Toward Scalable, Reproducible, and Open Ocean Acoustic Research

Valentina Staneva

Amanda Tan *Wu-Jung Lee* *Divya Panicker*

Valentina Stane

Valentina Staneva, University of Washington, USAva, University of Washington, USA

University of Washington

ASA Meeting, Victoria, BC, Nov 6, 2018

UNIVERSITY *of* WASHINGTON
eScience Institute

GORDON AND BETTY MOORE FOUNDATION

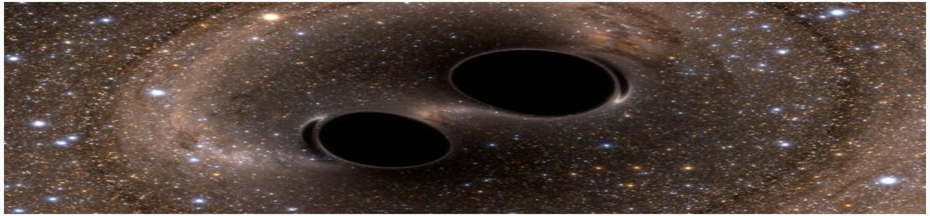
ALFRED P. SLOAN FOUNDATION

Streaming Sounds

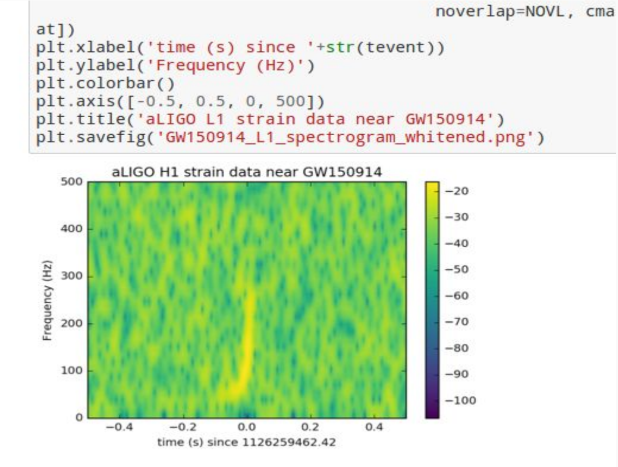
Ocean Observatories Initiative:

- 1 Hydrophone - up to 3TB per year
- 6 Hydrophones within network
- Many more ocean observatories:
  sonar, video,
  physical and chemical variables

Ligo Experiment

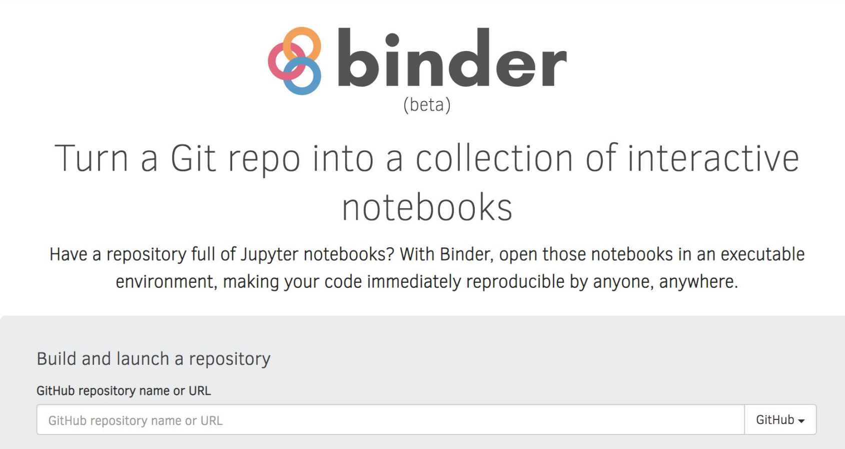


Jupyter Notebooks analyzing the data:



```
                                            noverlap=NOVL, cma
at])
plt.xlabel('time (s) since '+str(tevent))
plt.ylabel('Frequency (Hz)')
plt.colorbar()
plt.axis([-0.5, 0.5, 0, 500])
plt.title('aLIGO L1 strain data near GW150914')
plt.savefig('GW150914_L1_spectrogram_whitened.png')
```

[Ligo Tutorial](#) [Binder Notebook](#)



binder

(beta)

Turn a Git repo into a collection of interactive notebooks

Have a repository full of Jupyter notebooks? With Binder, open those notebooks in an executable environment, making your code immediately reproducible by anyone, anywhere.

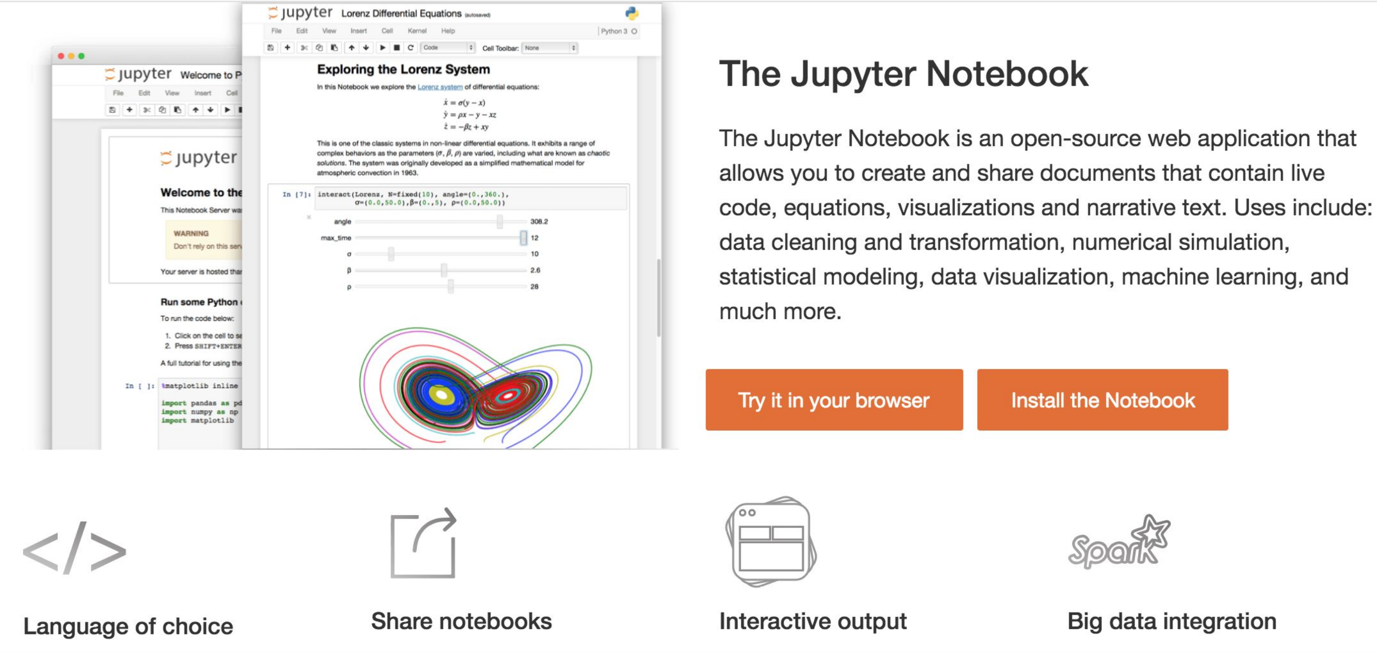
Build and launch a repository

GitHub repository name or URL

[mybinder.org](#)

Jupyter Notebooks

Project Jupyter


jupyter Welcome to P
File Edit View Insert Cell
jupyter
Welcome to the
This Notebook Server wa
WARNING
Don't rely on this serv
Your server is hosted than
Run some Python
To run the code below:
1. Click on the cell to se
2. Press SHIFT+ENTER
A full tutorial for using the
jupyter Lorenz Differential Equations (autosaved)
File Edit View Insert Cell Kernel Help
Python 3
Code
Cell Toolbar: None

Exploring the Lorenz System

In this Notebook we explore the Lorenz system of differential equations:

$$\dot{x} = \sigma(y - x)$$
$$\dot{y} = \rho x - y - xz$$
$$\dot{z} = -\beta z + xy$$

This is one of the classic systems in non-linear differential equations. It exhibits a range of complex behaviors as the parameters (σ, β, ρ) are varied, including what are known as *chaotic solutions*. The system was originally developed as a simplified mathematical model for atmospheric convection in 1963.

```
In [7]: interact(Lorenz, N=fixed(10), angle=(0.,360.),
                 σ=(0.0,50.0),β=(0.,5), ρ=(0.0,50.0))
```

angle 308.2
max_time 12
σ 10
β 2.6
ρ 28

```
In [ ]: %matplotlib inline

        import pandas as pd
        import numpy as np
        import matplotlib
```

The Jupyter Notebook

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

Language of choice

Share notebooks

Interactive output

Big data integration

Over 40 languages Supported

Many Notebook Hosting Platforms

Web Based:
Supports sound, video, widgets, visualizations, maps

Getaway to cloud computing

Literate Programming

Combining documentation and code in a single program.

“Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to human beings what we want a computer to do.”

- WEB (1981) - Latex + Pascal
- Mathematica Notebooks
- Reporting: Knitr + RPubs
- Notebooks
  - Jupyter, R Notebooks, Zeppelin, Sage, Beaker, ...
- Notebook Environments:
  - Binder, NBviewer, CoCalc, Colaboratory, Kaggle, ...

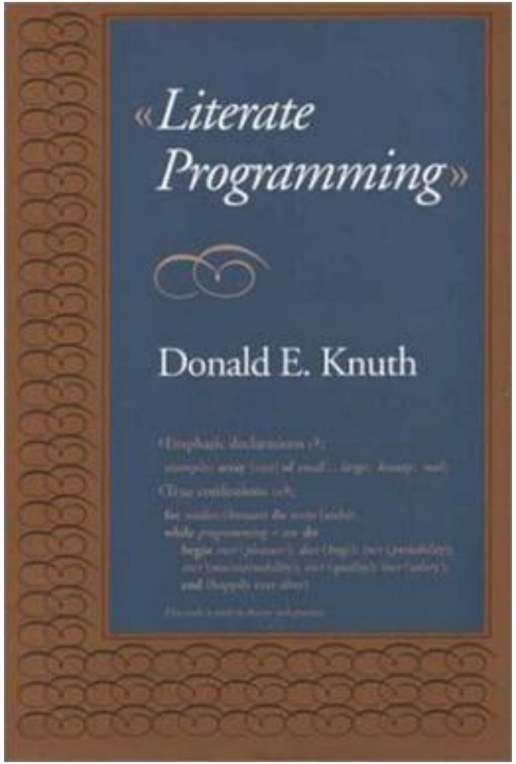


Image by Wikipedia

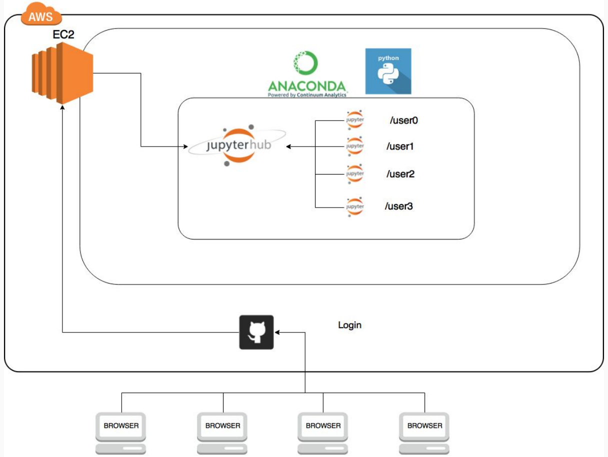
Scalable Education with JupyterHub

Hackweeks at University of Washington:

- Cabled Array Hackweek ~ 20 participants
- Oceanhackweek ~ 50 participants

Oceanhackweek 2018 Participants

- All tutorials hosted on JupyterHub on Amazon Cloud.
- Each user gets access to a Jupyter Notebook.
- Environment with all dependencies already installed.
- Instructors submit a conda environment files from which one docker image is built.



Zero to Jupyter Hub Tutorial

Image Source

Scalable Computing: GPU power

Azure Notebooks (Microsoft)

- 4GB RAM
- 1GB disk space
- Great Integration with Github
- R and Python

Cons: limited resources

Dask Tutorial Example

Colaboratory Notebooks (Google)

- 13GB RAM
- 33 GB disk
- GPU support
- Notebooks and data on Google Drive
- Integration with Github
- Simultaneous Editing
- Python only so far

Cons: not real filesystem

Kaggle Kernels

- 16GB RAM
- 5GB disk
- GPU support
- Upload/Edit/Download Notebooks
- Kaggle Datasets: public and private(20GB)
- Version Control Support
- R and Python

Cons: no Github integration

Deep Learning Example
CPU run: 1h 3min
GPU run: 3min

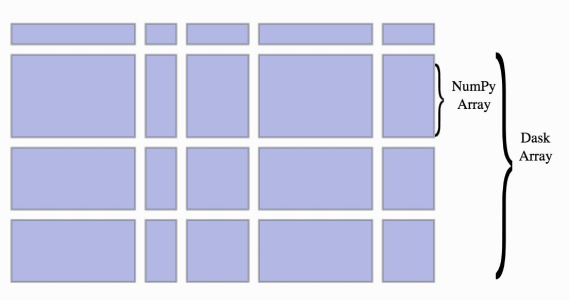
Scalable Computing: managing RAM

Chunked Data!

Data Formats: HDF5, netCDF, zarr, tiled tiff, ...

Libraries: h5py, dask, xarray, ...

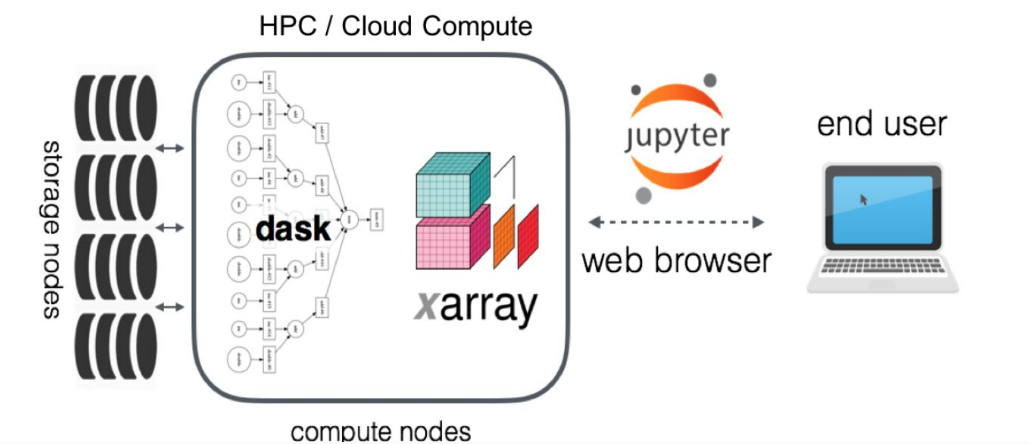
Local out-of-core computations + distributed computing.



[Pangeo Big Data Platform](#)

PANGEO

[Pangeo Physical Oceanography Lessons](#)



Coding Best Practices

- Using Version Control (Software Carpentry Lessons)
- Picking a license (https://choosealicense.com/)
- Project Organization and Packaging (Cookiecutter, Shablona)
- Virtualization (Conda, Docker, Vagrant, VirtualBox, VMWare, Cloud Images)
- Testing
  - Locally: Python - nose, pytest; R - testthat
  - Remotely: Travis, CircleCI, AppVeyor
- Documentation: Sphinx for Python, R Vignettes

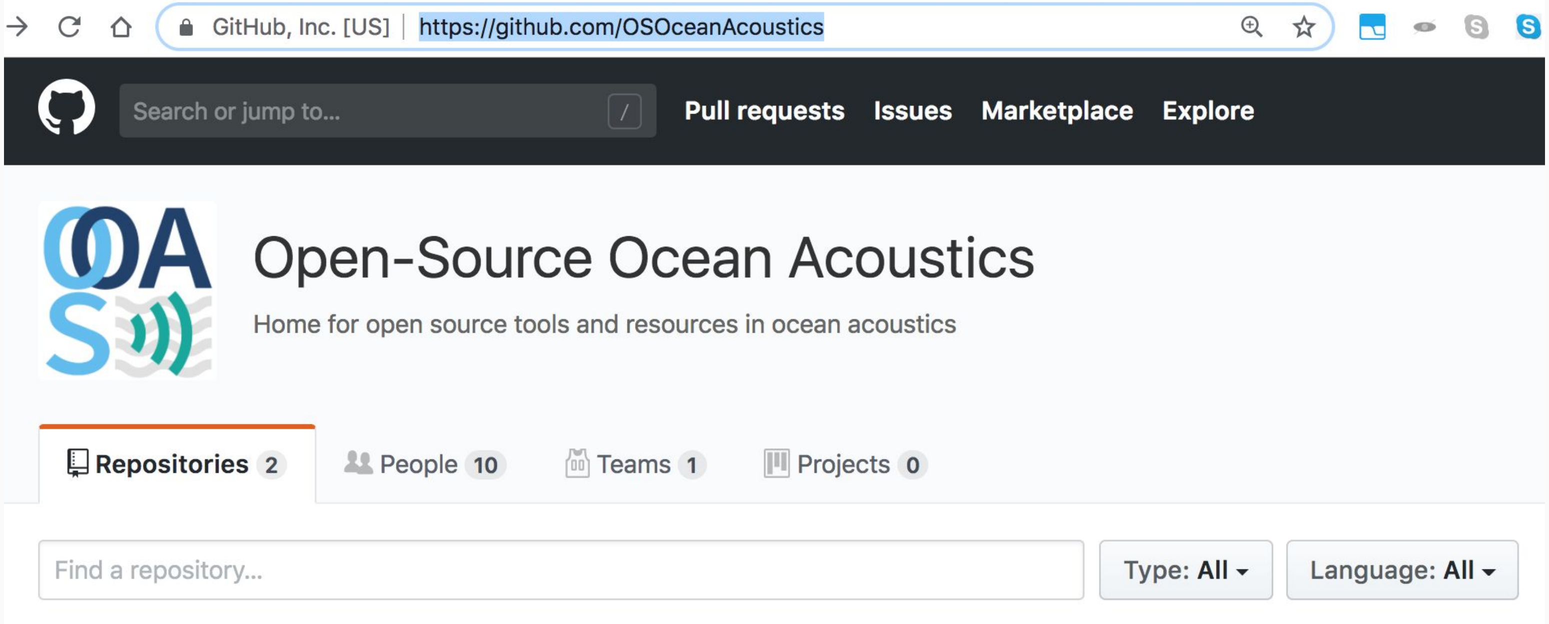
Learn by example!

Data Repositories

zenodo	figshare	DRYAD
Up to 50GB free Not-for-profit - EU funded	100GB free per manuscript Institutional plans For-profit	Publishing Fee - $120 Excess fees after 20GB Associated with articles Not-for-profit

- Cloud Storage: free to upload, fees to download
- Datasets receive Digital Object Identifier (DOI)
- Nature Journal Scientific Data: https://www.nature.com/sdata/

Join the Community!



GitHub, Inc. [US] | https://github.com/OSOceanAcoustics

Open-Source Ocean Acoustics

Home for open source tools and resources in ocean acoustics

https://github.com/OSOceanAcoustics

Thank You!