

# Making your life easier by embracing common concepts and tools to support repeatable computational experiments

C. Titus Brown

[ctbrown@ucdavis.edu](mailto:ctbrown@ucdavis.edu)

# Goals

- Introduce concepts
- Discuss ways to integrate & some experience
- Q&A!

# URLs

- Computing Workflows for Biologists: a Roadmap
  - <http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002303>
- Oslo workshop on Reproducibility
  - [ivory.idyll.org/blog/2016-norway-repeatability.html](http://ivory.idyll.org/blog/2016-norway-repeatability.html)
- mybinder.org

workflow  
engine

+

data  
analysis  
narrative

+

execution  
environment

+

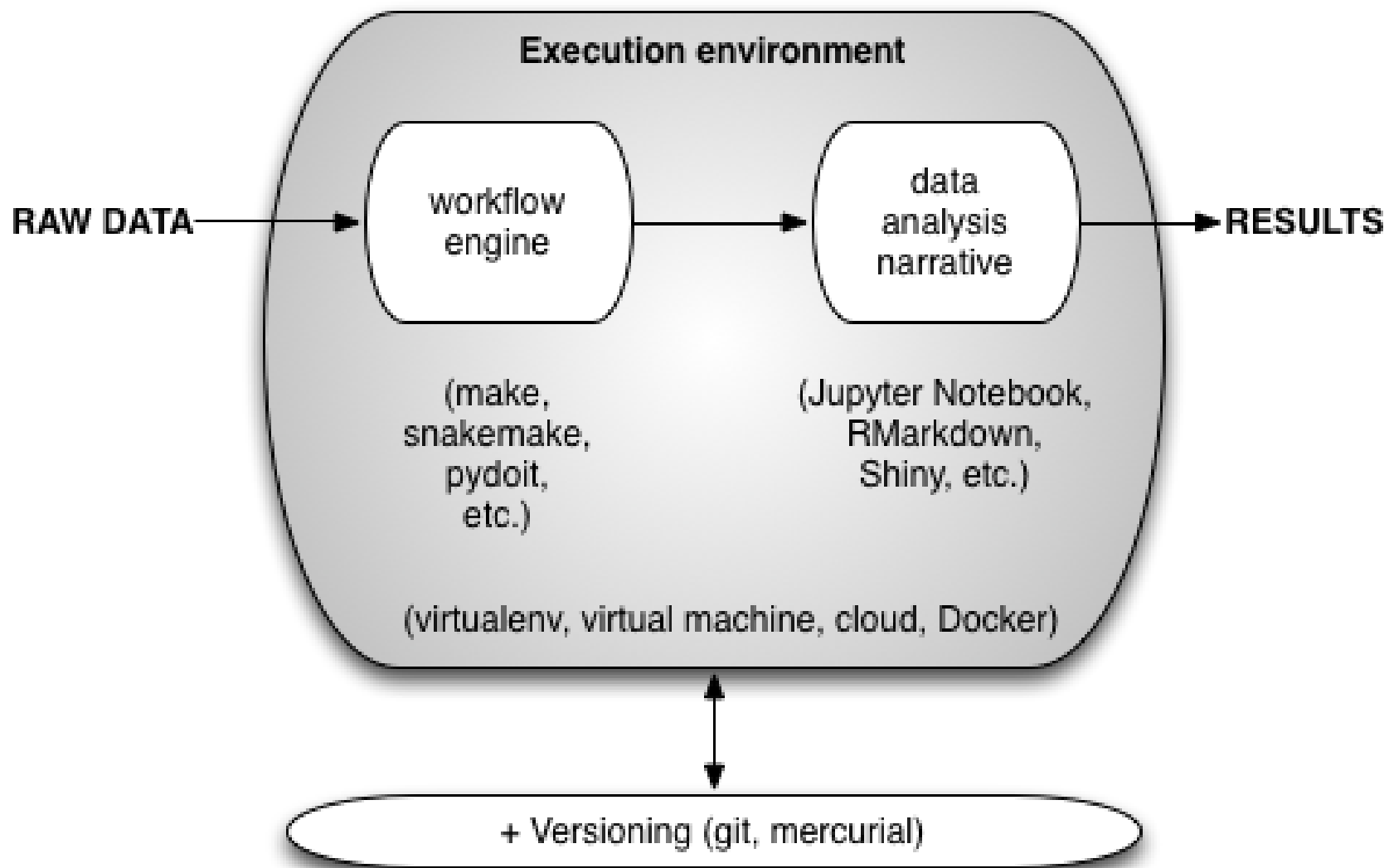
versioning

(make,  
snakemake,  
pydoit,  
etc.)

(Jupyter Notebook,  
RMarkdown,  
Shiny, etc.)

(virtualenv,  
conda,  
virtual machine,  
cloud,  
Docker,  
etc.)

(git,  
mercurial,  
etc.)



# OHMIGODTHISISALOTOFSTUFF: Advice

- Adopt what you can when you need it.
- Attack your BIG problems, and/or solve your easy-to-solve problems, as you need to.
- Invest periodically in training (leveling up happens incrementally)
- Organize your data; build a structure around it; refactor as you get annoyed by it. (Example!)

# Benefits

- Increased efficiency
- Surprisingly easy
- Rather convenient
- Improved correctness
- Easy methods
- Positions you well for non-professor jobs

# Problems!

- Where do you put the data? How do you bring the compute to the data?
- Long term reproducibility (platforms, libraries, and formats)



# Resources

- Software and Data Carpentry
- Bioinformatics Data Skills (book)
- Practical Computing for Biologists (book)