

Galaxy Interactive Environments

Björn Grüning and Eric Rasche

2015-07-08

Interactive Environment (IEs)

Interactive
Environments

BG & ER

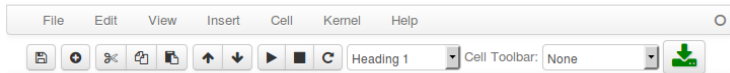
- New way to interact with your data, without leaving Galaxy
- Flexible, efficient, extensible, and interactive
- Full spectrum of use; Teaching, Research, and Development

Your favourite Data Science tools...

Interactive
Environments

BG & ER

IP[y]: Notebook ipython_galaxy_notebook



Welcome to the interactive Galaxy IPython Notebook.

You can access your data via the dataset number. For example, `handle = open(get(42), 'r')`. To save data, write your data to a file, and then call `put('filename.txt')`. The dataset will then be available in your galaxy history. Notebooks can be saved to Galaxy by clicking the large green button at the top right of the IPython interface.

More help and informations can be found on the project [website](#).

In [1]:

... Inside of Galaxy

Interactive
Environments

BG & ER

The screenshot displays the Galaxy web interface. At the top, a dark navigation bar contains the 'Galaxy' logo and several menu items: 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. Below this, the main interface is divided into three vertical panels. The left panel, titled 'Tools', features a search bar and a list of tool categories including 'Get Data', 'Send Data', 'Lift-Over', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Convert Formats', 'Extract Features', 'Fetch Sequences', 'Fetch Alignments', 'Statistics', and 'Graph/Display Data'. The middle panel, titled 'IP[y]: Notebook ipython_galaxy_notebook', has a menu bar with 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', and 'Help'. It contains a toolbar with various icons for file operations and a text area with the following text: 'Welcome to the interactive Galaxy IPython Notebook.' followed by instructions on how to access data via dataset numbers and save data to a file. Below the text is a code input field with the prompt 'In [1]:'. The right panel, titled 'History', shows a search bar and a list of datasets. The first entry is 'Unnamed history' with 1 shown and 2 previewed, totaling 47 bytes. Below this is a green box labeled '3: Pasted Entry' containing a table with 4 lines of data: 'format: txt, database: 2' and 'uploaded txt file'. The table has two columns, 'T' and 'A', with the following values: (1, 2), (3, 4), and (5, 6).

Galaxy

Analyze Data Workflow Shared Data Visualization Admin Help User

Tools

search tools

Get Data
Send Data
Lift-Over
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Fetch Alignments
Statistics
Graph/Display Data

Workflows
• All workflows

IP[y]: Notebook ipython_galaxy_notebook

File Edit View Insert Cell Kernel Help

Welcome to the interactive Galaxy IPython Notebook.

You can access your data via the dataset number. For example, `handle = open(get(42), 'r')`. To save data, write your data to a file, and then call `put('filename.txt')`. The dataset will then be available in your galaxy history. Notebooks can be saved to Galaxy by clicking the large green button at the top right of the IPython interface. More help and informations can be found on the project [website](#).

In [1]:

History

search datasets

Unnamed history
1 shown, 2 previewed
47 bytes

3: Pasted Entry
4 lines
format: txt, database: 2
uploaded txt file

T	A
1	2
3	4
5	6

IE Demonstration

Interactive
Environments

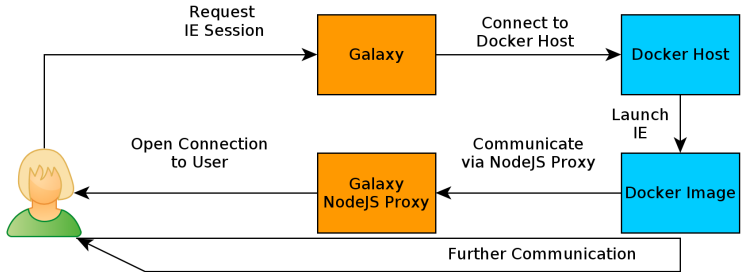
BG & ER

- We'll demonstrate some analysis ...
- `www.youtube.com/watch?v=U0FFkDuJxgk`

How does this magic work?

Interactive
Environments

BG & ER



Who should use IEs?

Interactive
Environments

BG & ER

- Everyone!
- IPython/RStudio are great for bioinformaticians and Data Scientists
- The upcoming iobio visualization IEs are great for Life Scientists

Why use IEs instead of ... Galaxy Tools/Viz?

Interactive
Environments

BG & ER

- Tools are not one-size-fits-all
- Visualisations are restrictive
- Complete freedom!

Why use IEs instead of ... “normal” service deployments?

Interactive
Environments

BG & ER

- IPython Notebooks are stored as history elements
- Notebooks are re-runnable, maintaining reproducibility
- API interactions required to access data are all 100% transparent
- Transparently integrates with standard Galaxy deployments and authentication schemes
- Notebooks are rendered into HTML for easy viewing/sharing, without launching an IE

Teaching

Interactive
Environments

BG & ER

- Ideal for teaching:
 - Researchers
 - Bioinformatics
 - Data Wrangling
 - and Scientific Programming
- Share notebooks with students inside of Galaxy
- Use “iterate programming” in IPython to teach students how analyses work, line-by-line

Research

Interactive
Environments

BG & ER

- Reproducible and transparent scripts
- Share “hotfix” scripts easily between bioinformaticians and researchers

Development

Interactive
Environments

BG & ER

- Rapidly prototype new scripts and tools for your organisation
- Immediately test them on your large, real datasets
- Does an existing visualisation not meet your goals?
Build a new one immediately in IPython/RStudio.

IEs

Interactive
Environments

BG & ER

Available Now

- IPython (included in Galaxy 15.05/Cloudman)
- RStudio (coming in Galaxy 15.07)

Coming Soon

- iobio BAM
- iobio VCF

In the Works

- Apache Zeppelin
- WebApollo
- Jupyter 3/4 (Python/R/Julia/Perl/Ruby)

IPython IE Features

Interactive
Environments

BG & ER

- Baked in Bioblend access to Galaxy
- Easily get data from/put data into Galaxy
- Bash and R “magics”
- Pre-installed: numpy biopython scikit-learn pandas
scipy sklearn-pandas bioblend matplotlib patsy
pysam khmer dendropy ggplot mpld3 sympy rpy2

RStudio IE Features

Interactive
Environments

BG & ER

- Easily get data from/put data into Galaxy
- R version 3.2.1
- Knitr/Sweave available
- R Packages: RCurl, XML, markdown, shiny, ggvis, dplyr, ggplot2, plyr, reshape2, devtools, RODB, maps, pheatmap, readr, tidyr, dplyr, RJSONIO, shinyapps, knitr
- Bioconductor: edgeR, Rgraphviz, biomaRt, topGO, limma, DESeq2, cummeRbund, Biostrings, GenomicRanges, Rsamtools, affy

Thanks

Interactive
Environments

BG & ER

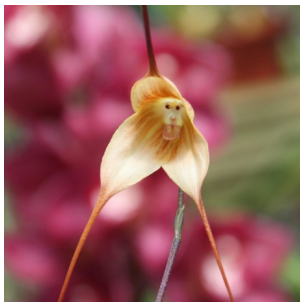
A big thank you to:

- John Chilton for his help getting the IE codebase merged into Galaxy originally
- Enis Afgan for getting the IEs into Cloudman
- the Galaxy Team for supporting this exciting new feature we've developed.

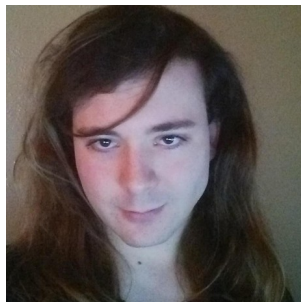
Q&A

Interactive
Environments

BG & ER



`github.com/bgruening`



`github.com/erasche`

- IPython <http://bit.ly/gxIEipython>
- RStudio <http://bit.ly/gxIERstudio>