Resource for bioinformatics analyses and more

Felix HOMA **INRA**
Jean François DUFAYARD **CIRAD**

Bioinformatic in biologist teams

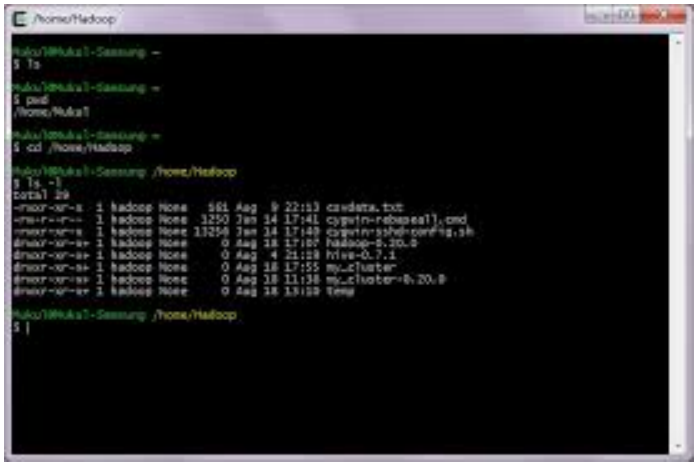
Problem including bioinformatic questions

- Hard to find the right interlocutor
- Reaction times are variable

- He isn't omniscient
- He isn't omnipotent
- His time is limited

The "Do it yourself" procedure

Sometimes the analysis are "simple" and consists in using classical tools



Communication

- Learning basic Unix commands
- Learning cluster specific commands
- Assume the technical and scientific survey
- Deal with file formats
- Deal with data transfers
- Scripting

Galaxy

Galaxy deals with some difficulties of "do it yourself"

Communication

Galaxy

Communication

Galaxy is Free web service integrating a wealth of tools, compute resources, a great volume of data. It decrease the barrier to entry into data analysis by improving accessibility of the Galaxy platform.

Galaxy

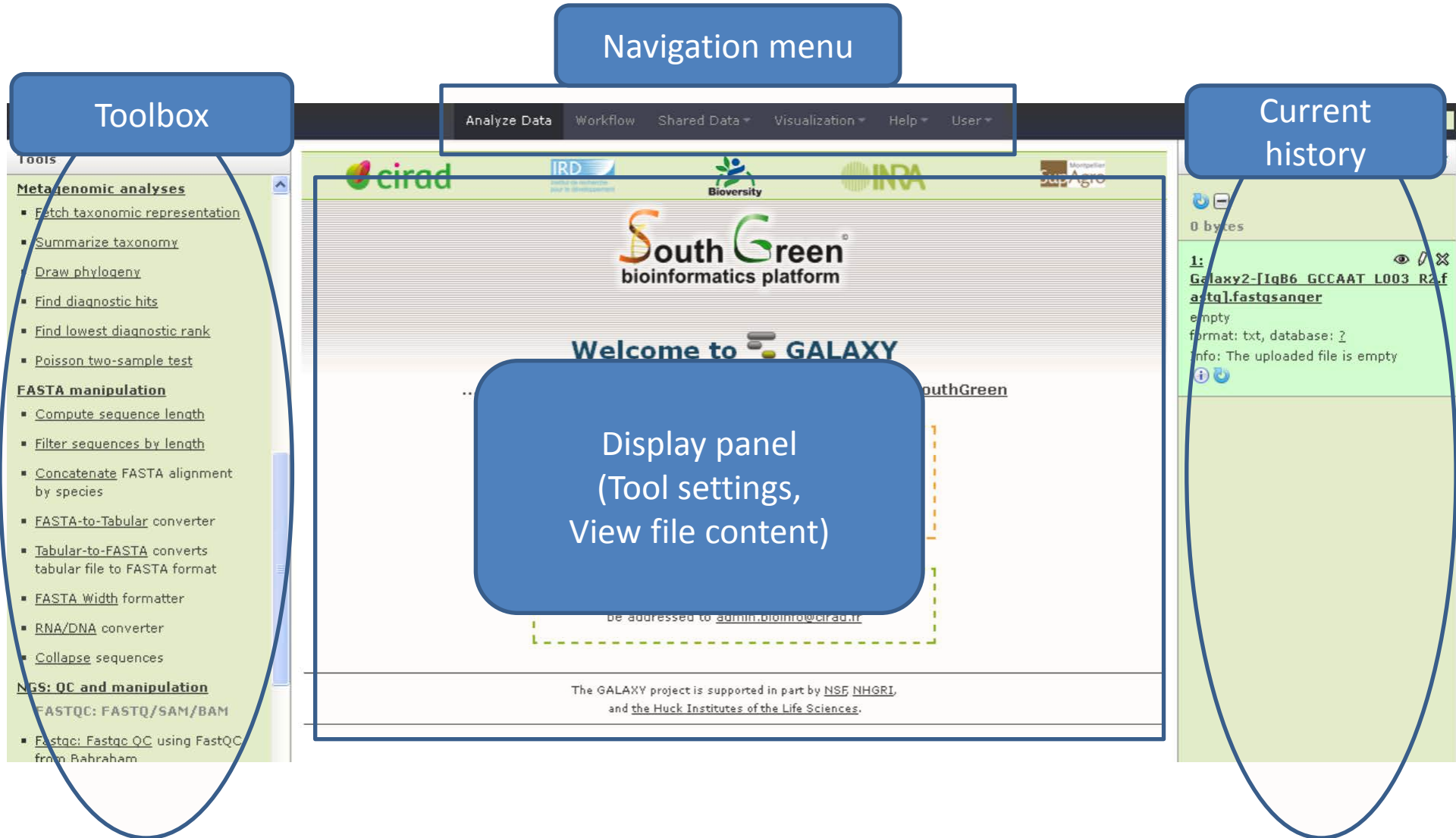
- **Accessibility**
  - User friendly
  - Every tool under the same interface
  - No programming skill is needed

- **Reproductibility**
  - Allows to remake analysis and obtain expected result
- **Transparency**
  - Analysis can be documented and published
- **Integrative**
  - Includes many tools (computer science, statistics, text manipulation, visualization…)
- **Widely used** : galaxy project Screencasts

Objectives

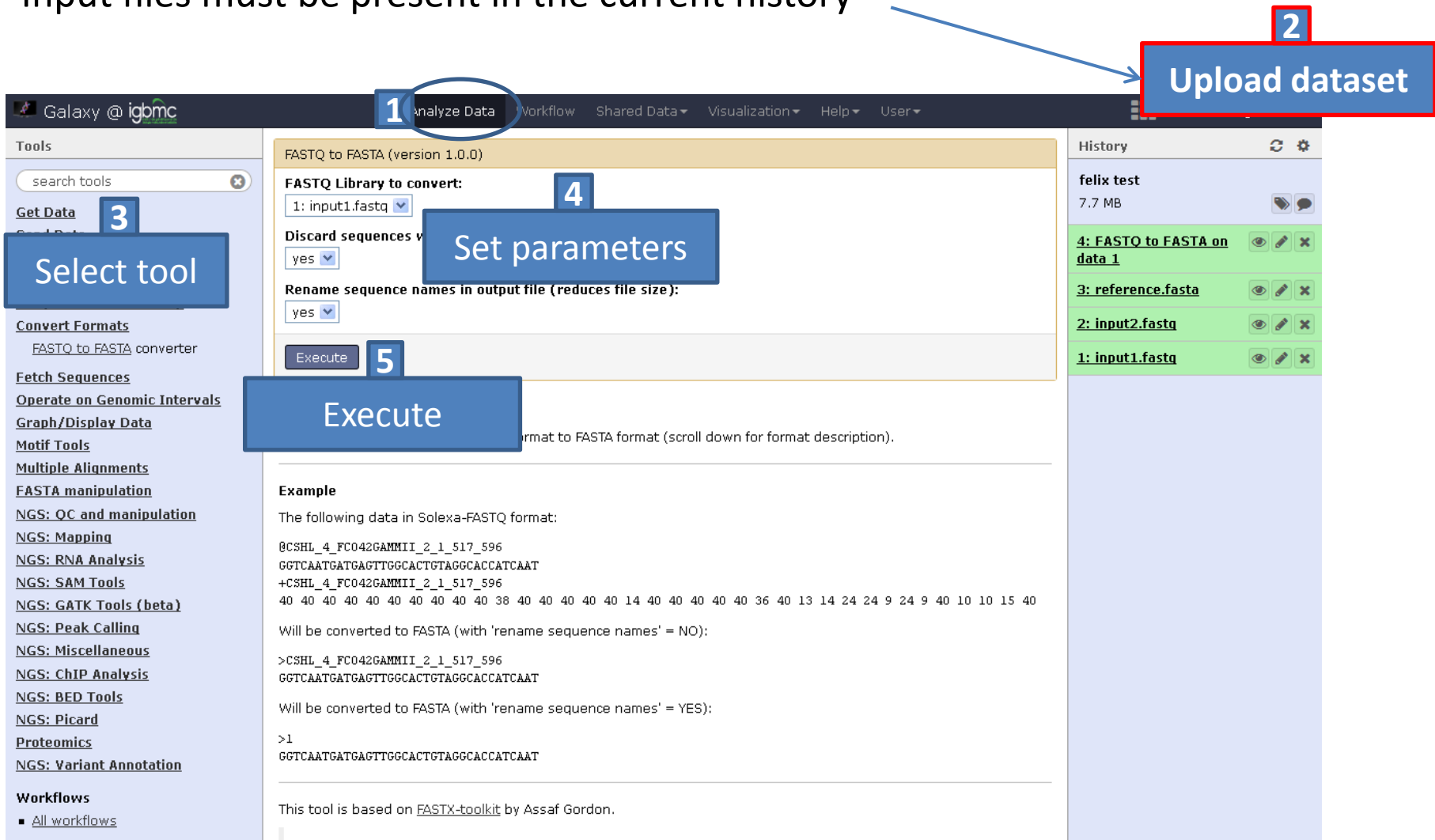
- *Objectives of training:*
  - To have a first **overview** of Galaxy user interface
  - To be able to **manage data** in Galaxy
  - To be able to **create** and **use Workflows**
  - To be able to **share galaxy items**
  - **This session is not about SPECIFIC ANALYSIS**

Galaxy : Global view

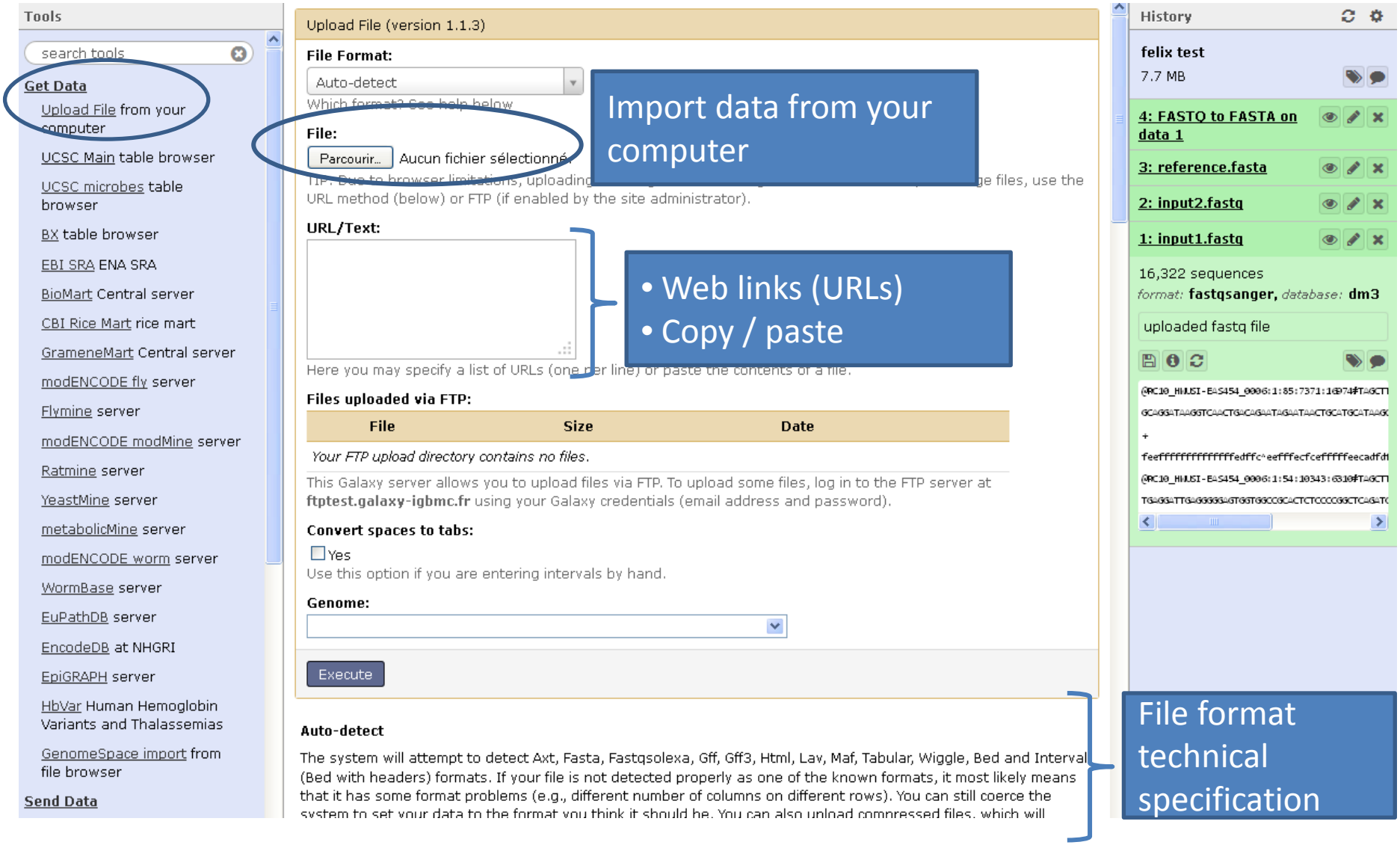


Galaxy : First run

Input files must be present in the current history

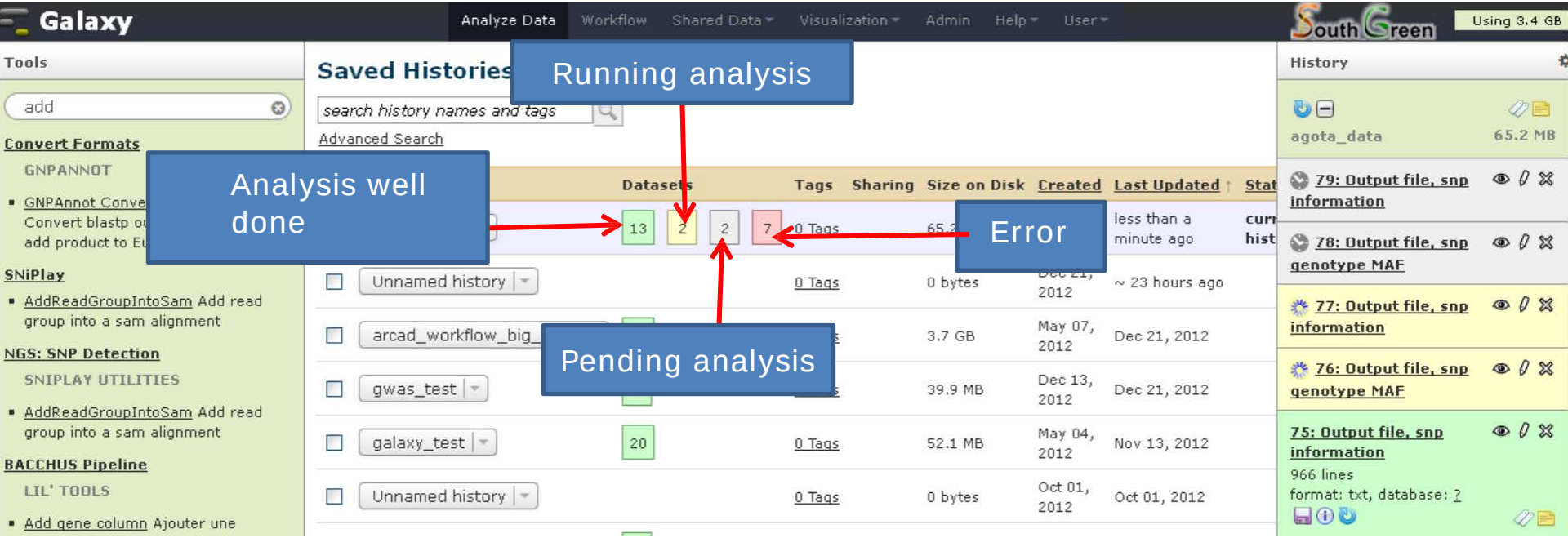


Galaxy : Upload a dataset

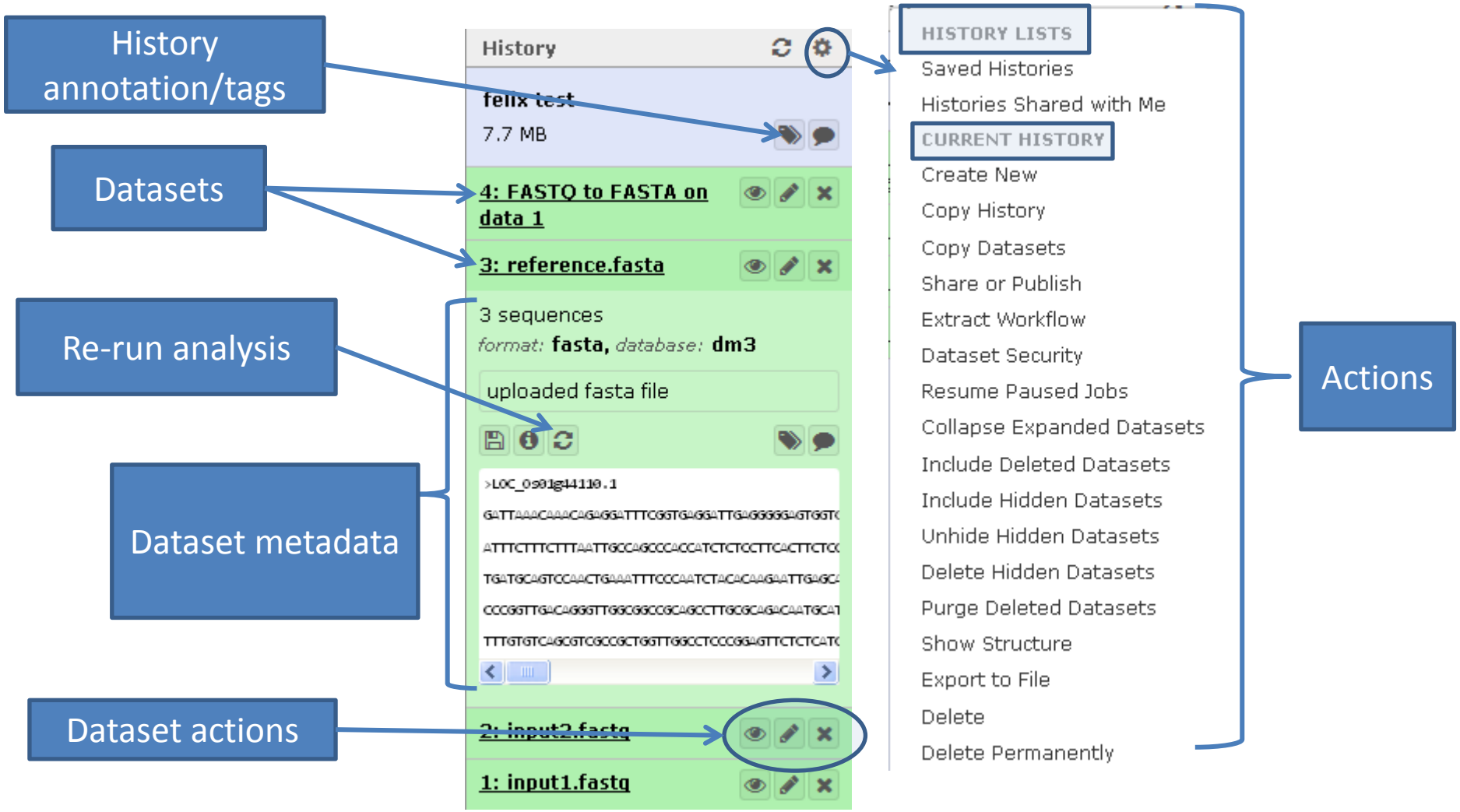


Galaxy : Color code

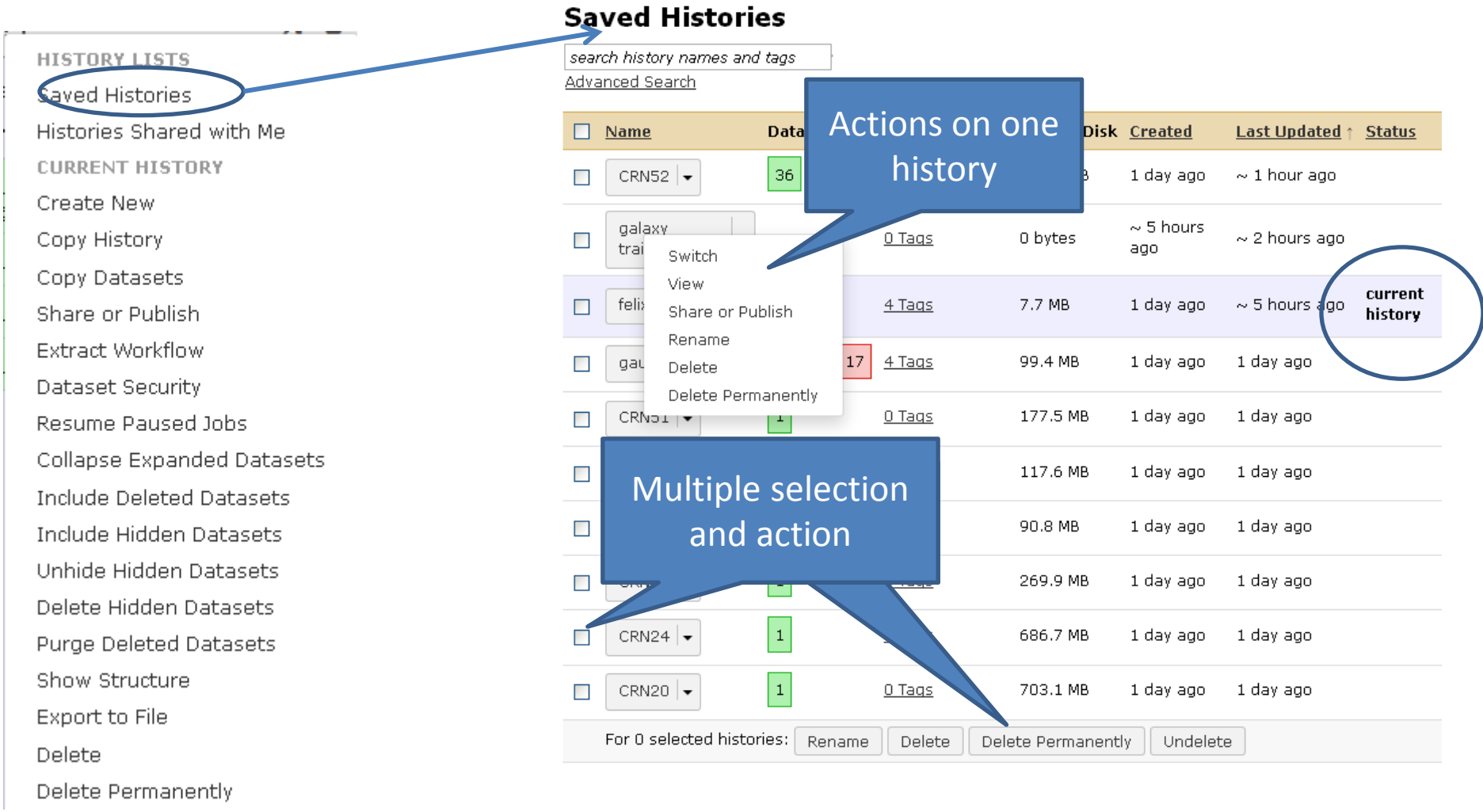
Galaxy associate a color for each step of analysis



Galaxy : Current history/datasets



Galaxy : Histories management



Galaxy : Best practices

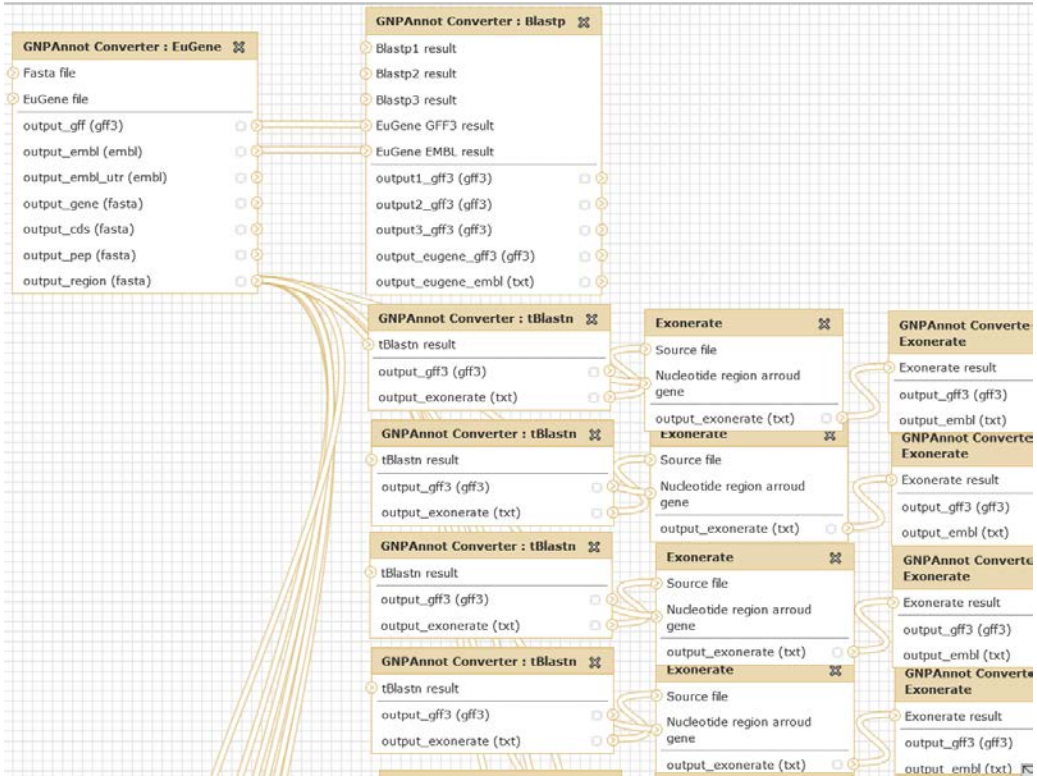
- Never use any **accents** or **special characters** in any metadata of Galaxy : source of bugs
- Create new history for each new analysis
- Always give an **explicit name** to interesting histories/datasets : 2 histories/datasets can wear the same name
- **Delete** not helpful Galaxy items: thingk about community
- Always annotate with **tags** interesting histories and datasets : allow easy access to data, improve reproductibility of analysis

Practice

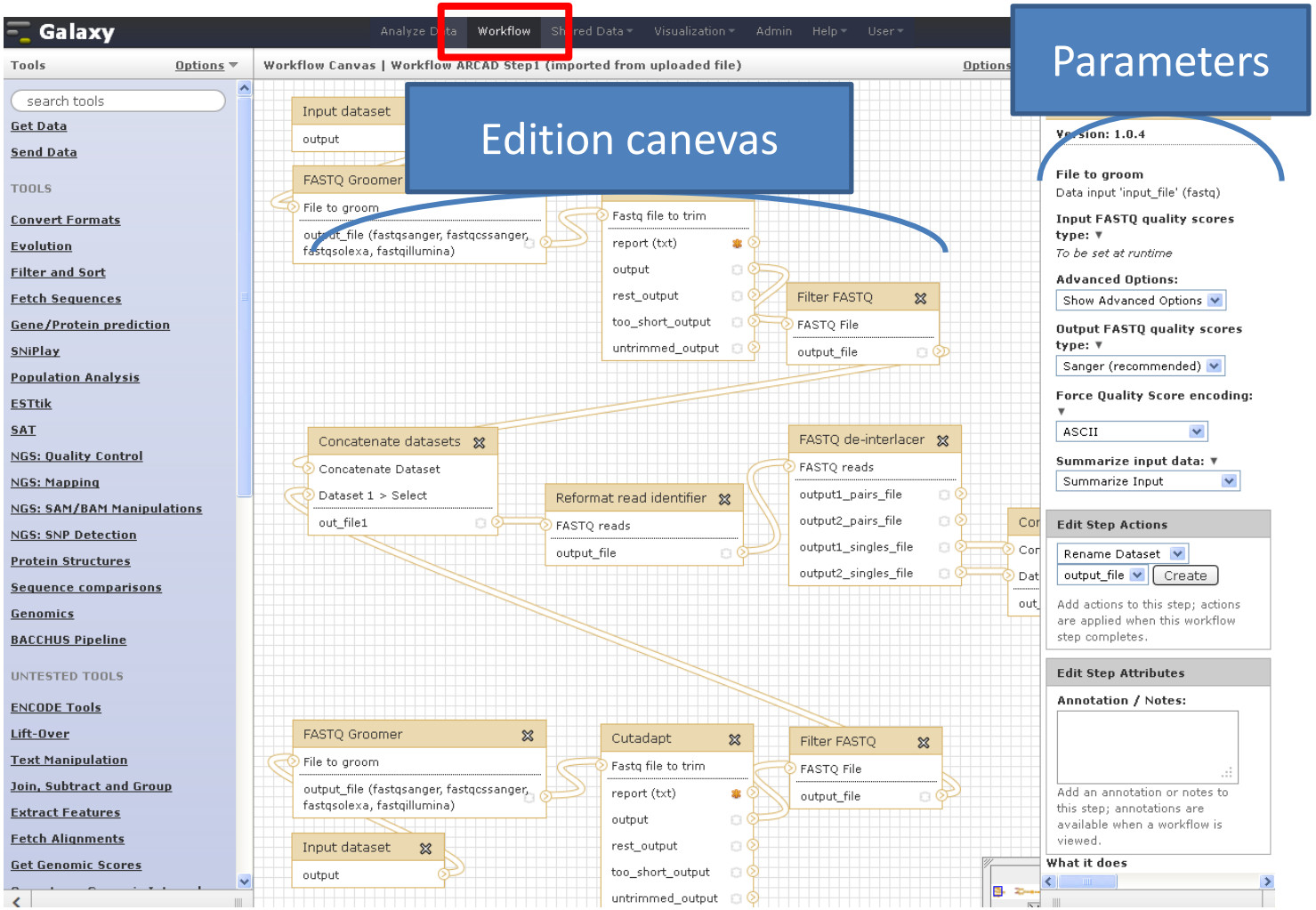
Practical session: Step 1
Changement de format
Création, renommage d'historique

Workflows

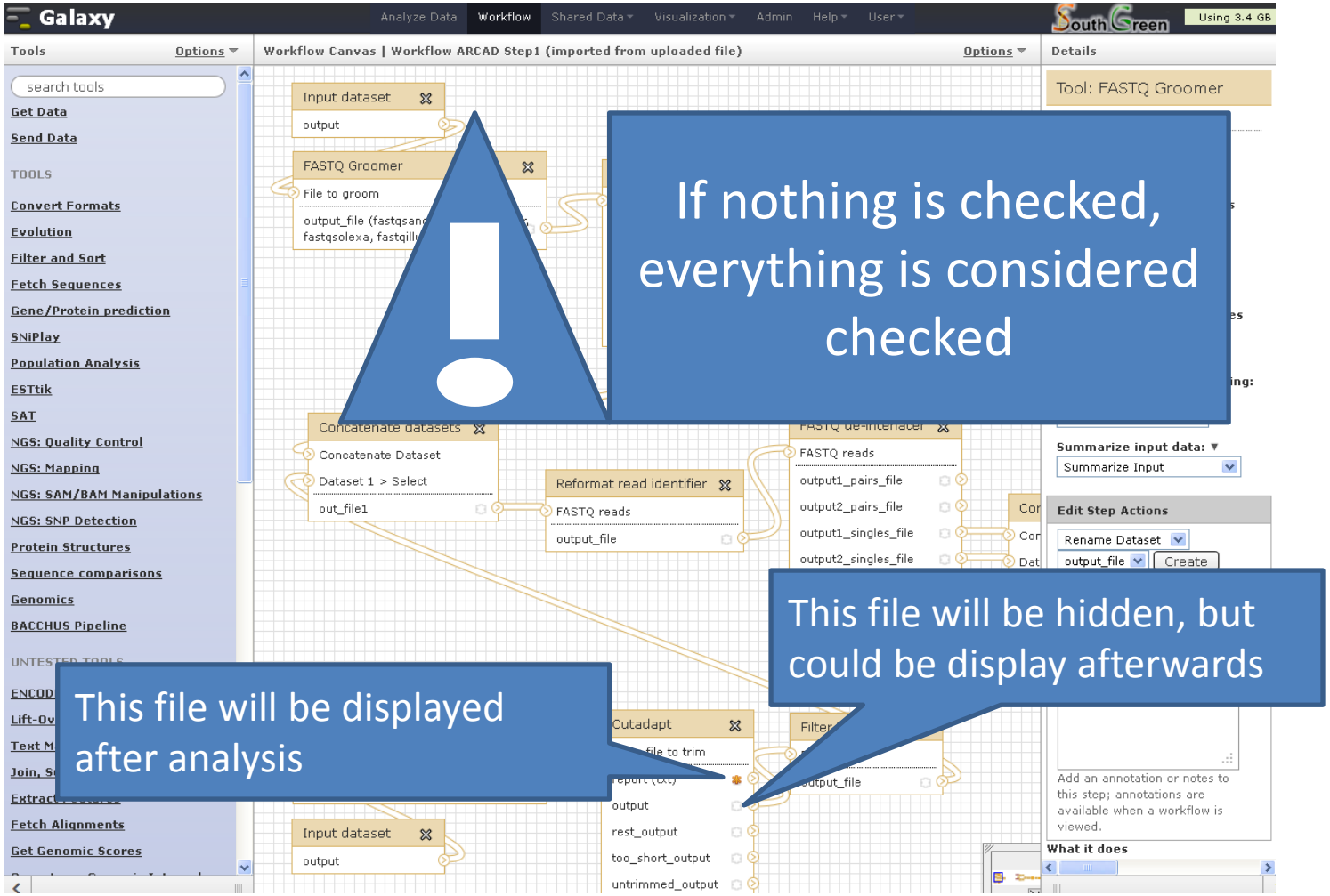
- Means a simple or very complex series of operations, carried out automatically. Operations refer to analysis steps, data reformatting steps.



Workflow : build and configure



Workflow : build and configure

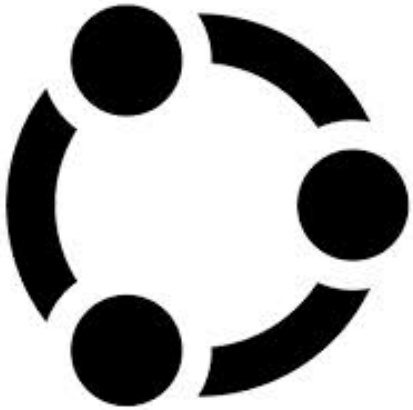


Practice

Practical session: Step 2

Shareable Items

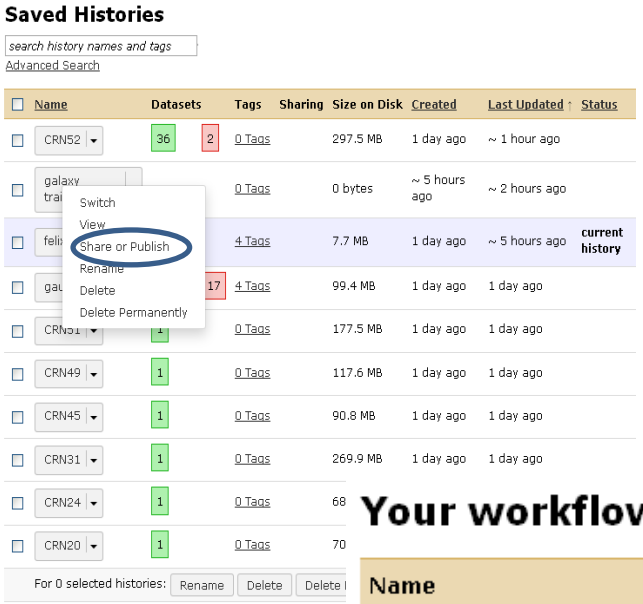
- Galaxy items
  - History
    - Datasets
  - Workflow
  - Pages

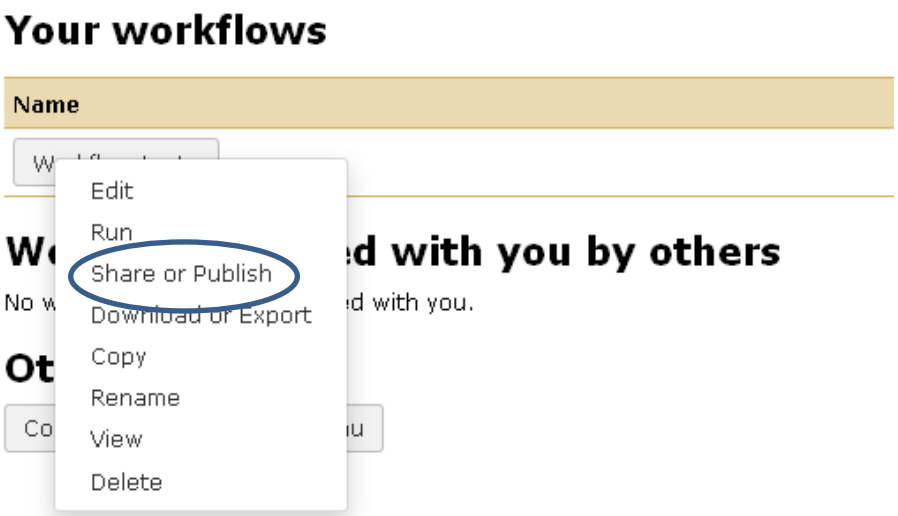


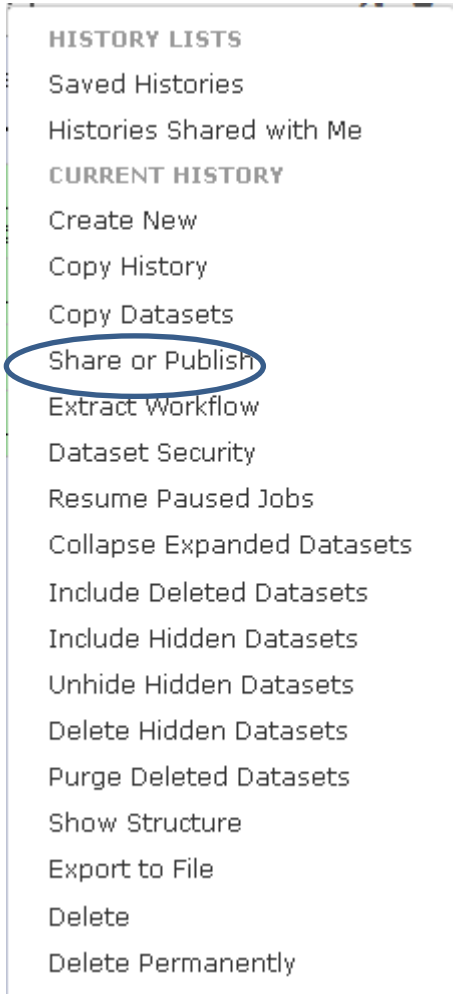
Sharing with galaxy

- Share Galaxy items (3 ways)
  - **By link**: allowed users can make a local copy in their own workflow space, using provided link
  - **By publication**: every user can access to your workflow in the menu “Shared Data ->Published Workflows”
  - **By email**: share the workflow directly with a specific individual user

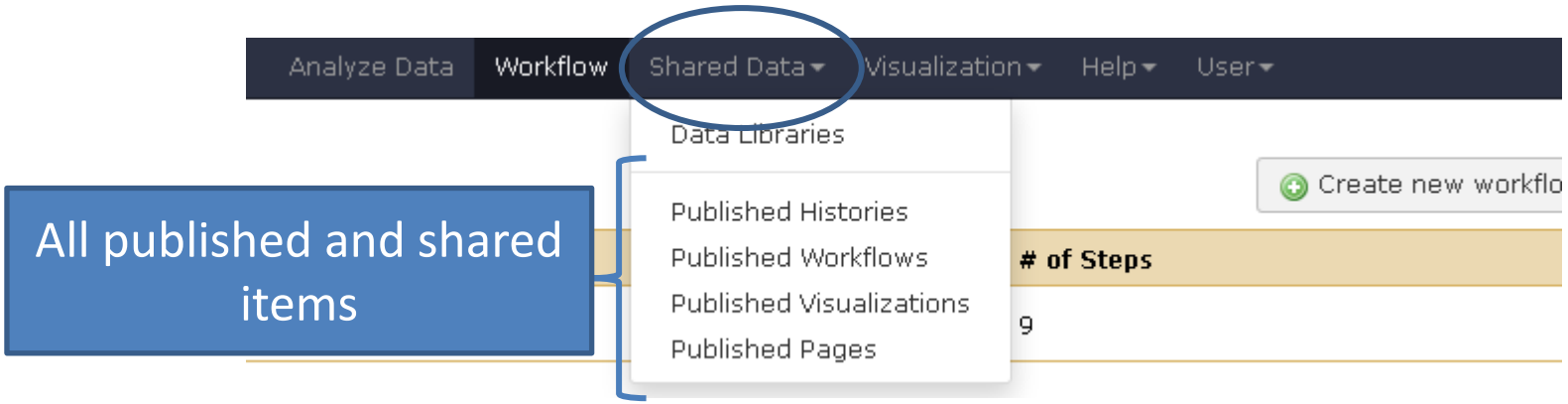
Sharing with galaxy







Sharing with galaxy : access to shared data



Practice

Practical session: Step 3

More with Galaxy

- Security of data
  - Manage permission of shared data
- Use Galaxy pages to initiate scientist paper
- Install tools and new visualization
- Custom Galaxy look and feel
- Use Galaxy with the cloud
- Interaction Galaxy and homemade tools