

The Project Pythia Cookbook Initiative

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



Project Pythia: what / why?

A Community Learning and Knowledge-Sharing
Resource for the Geosciences

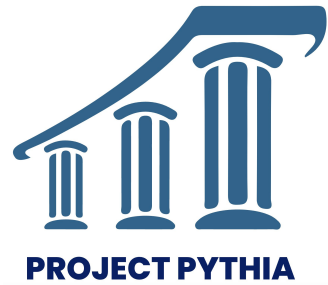


In ancient Greek mythology, the god **Apollo** is said to have slain the monstrous **Python**. Apollo's temple was later served by the **Oracle** of **Delphi**, who was known as the **Pythia**.

Project Pythia exists to help geoscientists do their science more effectively, centered on the scientific Python / Pangeo ecosystem of tools

<https://projectpythia.org>

How do I...



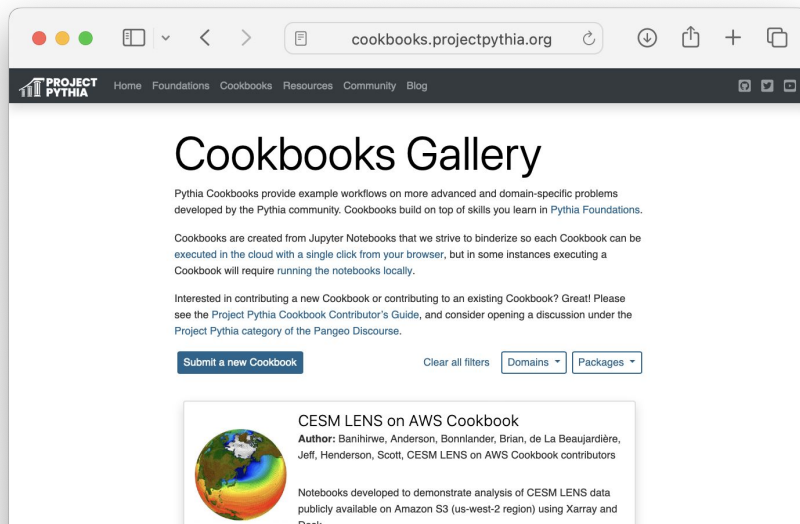
- *Get started* with data access, comprehension, analysis, visualization?
- Know what packages to use?
- *Scale up* from a laptop to a Big Data compute engine?
- *Actually implement a calculation* that is only conceptually described in the literature?
- *Connect with other people* struggling with the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

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**”



*Reducing
knowledge
barriers*

*Nurturing an
inclusive
community*

*Democratizing
access to data*

Current state of the Pythia Cookbook collection

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

WE NEED MORE OCEANOGRAPHY!

<https://cookbooks.projectpythia.org>



The screenshot shows the Project Pythia Cookbooks Gallery website. The page features a navigation bar with links for Home, Foundations, Cookbooks, Resources, Community, and Blog. The main heading is "Cookbooks Gallery". Below the heading, there is a brief introduction to Pythia Cookbooks and a section for submitting new cookbooks. The gallery displays three featured cookbooks:

- CESM LENS on AWS Cookbook**: Authored by Banihirwe, Anderson, Bonniander, Brian, de La Beaujardière, Jeff, Henderson, Scott. Description: Notebooks developed to demonstrate analysis of CESM LENS data publicly available on Amazon S3 (us-west-2 region) using Xarray and Dask. Tags: climate, dask, Intake-esm, xarray.
- CMIP6 Cookbook**: Authored by Abernathy, Ryan, Drake, Henri, Ford, Robert R. Description: Examples of analysis of Google Cloud CMIP6 data using Pangeo tools. Tags: climate, Intake-esm, xesmf.
- HRRR AWS Cookbook**: Authored by Tyle, Kevin. Description: A cookbook for working with AWS-served HRRR model output data. Tags: AWS-cloud, HRRR-model, xarray, zarr.

At the bottom, a "Radar Cookbook" is partially visible.



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



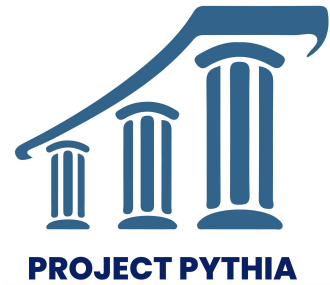
GitHub Actions

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...

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



GitHub Actions

GitHub Pages

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

zenodo

Cookbooks are collaborative scholarly objects

What problems are Cookbooks trying to solve?

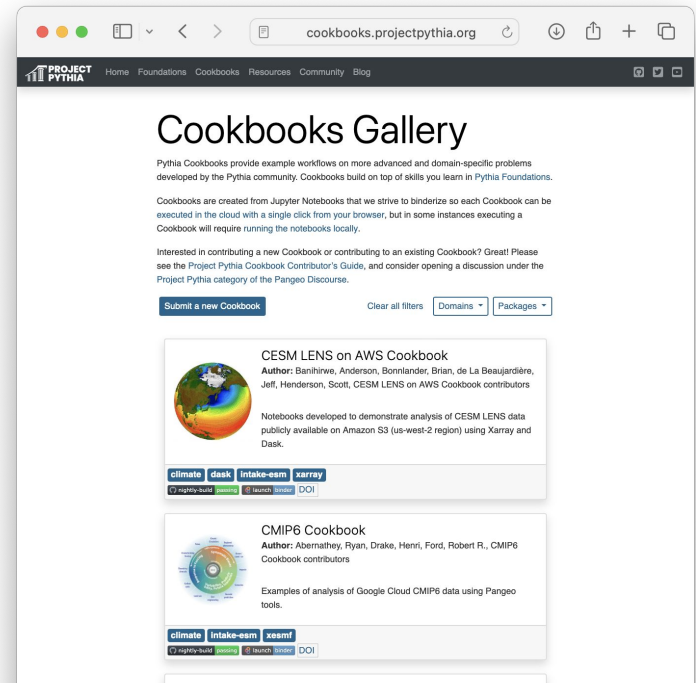
Jupyter Notebooks are awesome, but...

- **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

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>

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!

