



# **A Practical Introduction to Reproducible Computational Workflows**

Make your code reproducible by  
anyone, anywhere

# Tools and Infrastructure



Computational Notebooks:  
Jupyter Notebook  
Jupyter Lab



Cloud environment to run  
computational notebooks  
(including RStudio)



Source package and  
environment  
management system



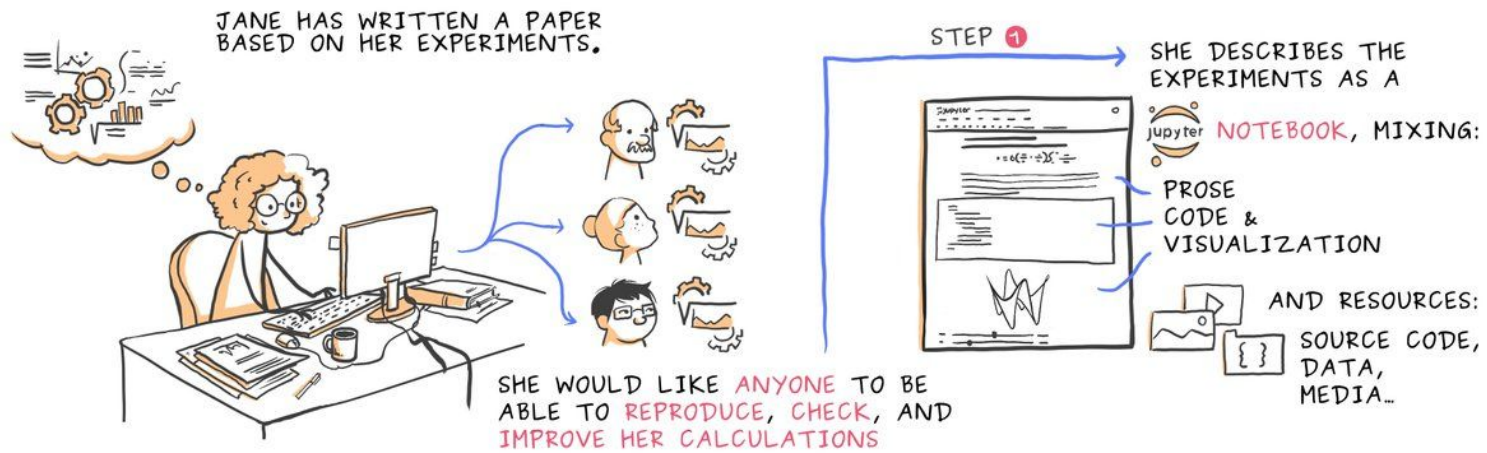
Source code  
repository

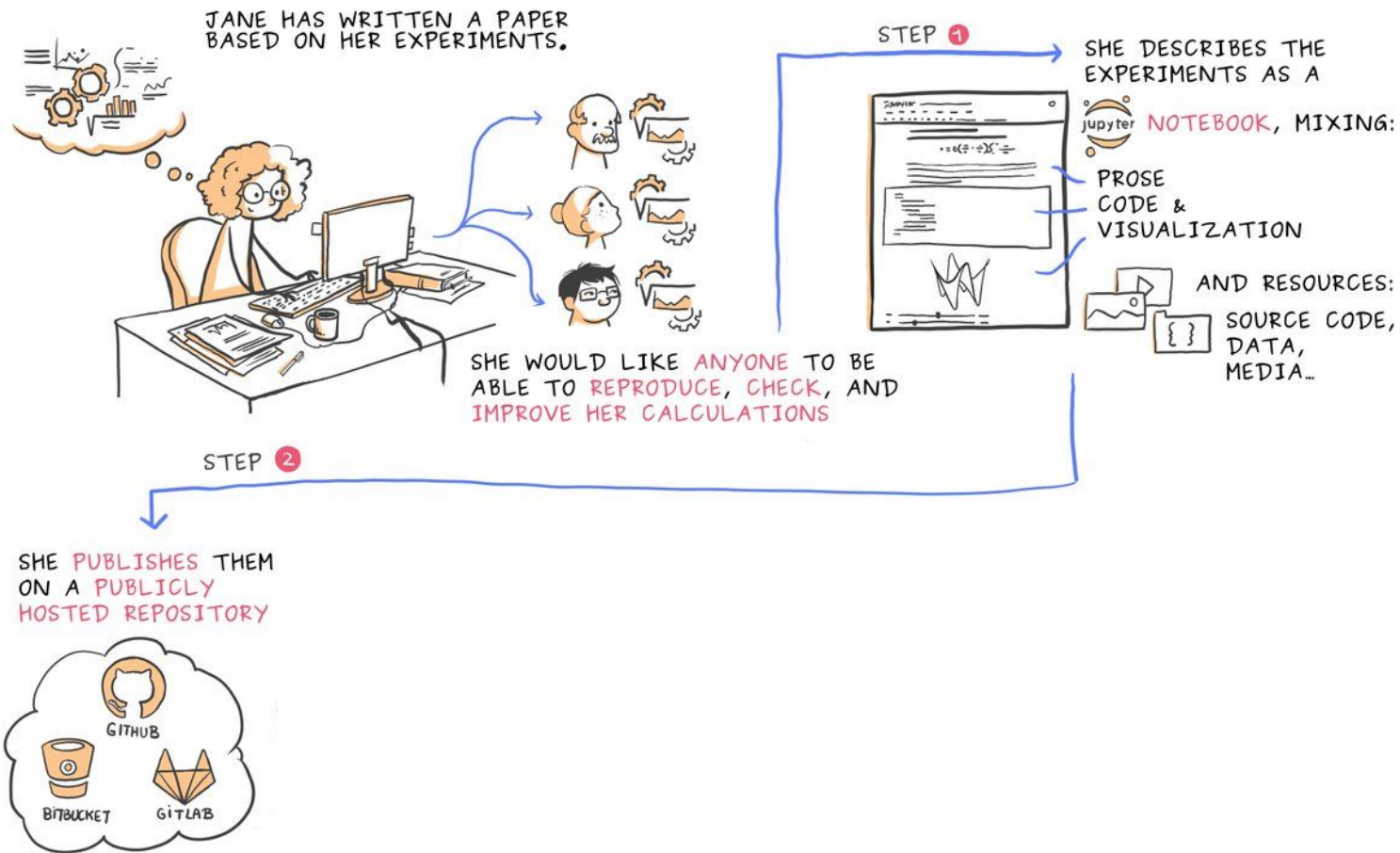
# The binder Project

A community that builds **free and open-source** tools  
for **reproducible, sharable scientific environments**  
that are **workflow- and platform-agnostic**.

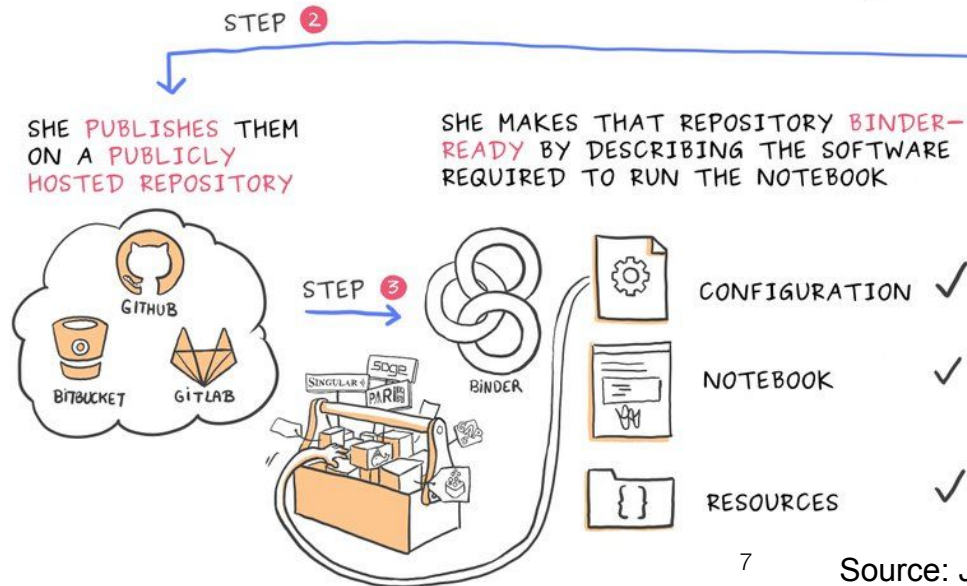













JANE HAS WRITTEN A PAPER  
BASED ON HER EXPERIMENTS.

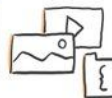


SHE WOULD LIKE **ANYONE** TO BE  
ABLE TO **REPRODUCE, CHECK, AND**  
**IMPROVE** HER CALCULATIONS

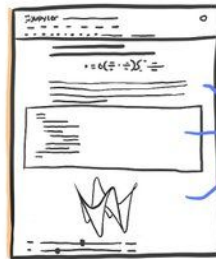
STEP 1

SHE DESCRIBES THE  
EXPERIMENTS AS A  
 **NOTEBOOK**, MIXING:

PROSE  
CODE &  
VISUALIZATION

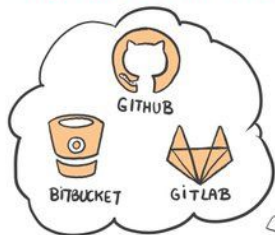


AND RESOURCES:  
SOURCE CODE,  
DATA,  
MEDIA...

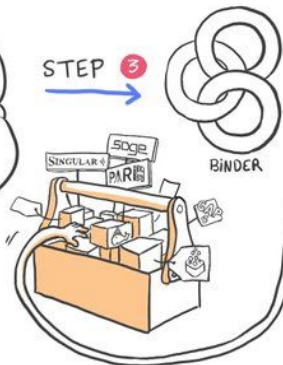


STEP 2

SHE **PUBLISHES** THEM  
ON A **PUBLICLY**  
**HOSTED** REPOSITORY



STEP 3

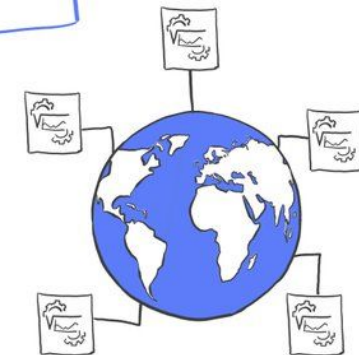


SHE MAKES THAT REPOSITORY **BINDER-READY** BY DESCRIBING THE SOFTWARE  
REQUIRED TO RUN THE NOTEBOOK

CONFIGURATION ✓

NOTEBOOK ✓

RESOURCES ✓



EVERYONE CAN NOW RUN AND  
REPRODUCE HER COMPUTATIONS

STEP 4






# How to configure a Repository for Binder?

“Binderize a repository”



Mybinder.org uses the files in **/binder** folder to create a conda environment.

The software dependencies are specified in **environment.yml**.

Branch: master ▾ [reproducible-computational-workflows](#) / **binder** /

 **pwrose** added postBuild file for py3Dmol

..

 <a href="#">environment.yml</a>	added binder and environment.yml
 <a href="#">postBuild</a>	added postBuild file for py3Dmol

environment.yml

```
name: ismb2019
channels:
  - conda-forge
  - rmg
dependencies:
  - python=3.7
  - numpy=1.16.4
  - pandas=0.24.2
  - matplotlib=3.1.0
  - seaborn=0.9.0
  - scikit-learn=0.21.2
  - ipywidgets=7.5.0
  - nodejs=11.14.0
  - jupyter=1.0.0
  - jupyterlab=1.0.2
  - nb_conda_kernels=2.2.2
  - py3dmol=0.8.0
```

Documentation: <https://mybinder.readthedocs.io>

# Jupyter Lab Plugins

- Some interactive packages require a Jupyter Lab plugin
  - These plugins are specified in the **postBuild** file, e.g., for [ipywidgets](#)

```
#!/bin/bash
jupyter labextension install @jupyter-widgets/jupyterlab-manager@1.0
```

Branch: master ▾

[reproducible-computational-workflows](#) / binder /



pwrose added postBuild file for py3Dmol

..



[environment.yml](#)

added binder and environment.yml



[postBuild](#)

added postBuild file for py3Dmol

# How to Launch Repo on <https://mybinder.org>?

<https://github.com/ISMB-ECCB-2019-Tutorial-AM4/reproducible-computational-workflows>

## Build and launch a repository

GitHub repository name or URL

<https://github.com/ISMB-ECCB-2019-Tutorial-AM4/reproducible-computational-workflows>

GitHub ▾

Git branch, tag, or commit

Git branch, tag, or commit



Path to a notebook file (optional)

Path to a notebook file (optional)

File ▾

launch

Copy the URL below and share your Binder with others:

<https://mybinder.org/v2/gh/ISMB-ECCB-2019-Tutorial-AM4/reproducible-computational-workflows/master>



Copy the text below, then paste into your README to show a binder badge:



# Make your code reproducible by anyone, anywhere

Share a link with your collaborators

Copy the URL below and share your Binder with others:

`https://mybinder.org/v2/gh/pwrose/reproducible-computational-workflows/master`



Add a Jupyter Notebook launch button to your repository README.md file

Copy the text below, then paste into your README to show a binder badge:



`\[\[Binder\]\(https://mybinder.org/badge\_logo.svg\)\]\(https://mybinder.org/v2/gh/pwrose/reproducible-co`



# Launch Link for a Specific Notebook

## Build and launch a repository

GitHub repository name or URL

GitHub ▼

Git branch, tag, or commit



Path to a notebook file (optional)

File ▼

launch

Copy the URL below and share your Binder with others:



<https://mybinder.org/v2/gh/pwrose/reproducible-computational-workflows/master?filepath=3-jupyter%2Fjupyter-test.ipynb>

# How to create a Launch Link for Jupyter Lab?

By default binder links to Jupyter Notebook

Copy the URL below and share your Binder with others:

```
https://mybinder.org/v2/gh/pwrose/reproducible-computational-workflows/master
```



To launch Jupyter Lab append **?urlpath=lab** to the launch link:

<https://mybinder.org/v2/gh/pwrose/reproducible-computational-workflows/master?urlpath=lab>

To launch a specific notebook in Jupyter Lab append

**?urlpath=lab/tree/path-to-notebook/notebook.ipynb** to the launch link:

<https://mybinder.org/v2/gh/pwrose/reproducible-computational-workflows/master?urlpath=lab/tree=3-jupyter/Jupyter-test.ipynb>



# Hands-on Session

## Binderize your Repository

<https://github.com/ISMB-ECCB-2019-Tutorial-AM4/reproducible-computational-workflows/5-binder/README.md>

# Other Options to Share Jupyter Notebooks

- CyVerse VICE (Visual Interactive Computing Environment)
  - <http://learning.cyverse.org/projects/vice/en/latest/index.html>
  - **Free**, account sign-up required
  - More CPUs, memory and persistent storage (default 100GB)
  - Run notebooks for up to 48 hours
  - Suitable for production and compute intensive tasks
  - Supports Jupyter Lab, RStudio, Shiny applications
  - Funded by the US National Science Foundation
- Demo
  - <https://github.com/sbl-sdsc/mmtf-genomics>
- **CyVerse presentation by Jason Williams, Wednesday, July 24, 2 pm**
  - <https://www.iscb.org/ismbeccb2019-program/special-sessions#sst02>

# Other Options to Share Jupyter Notebooks

- Commercial
  - Suited for ad hoc sharing of single Notebooks, not reproducible research
  - Google Collaboratory (uses Google Drive to share single Notebooks)
    - <https://colab.research.google.com>
    - Pros: more CPU, RAM and GPUs
    - Cons: limited options to customise execution environment
  - Microsoft Azure Notebooks
    - <https://notebooks.azure.com/>

# Resources

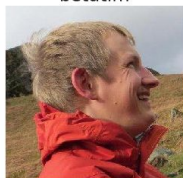
- Binder documentation
  - <https://mybinder.readthedocs.io>
- Binder example repository
  - <https://github.com/binder-examples>
- Binder 2.0, a Tech Guide
  - <https://blog.jupyter.org/binder-2-0-a-tech-guide-2017-fd40515a3a84>
- Binder 2.0 - Reproducible, interactive, sharable environments for science at scale, Project Jupyter, et al. (2018) Proc. of the 17th Python in Science Conf. (SCIPY 2018)
  - [http://conference.scipy.org/proceedings/scipy2018/pdfs/project\\_jupyter.pdf](http://conference.scipy.org/proceedings/scipy2018/pdfs/project_jupyter.pdf)

# The Jupyter/Binder Team

andrewosh



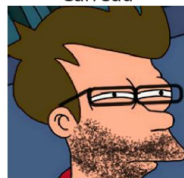
betatim



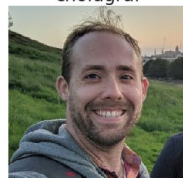
captainsafia



Carreau



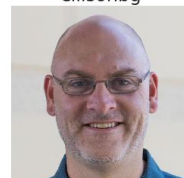
choldgraf



consideratio



ellisonbg



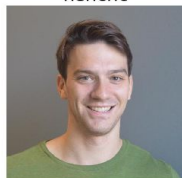
fperez



freeman-lab



henhc



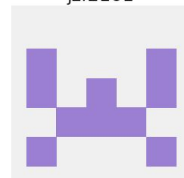
JamiesHQ



jhamrick



jzf2101



lheagy



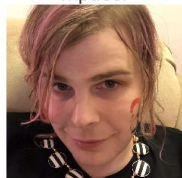
mbmilligan



minrk



mpacer



parente



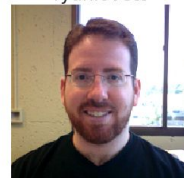
rgbkrk



Ruv7



ryanlovett



sgibson91



takluyver



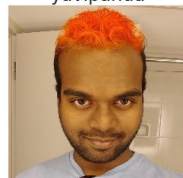
tgeorgeux



willingc



yuvipanda



Zsailer

