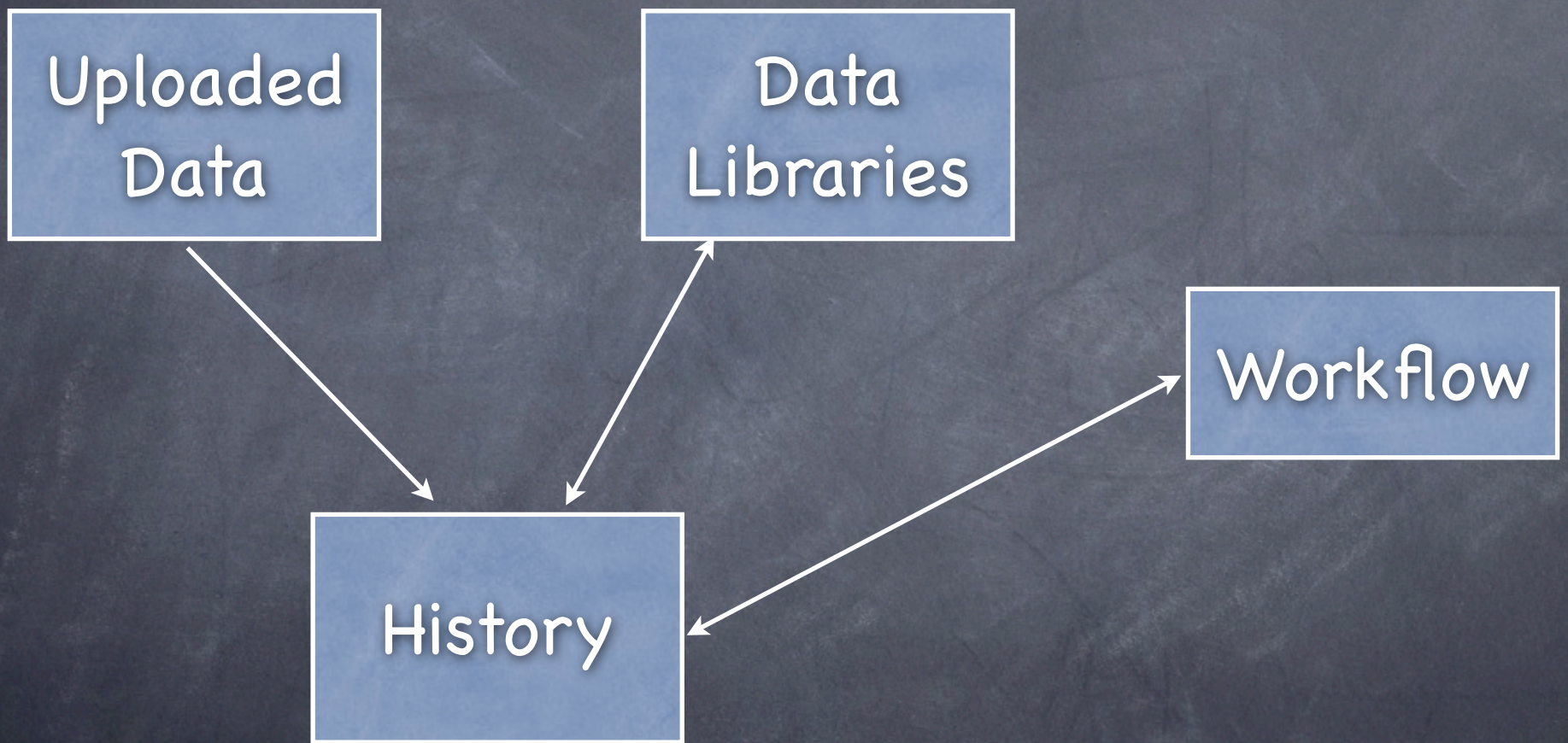
Tool selection | Selected tool form | Workflow (history)

Talleres Internacionales de Bioinformática - UNAM- Enero 2012

Basic Galaxy terminology

- **Analysis step**: A tool which accepts input data and generates output data.
- **History**: All ordered analysis steps plus the data plus the settings on each step.
- **Workflow**: Ordered processing steps without the data ("blueprint" of a history)
- **Datasets**: The input and output data of analysis steps.
- **Data Libraries**: Specific datasets organized for reference.

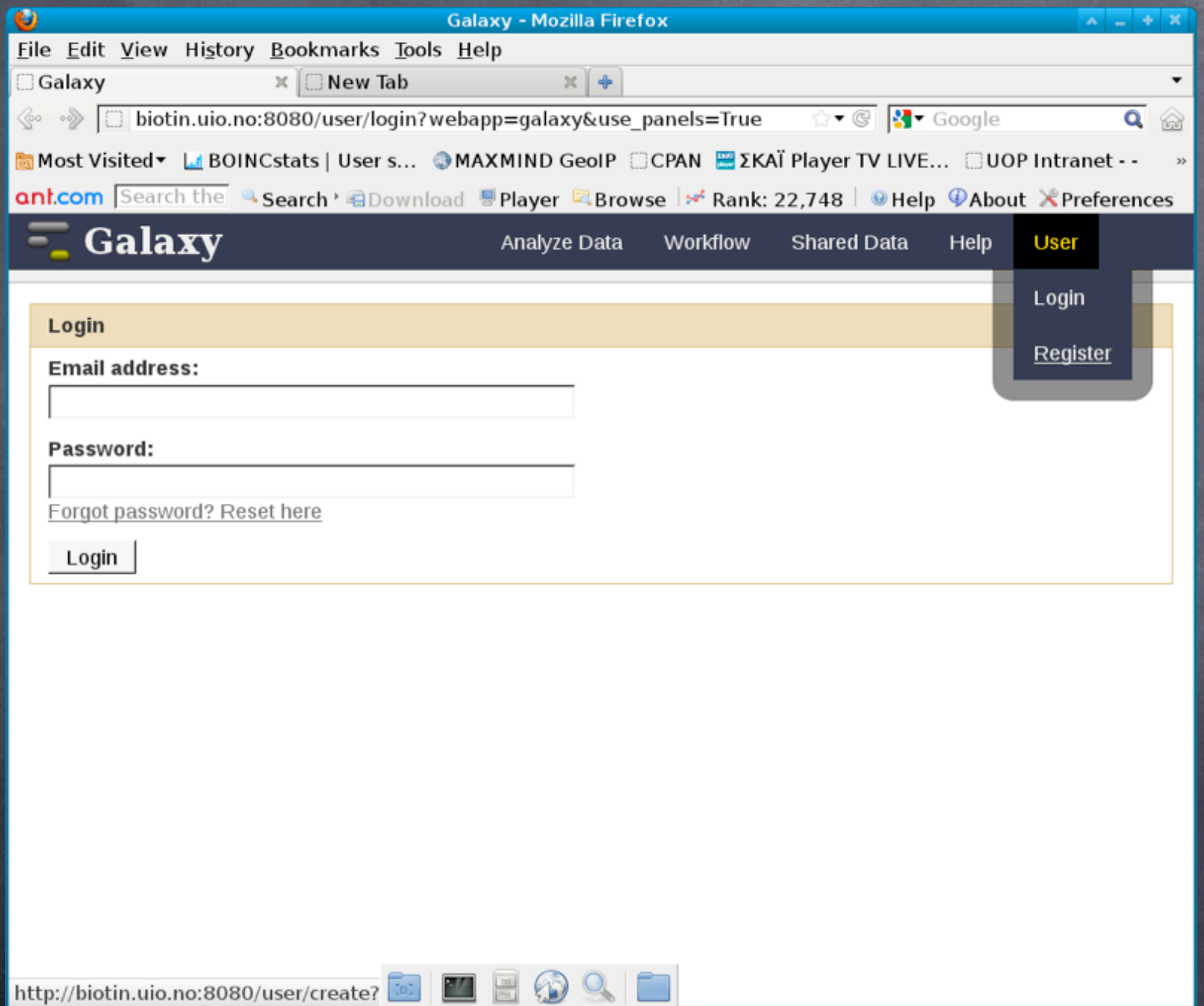
Building blocks and information flow



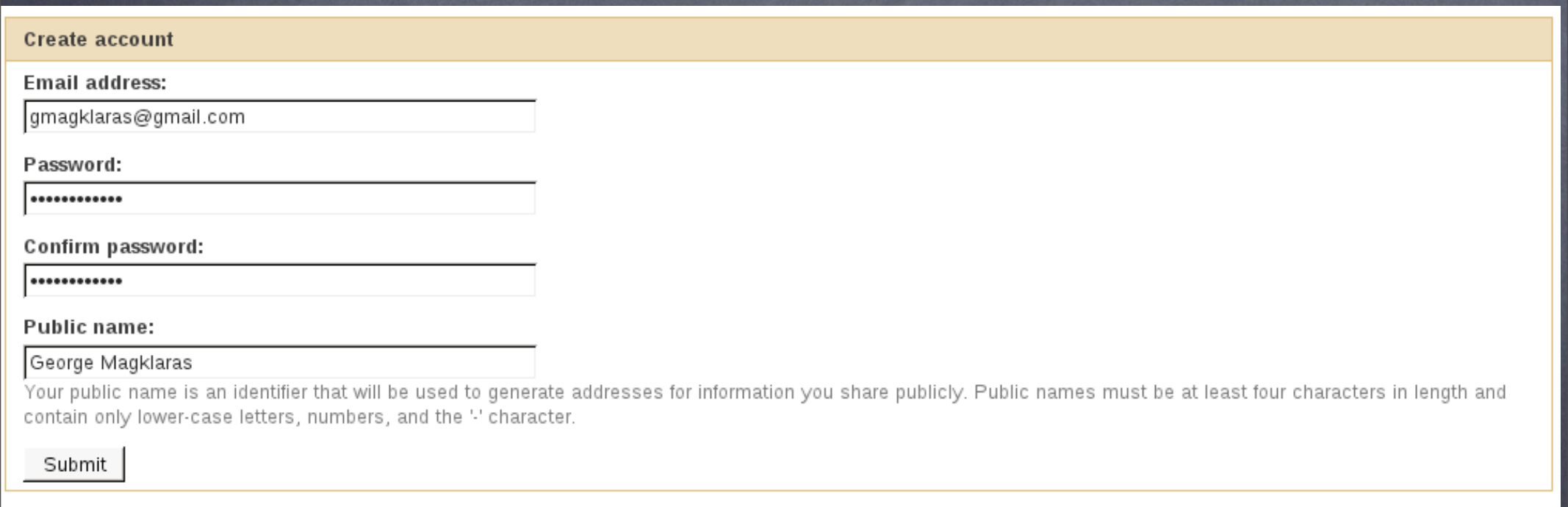
User registration

- You should register in order to get the most out of the Galaxy environment. It allows you to:
  - build and access workflows
  - have access to non-public data files and workflows.
  - Save your datasets and workflow histories.

User registration (2)



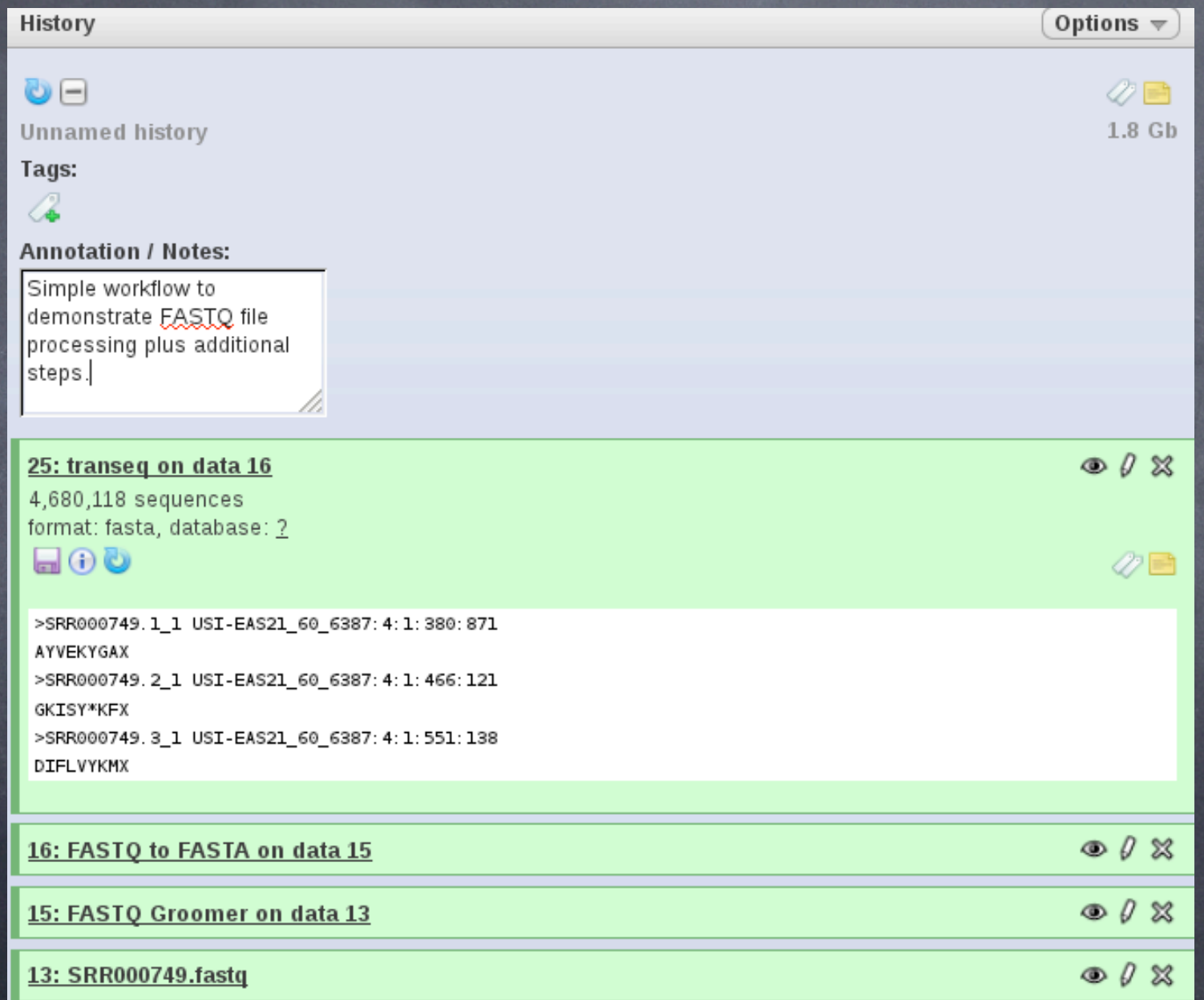
User registration (3)



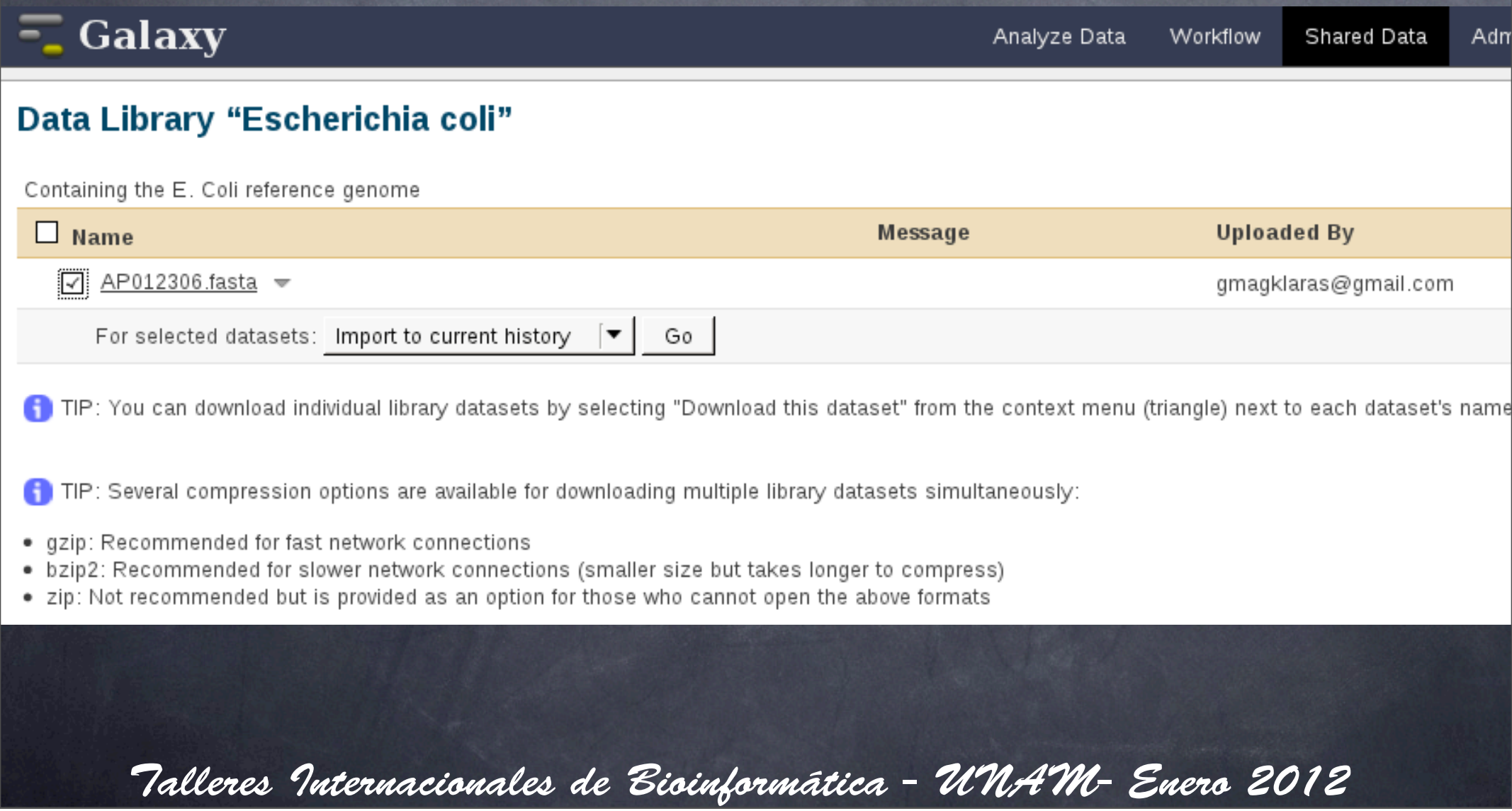
Galaxy "history"

- Your workflow´s scratchpad
- You record and annotate your steps
- You can generate, report, export data to it from data libraries
- You can save, publish your history

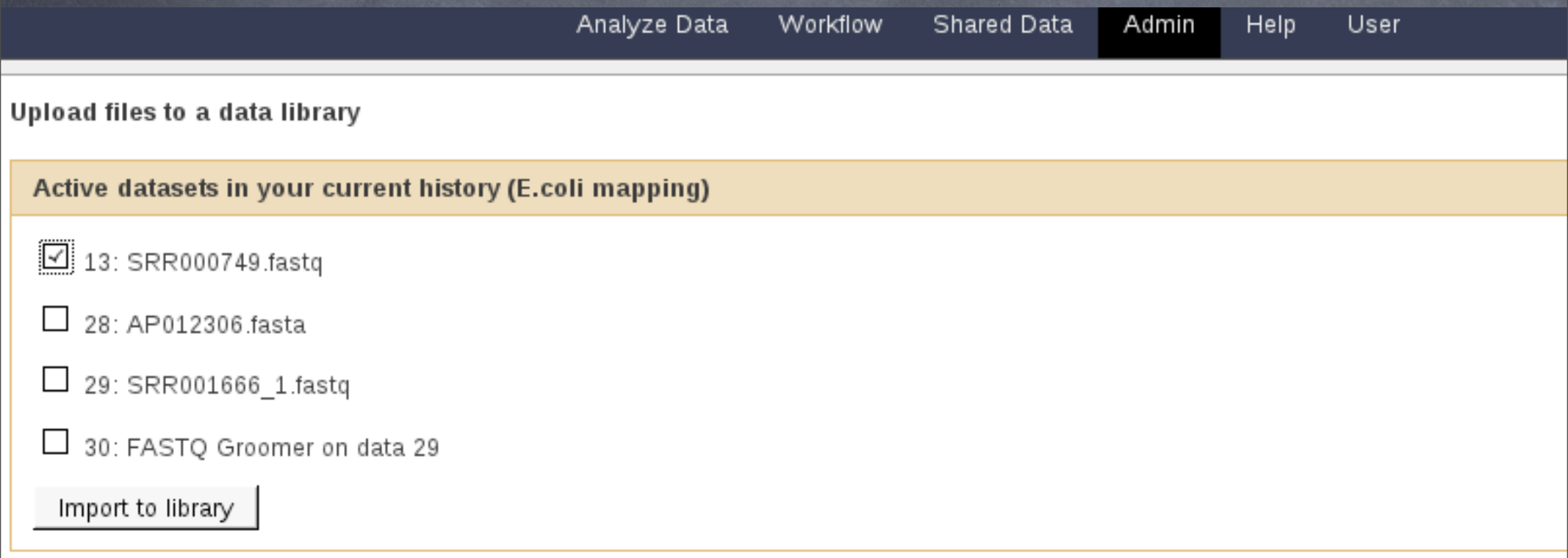
History annotation



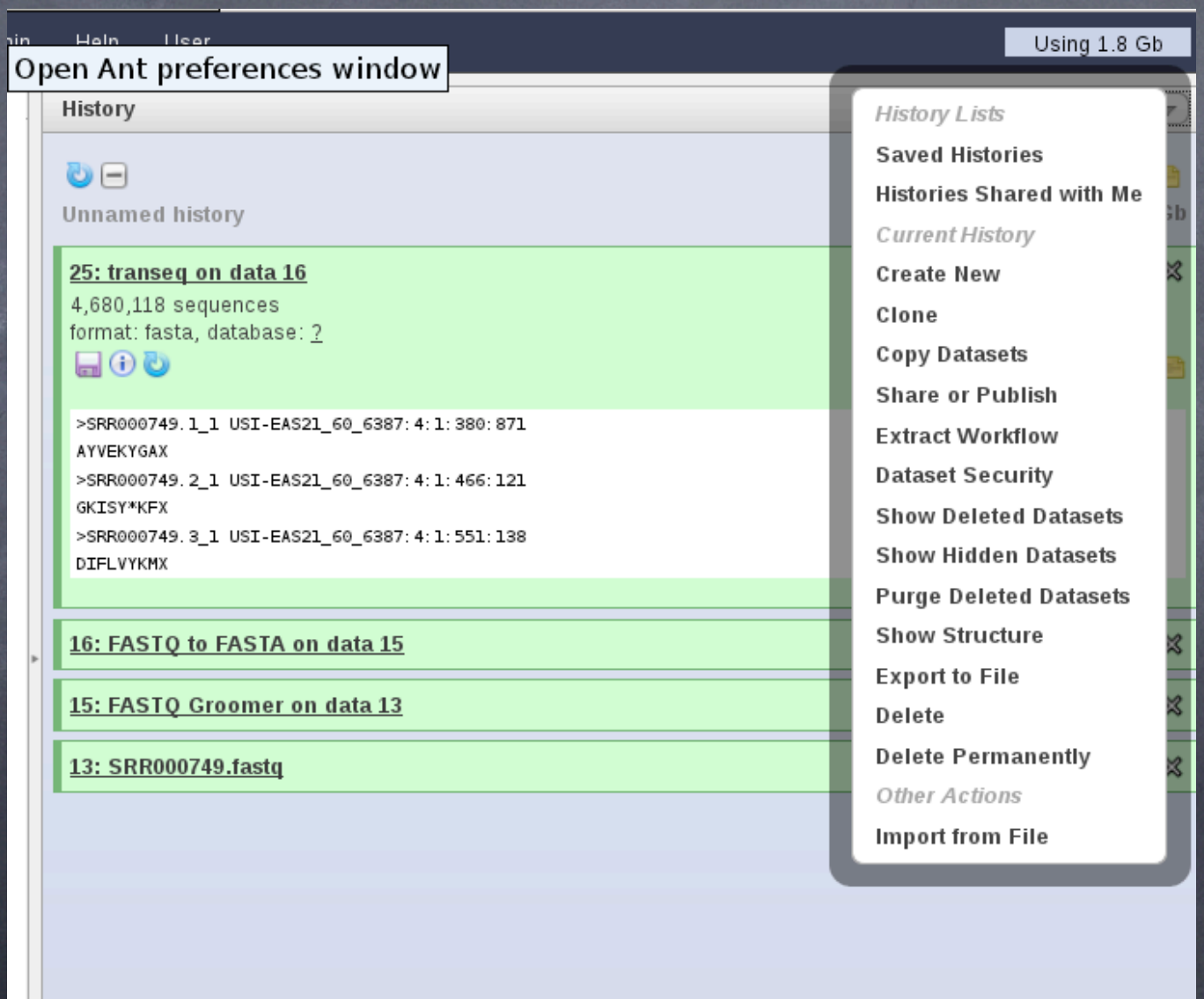
Importing data to history



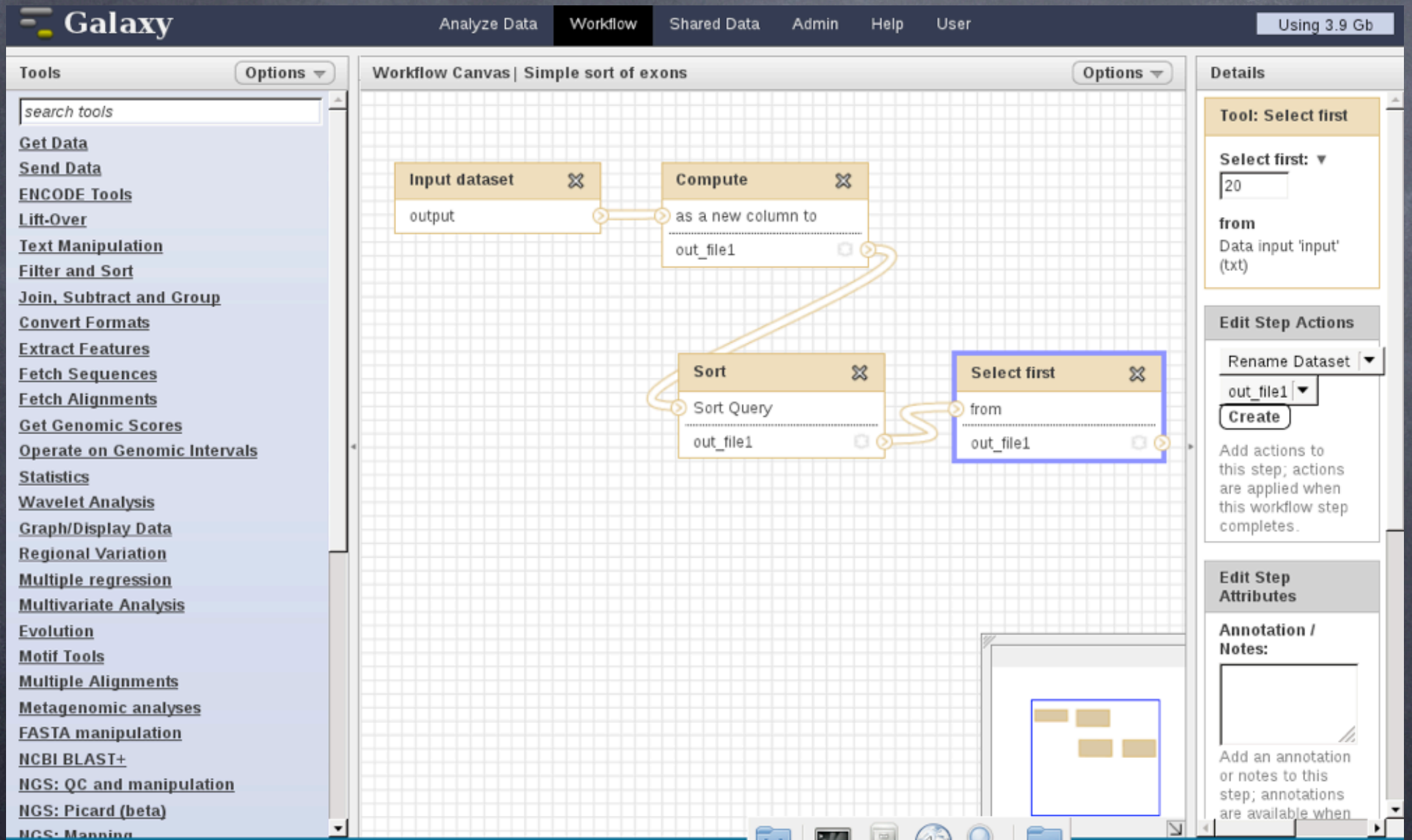
Exporting data from history



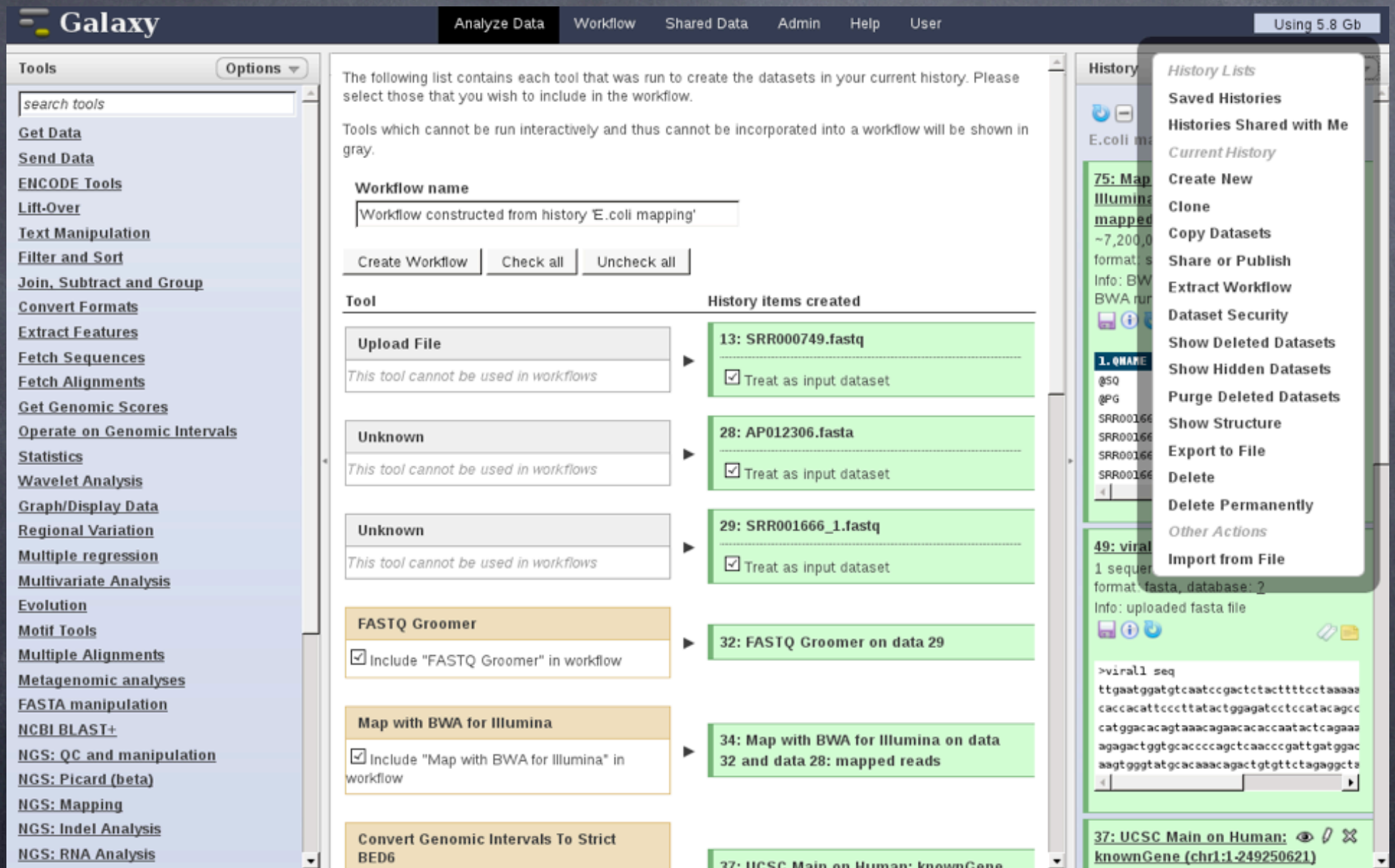
Exporting/publishing a history



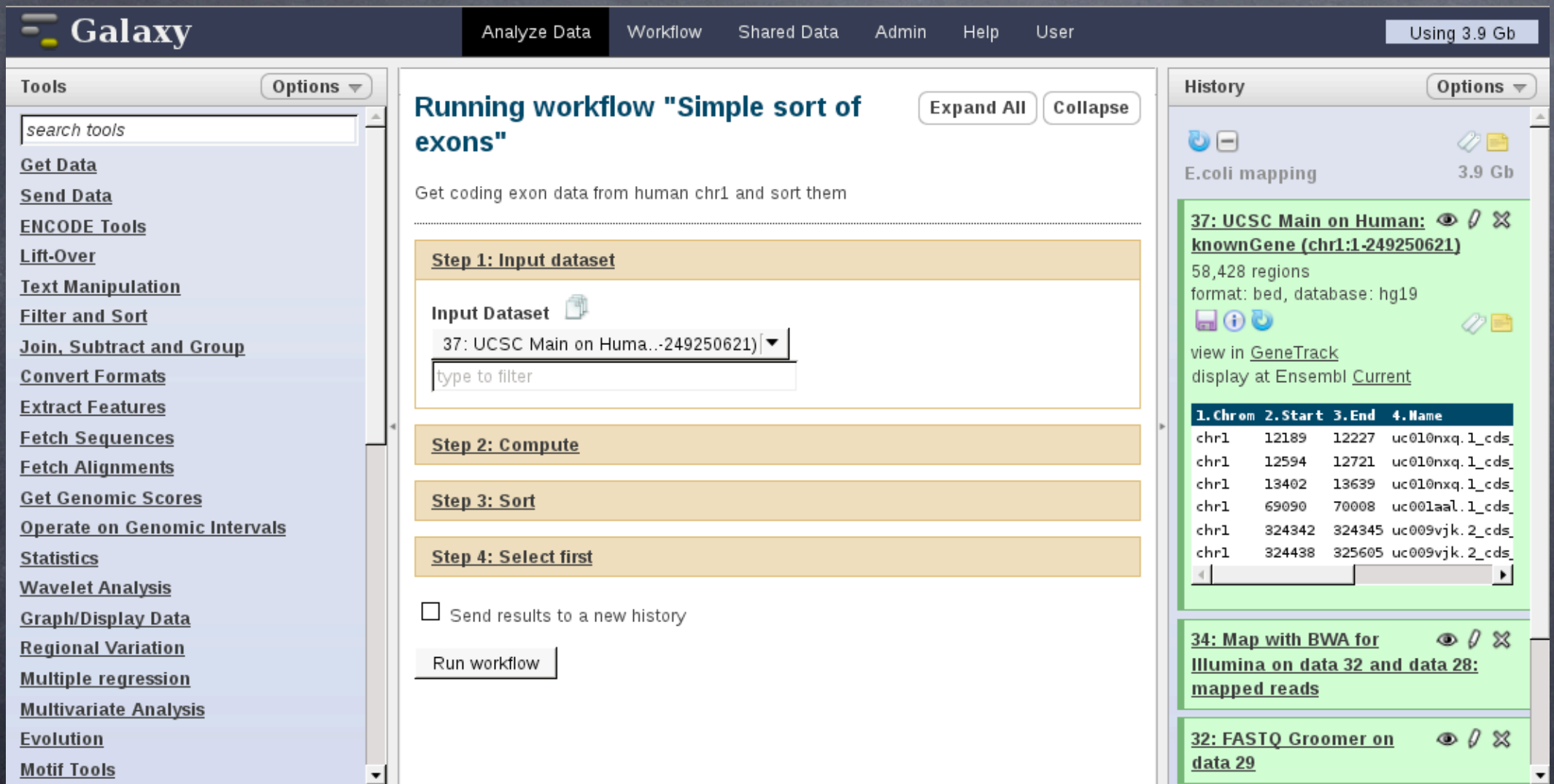
How to make a workflow (1)



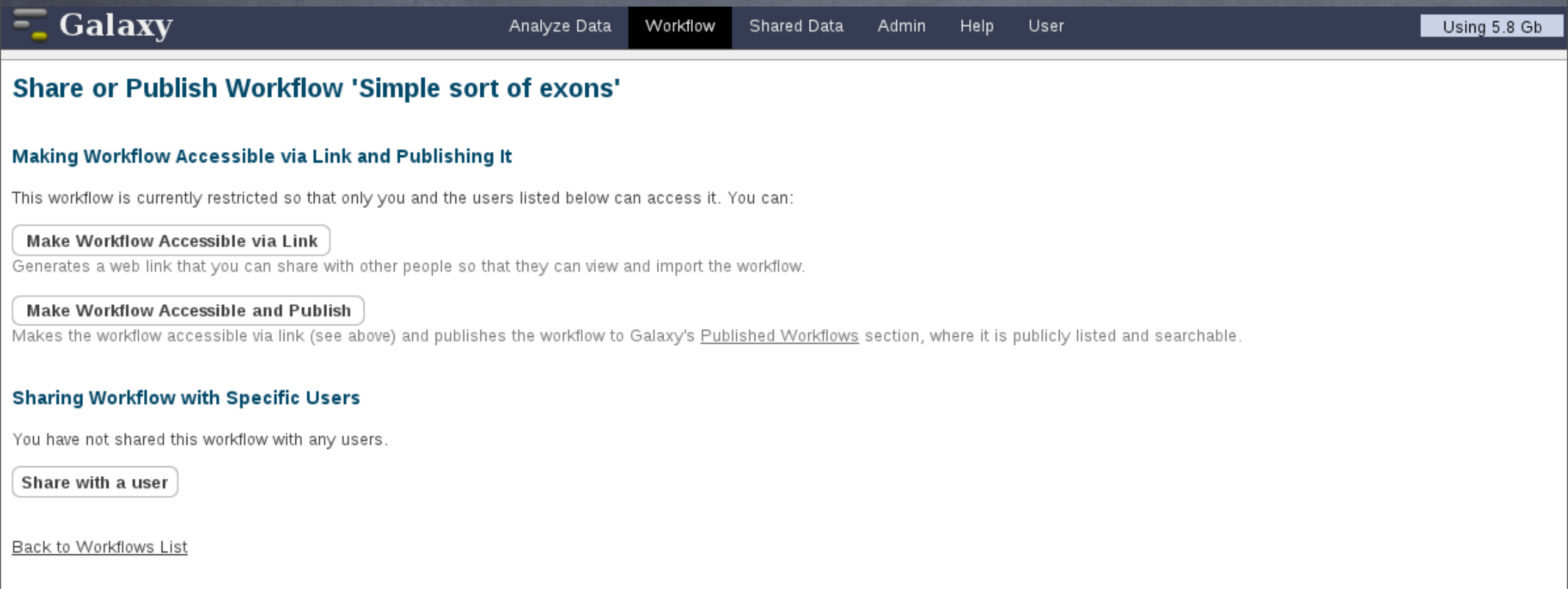
How to make a workflow (2)



Running (using) the workflow

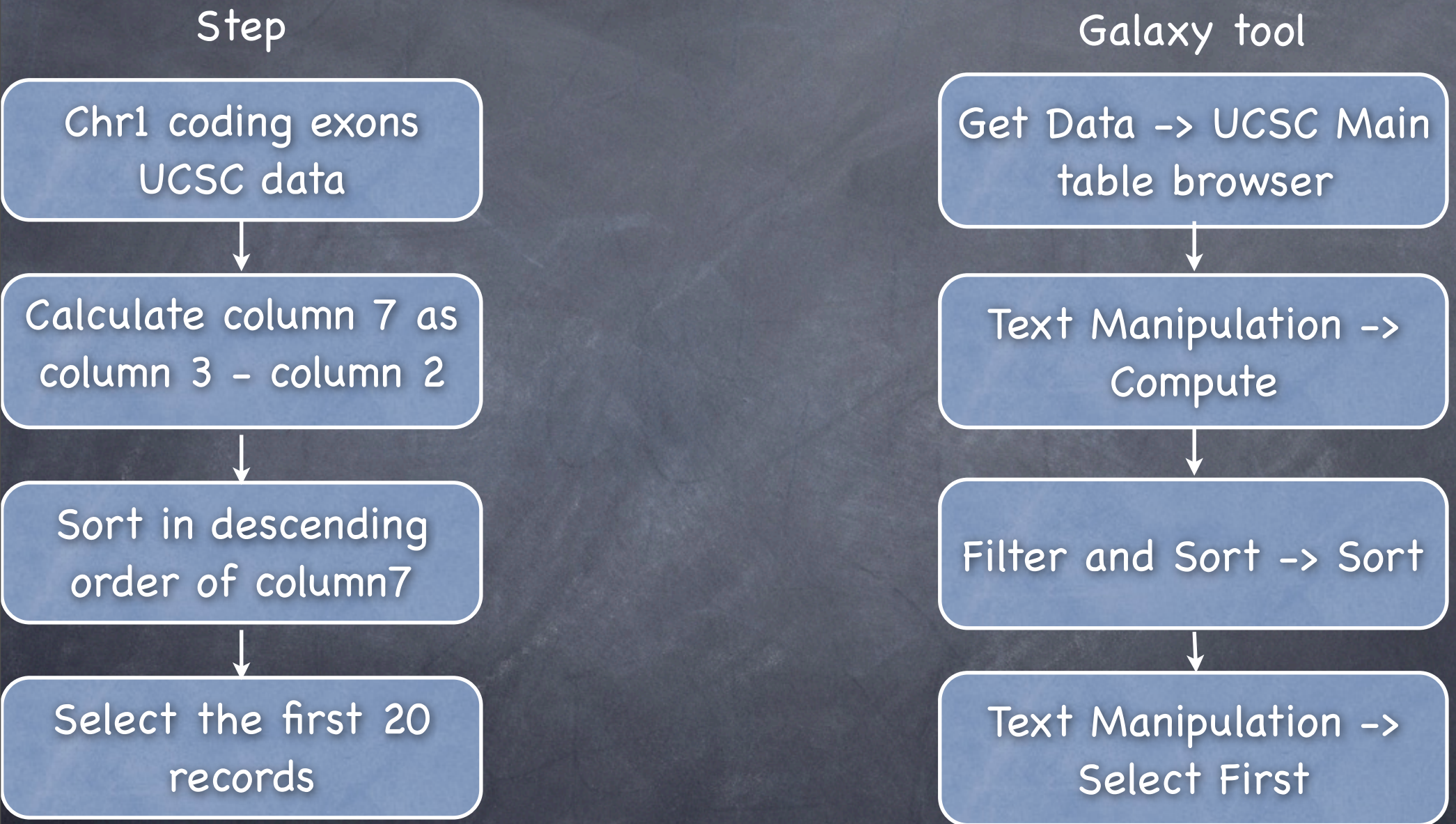


Publishing a workflow

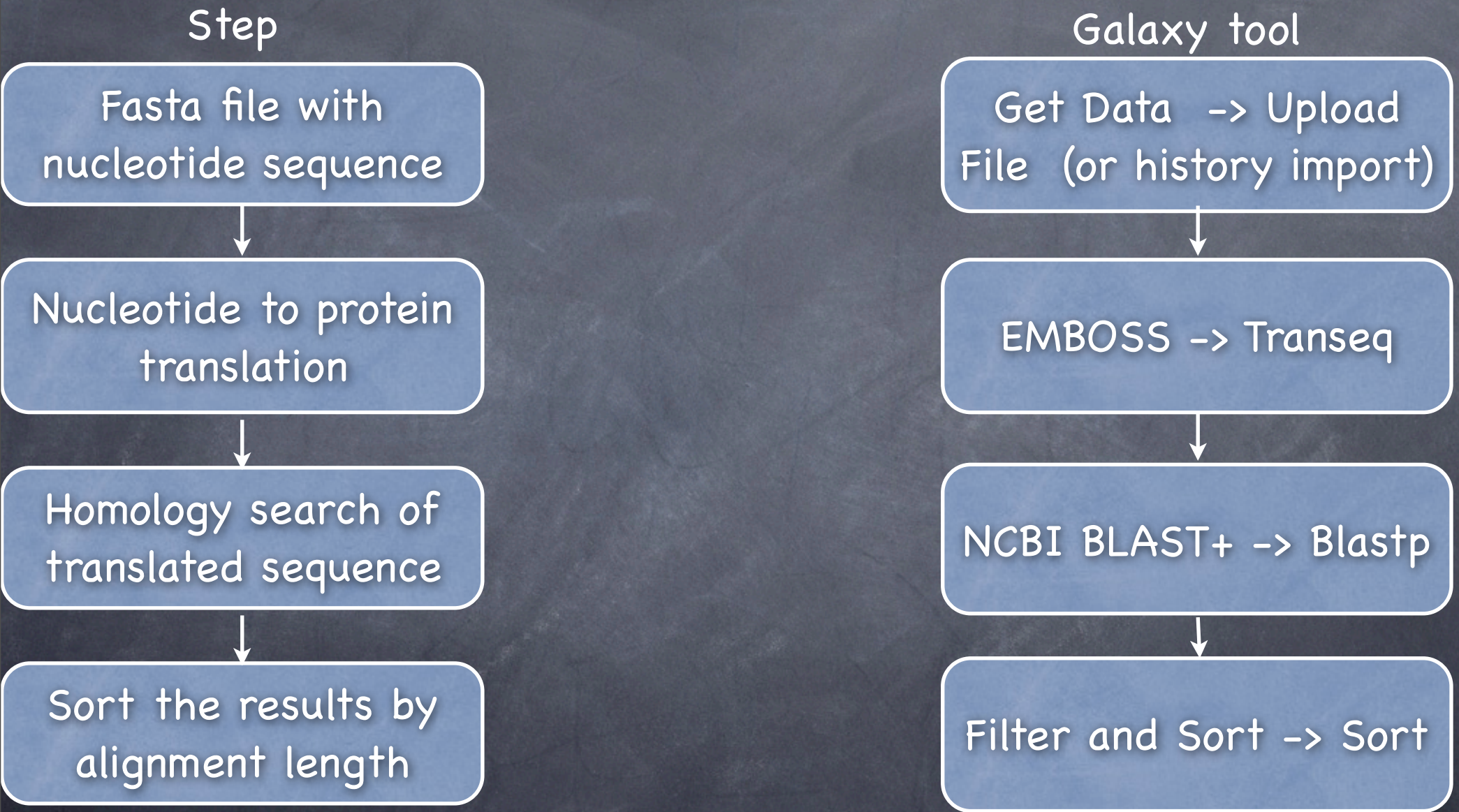


From the top menu bar: "Shared Data" -> "Published Workflows"

DEMO 1: Get coding exons on Human Chr1 and post-process on start and end positions

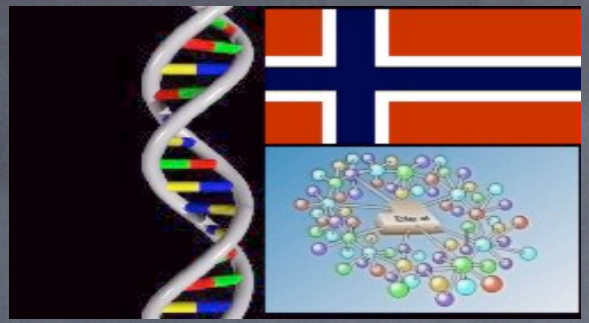


DEMO 2: Characterize a sequence fragment by BLAST search and sort the hits with by alignment length



Exercise

- **Step 1**: User register and login to the Galaxy server (http://biotin.uio.no:8080)
- **Step 2**: Locate the Published Workflow called "NGS example". What does it do?
- **Step 3**:Locate the Published Data Library called "Escherichia Coli" which contains an Illumina experiment and a reference Genome.
- **Step 4**: Import both files from this Data Library into your current history.
- **Step 5**: Now run the "NGS example" workflow with the imported files.
- **Step 6**: Can you add steps to your history, make a workflow out of it and publish it for other users to use?



Questions?

admin@embnet.uio.no

and

http://www.embnet.org/join/ContactRegistration