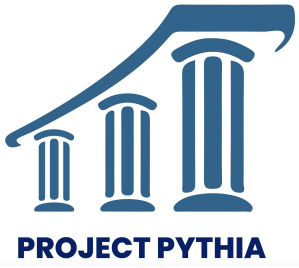



# Project Pythia's vision

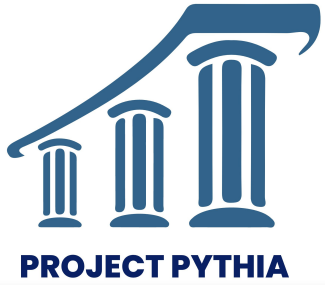
## Building an Inclusive Geoscience Community through Accessible, Reusable, and Reproducible Workflows

**Brian E. J. Rose**, John Clyne, Ryan May, James Munroe, Amelia Snyder, Orhan Eroglu, Kevin Tyle, Drew Camron, Max Grover, Julia Kent, Robert Ford





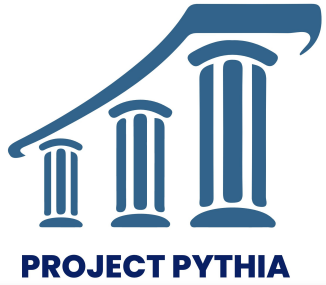

In ancient Greek mythology, the god Apollo is said to have slain the monstrous Python. Apollo's temple was later served by the Oracle of Delphi, who was known as the Pythia.




https://projectpythia.org/about.html

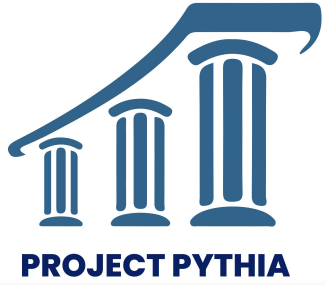
In ancient Greek mythology, the god Apollo is said to have slain the monstrous Python. Apollo's temple was later served by the Oracle of Delphi, who was known as the Pythia.

*Like the Oracle of old, Project Pythia will help you make sense of things that seem mysterious. We will not, however, slay your Python, but instead help you improve it – and help you to help others do the same!*




https://projectpythia.org/about.html

# Project Pythia: what / why?




A Community Learning Resource for Geoscientists

1. Geoscience community moving to open source software and cloud computing for analysis to better support open, reproducible science
2. Python ecosystem, cloud computing are complex and dynamic environments
3. Geoscientists are not computer scientists (but are increasingly reliant on a complex and ever-changing collection of computing technologies)
4. **Little training material exists that is focused specifically on needs of geoscientists**

5. ***Wide. Open. Science.*** requires new skills!

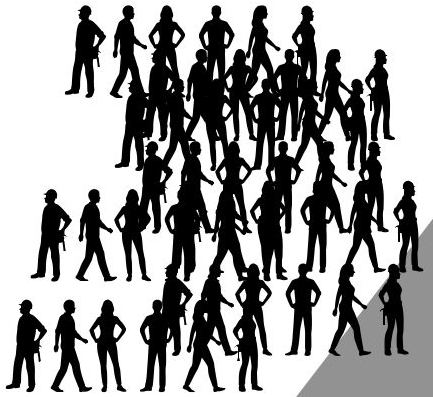
https://projectpythia.org

# Why?

Data!

PROJECT PYTHIA

*Most Geoscientists...*

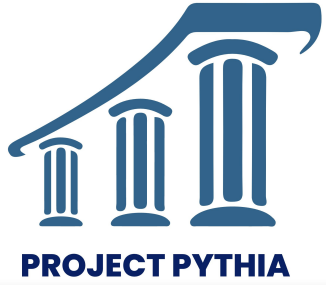


A mountain of languages, software packages, environments, and tools

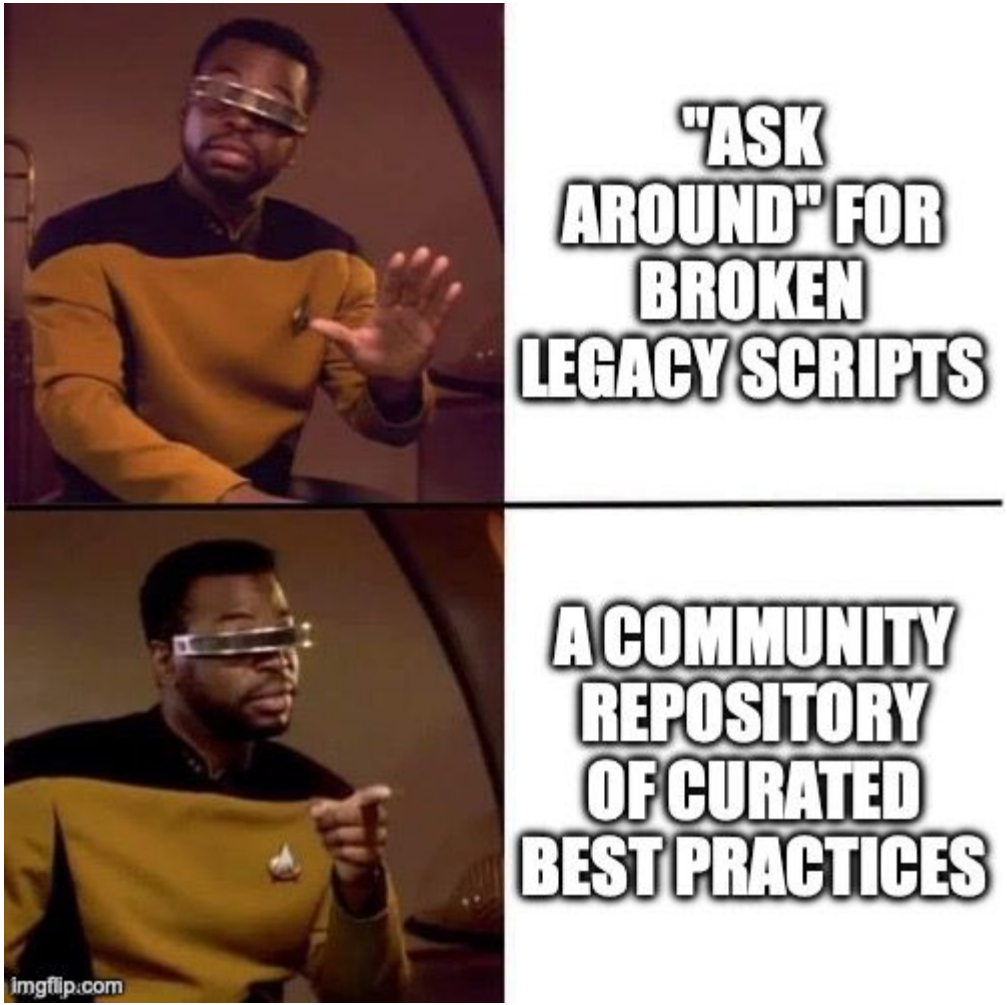
*Happy Pangeans*

*To **reduce barriers** to **tool adoption** and **participation**;*
*To **empower** more people to do **better**, more **reproducible data-heavy science**;*
*To **organize** the community around a **well-maintained** set of **learning resources***

# How do I…



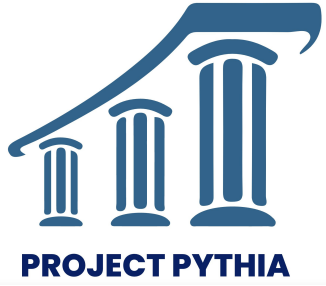

- *Get started* with data access, comprehension, analysis, visualization?
- Know what packages to use?
- *Scale up* from a laptop to a Big Data compute engine?
- Actually *implement a calculation* that is only conceptually described in the literature?
- *Connect with other people* struggling with the same datasets?
- Ensure my results are *reproducible* today, and might still be reproducible next year?




[me in grad school]

PROJECT PYTHIA

# Envisioning a better future




- Crowd-sourced best practices are curated and maintained in a findable, open, inclusive community archive
- Backed by cloud-based reproducibility infrastructure
- Code examples “just work” on real data
- Pathways to scale up from working examples to new science

We call these curated **collections of recipes** for transforming raw data into expressive results **“Cookbooks”**


[most notebooks found in the wild]

PROJECT PYTHIA

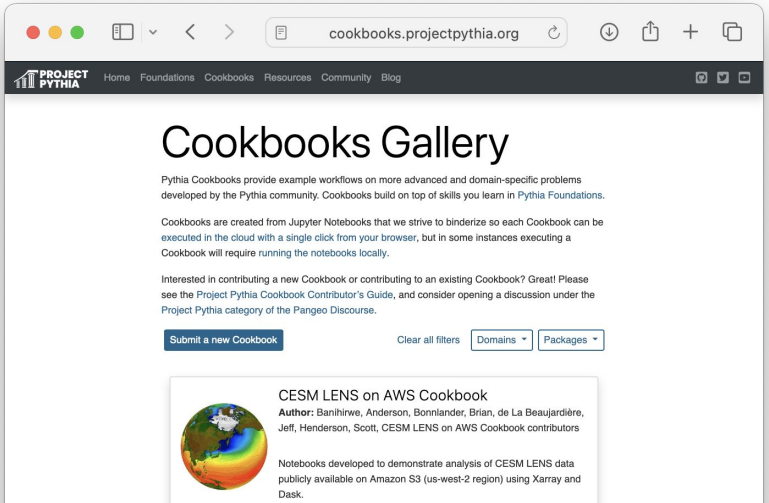
# A sustainable, inclusive Open Science community enabled by Cookbooks

*A vibrant clearinghouse of well-organized, accessible, and inter-linked Cookbooks backed by a performant, portable infrastructure and an enthusiastic social culture of open knowledge-sharing*

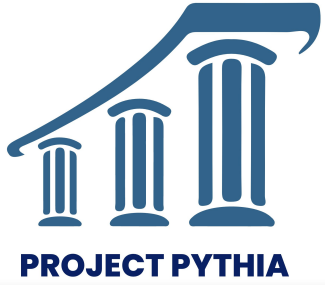
PROJECT PYTHIA

Democratizing access to data

Reducing knowledge barriers

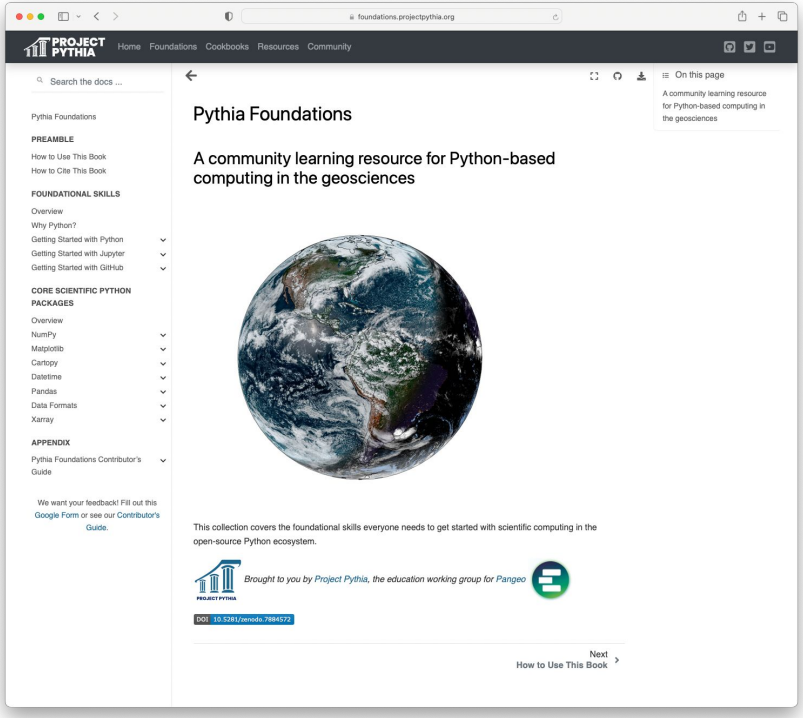


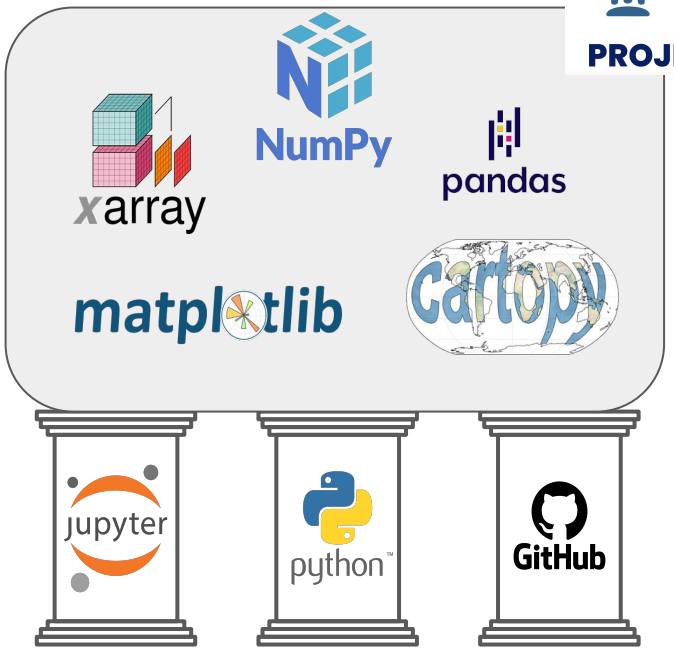

Nurturing an inclusive community




# OK, but what has Project Pythia actually done?

# Pythia Foundations: the on-ramp for new users, aiming for broad, diverse participation




https://foundations.projectpythia.org

***Binderized*** for one-click interactive learning

# Pythia Cookbooks


https://cookbooks.projectpythia.org

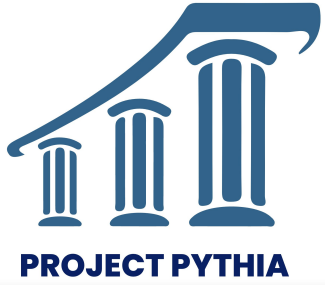
PROJECT PYTHIA

*Cookbooks are community-contributed collections of advanced or domain-specific tutorials and example workflows*

*Essential features of Pythia Cookbooks:*

- *Explicitly build upon Foundations*
- *Demonstrate **real workflows** on publicly available data*
- *Binderized for interactive learning*
- *Backed by **automated testing infrastructure** to ensure that the example code "just works" and stays relevant*

*Starting points for new geoscience analysis using the Python stack*

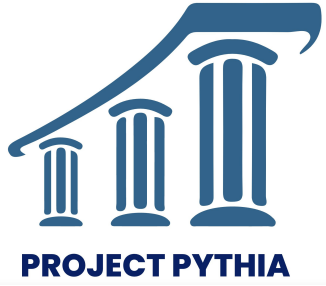



# Why Cookbooks?

## *What problems are Cookbooks trying to solve?*

**Jupyter Notebooks are an awesome way to share scientific workflows, but…**

# What problems are Cookbooks trying to solve?




**Jupyter Notebooks are awesome, but…**

- **Ambiguity**: Jupyter notebooks don't fully describe their own execution environment

jupyter {book}

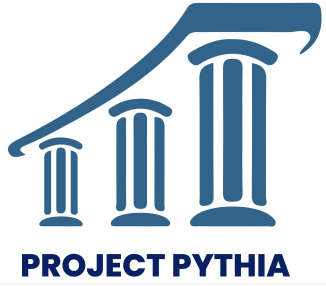
binder

A great tool for packaging Notebooks and conda environment descriptions into easy-to-navigate Web pages, with Binder links for execution

*Cookbooks are executable and reproducible*

# What problems are Cookbooks trying to solve?

**Jupyter Notebooks are awesome, but…**




- **Obsolescence**: most Notebooks found “in the wild” will not run and/or will not reproduce themselves


We need a CI service that can perform regular “health-checking” of notebook code!

*Cookbooks are versioned and maintained*

# What problems are Cookbooks trying to solve?

**Jupyter Notebooks are awesome, but…**

PROJECT PYTHIA

- **Collaboration and Attribution**: Notebooks don't play very well with GitHub pull requests




*We need to execute notebooks and generate + deploy a preview of the rendered book to facilitate review and merge cycles*

*Cookbooks are collaborative scholarly objects*

# What problems are Cookbooks trying to solve?

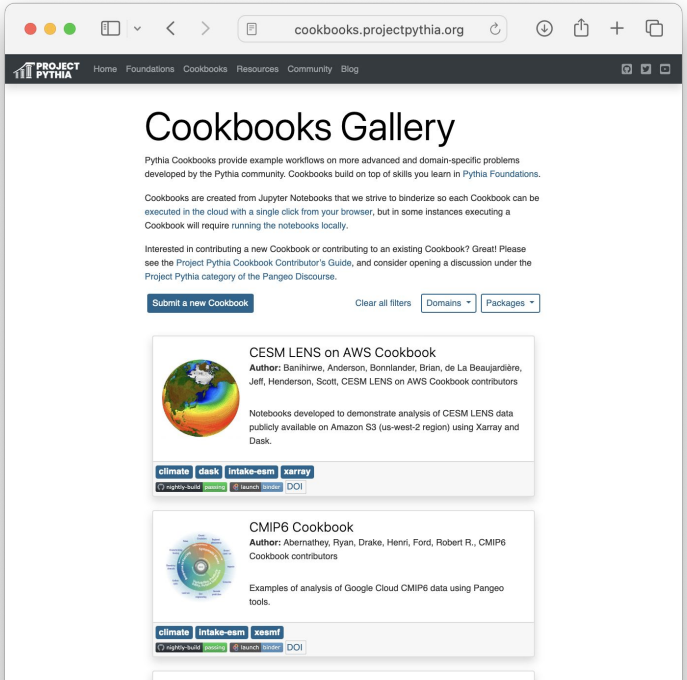
**Jupyter Notebooks are awesome, but…**

- **Findable and Accessible**: using Notebooks to share knowledge about scientific workflows requires an audience!

*We should have a community repository for sharing workflows that represent established best practices!*

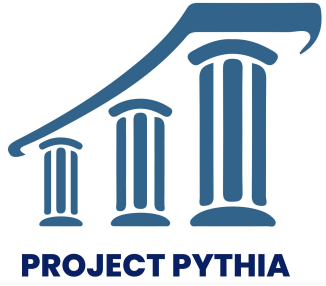
*And it should be organized and filterable*

*Cookbooks are open and community-owned*



# What problems are Cookbooks trying to solve?

**Jupyter Notebooks are awesome, but…**



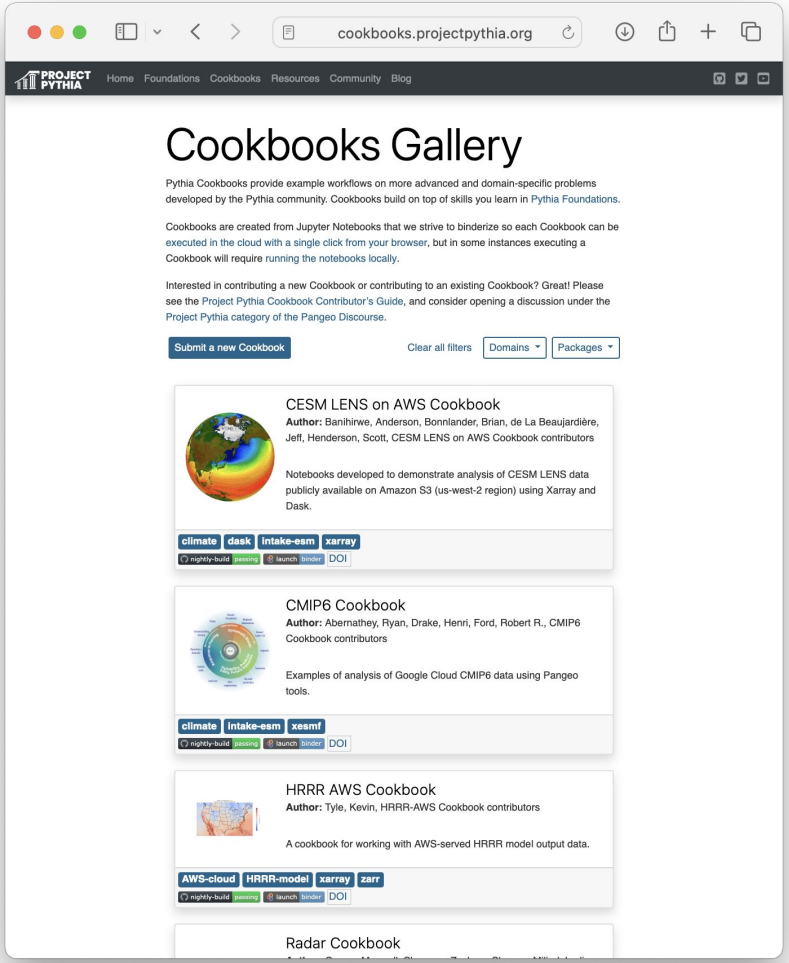

- **Scalability**: tutorials that run in a limited sandbox don't offer clearest paths to doing new science on real data


*We need to be able to route notebook execution to the appropriate compute resource for its content!*

*Cookbooks are portable – bring the compute to the data*

# Current state of the Cookbook collection

- **17 Cookbooks** comprising 95 individual chapters (notebooks)... and growing
- Books organized around science topic, datasets, or tool
- Climate modeling, remote sensing, radar meteorology, machine learning, parallel computing, visualization, …




https://cookbooks.projectpythia.org

# What's next?

**New 3-year NSF GEO OSE award**: Project Pythia and Pangeo: Building an inclusive geoscience community through accessible, reusable, and reproducible workflows

**New partners:** 2i2c, USGS

1. More cookbooks, more people in the kitchen!
   - More exemplar Cookbooks touching high-value datasets across the geosciences
   - Annual Cookbook **hackathon**; travel funds to recruit a broader group into open science
2. Scalable infrastructure for a growing Pythia user community
   - NSF-funded cyberinfrastructure (e.g. Jetstream2)
   - Commercial cloud
3. Cookbooks as scholarly objects
   - Formalization of review criteria
   - Elevation of reproducible and maintained notebooks as citable scholarly output
4. Community governance
   - Move from PI directed management to community-driven Steering Council model
   - Get involved!

https://projectpythia.org

# Project Pythia is a big tent

People who have contributed time, ideas, discussion, code, or content

Abigail Bodner
Alfonso Ladino
Alejandro Coca-Castro
Amelia Snyder
Andrew Huang
Anderson Banihirwe
Angus Hollands
Anissa Zacharias
Anne Fouilloux
Bane Sullivan
Brian Bonnlander
Brian Rose
Chia-Wei Hsu
Chris Cardinale
Christopher Dupuis
Crystal Worley
Curtis L. Walker
Daphne Quint
Dean Henze
Deepak Cherian
Demetris Roumis
Drew Camron
Elena Romashkova
Emilio Mayorga
Eric Nienhouse
Erin Rhoades
Heather Cracker
James Morley
James Munroe
Jasmine Sandhu
John Clyne
Jonathan Kingslake
Julia Kent
Julia Signell
Kevin Paul
Kevin Tyle
Lily Kailyn
Max Grover
Max Jones
Michael Barletta
Michaela Sizemore
Milind Sharma
Naomi Goldenson
Negin Sobhani
Nicole Corbin
Nihanth Wagmi Cherukuru
Orhan Eroglu
Phil Austin
Raphael Hagen
Rich Signell
Rowan Cockett
Ryan Abernathey
Ryan May
Pritam Das
Seth McGinnis
Scott Collis
Scott Pearse
Thomas Martin
Tom Nicholas
Yuta Norden
Zachary Sherman

# Join us at the Pythia Cook-off 2024

June 11-14 2024 @ NCAR and online

***A Cookbook-flavored hack week***

Bring your science, we'll help you turn it into brand new Cookbooks!

Help maintain and extend existing content

Develop your scientific Python development and educational skills

Learn Pythia's GitHub + Jupyterbook infrastructure

Connect with the Pythia and Pangeo community

**You are welcome!**

# June 11-14

In-person @ NCAR Mesa Lab
Hybrid-virtual via Zoom
Boulder, Colorado, US

projectpythia.org