*Lessons from the field*
*Developing the Tutorial:*

# TIME SERIES OF **SATELLITE** DATA USING **PYTHON** IN THE **CLOUD**

FARALLON INSTITUTE


**Marisol García-Reyes**

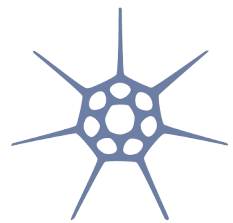
Farallon Institute \ marisolgr@gmail.com


https://bssw.io

better scientific software

# ROADMAP

- Motivation
- The tutorial
- Lessons learned



FARALLON INSTITUTE

# OPEN SCIENCE

- Sharing Data, Methods & Code
- Ensure accessibility to people that does not have it
  - Limited resources or
  - Limited expertise (not in the same field)
- Advancement & innovation by bringing diversity of ideas

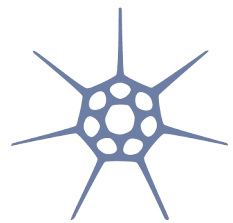
FARALLON INSTITUTE

# MAKE SATELLITE DATA & CLOUD COMPUTING MORE ACCESSIBLE TO NON-EXPERTS

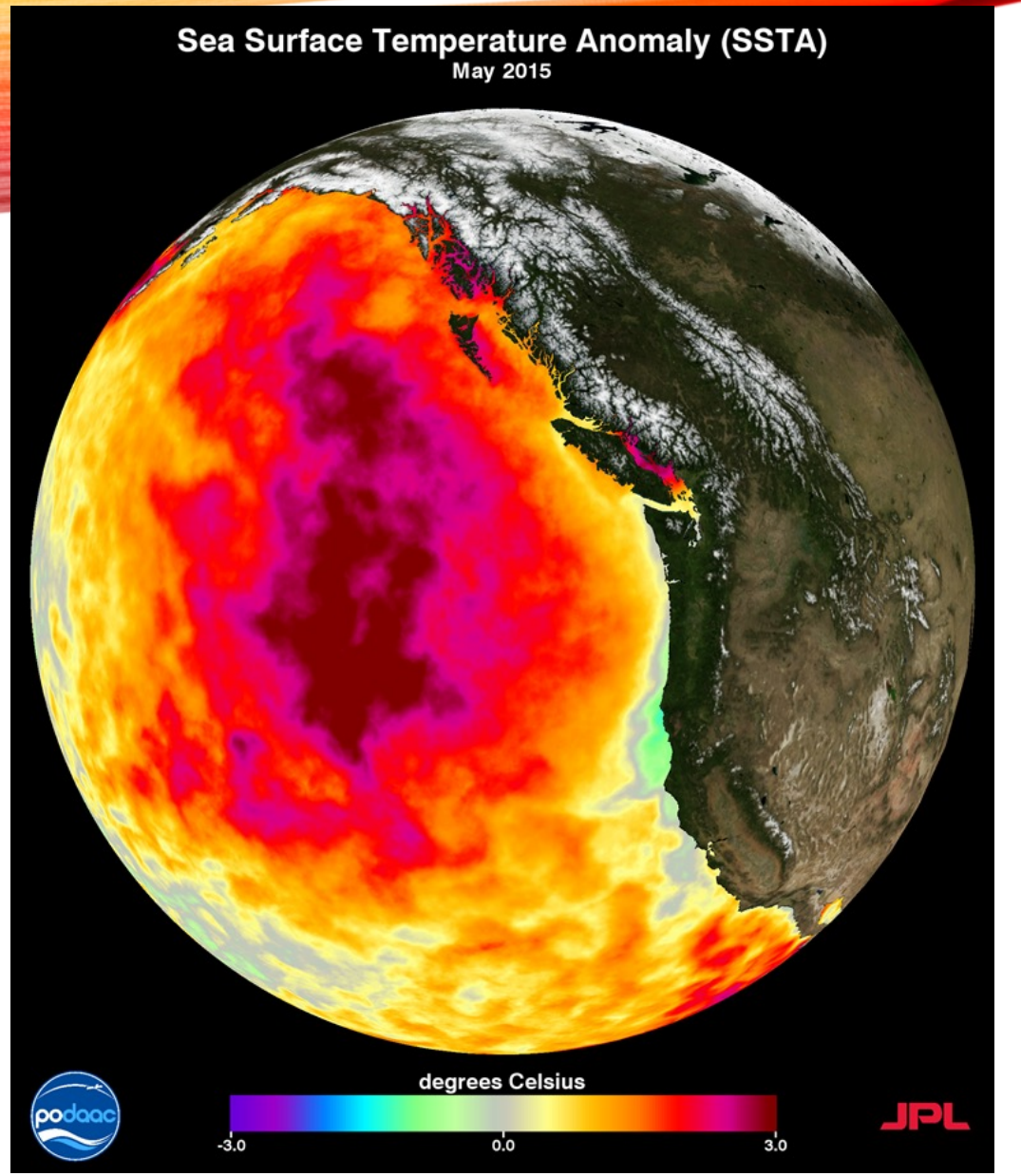
One time series at a time

# ROADMAP

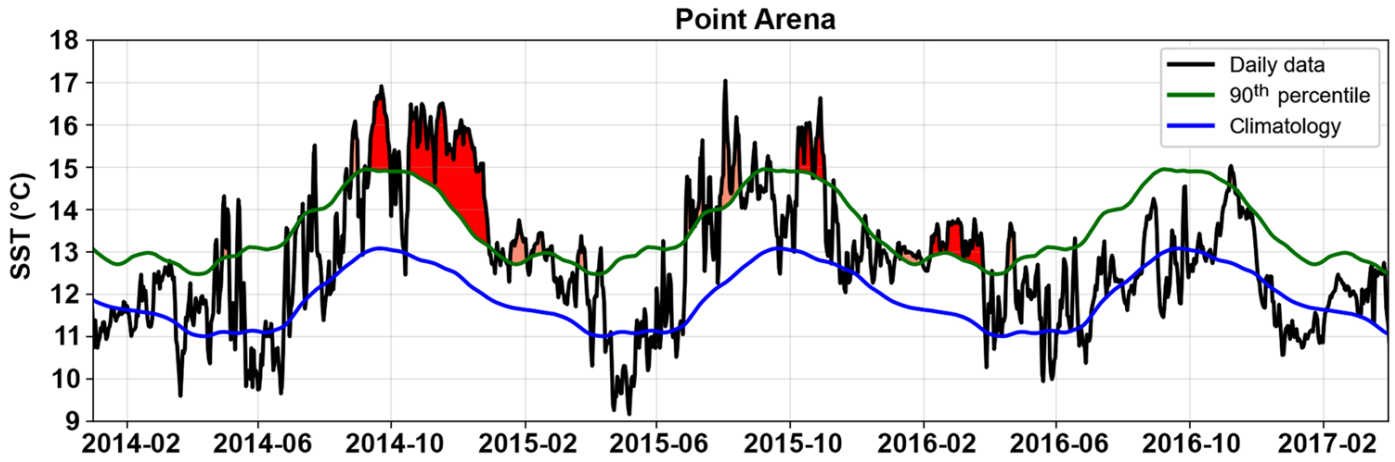
- Scientific Motivation
- The tutorial itself
- Lessons learned



FARALLON INSTITUTE




Gentemman et al. 2017



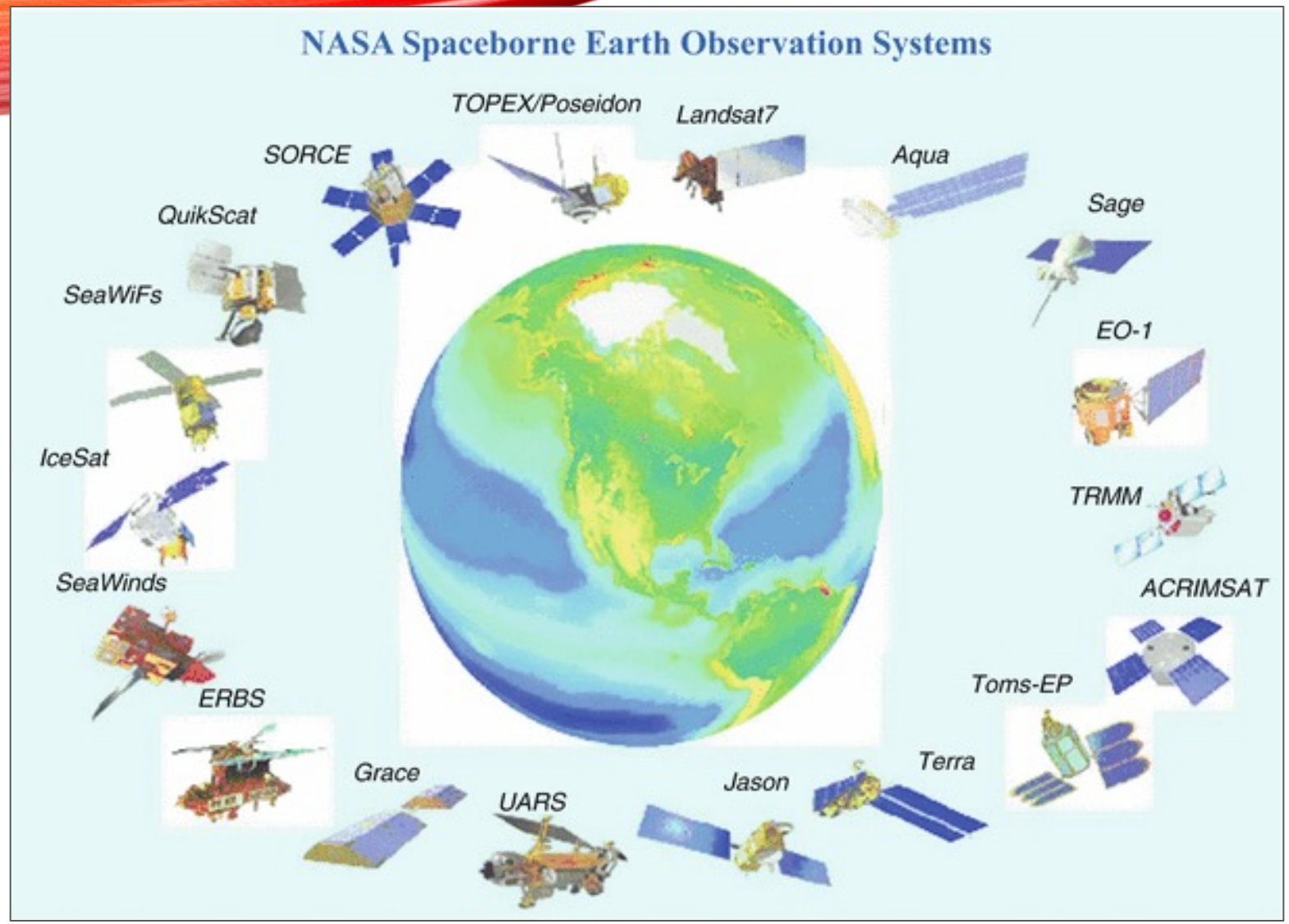

Sanford et al. 2019

Many web-based user-friendly interfaces to

**acquire & plot images & short time series**

of satellite & satellite-based data.

FARALLON INSTITUTE




Salomonson, 2015

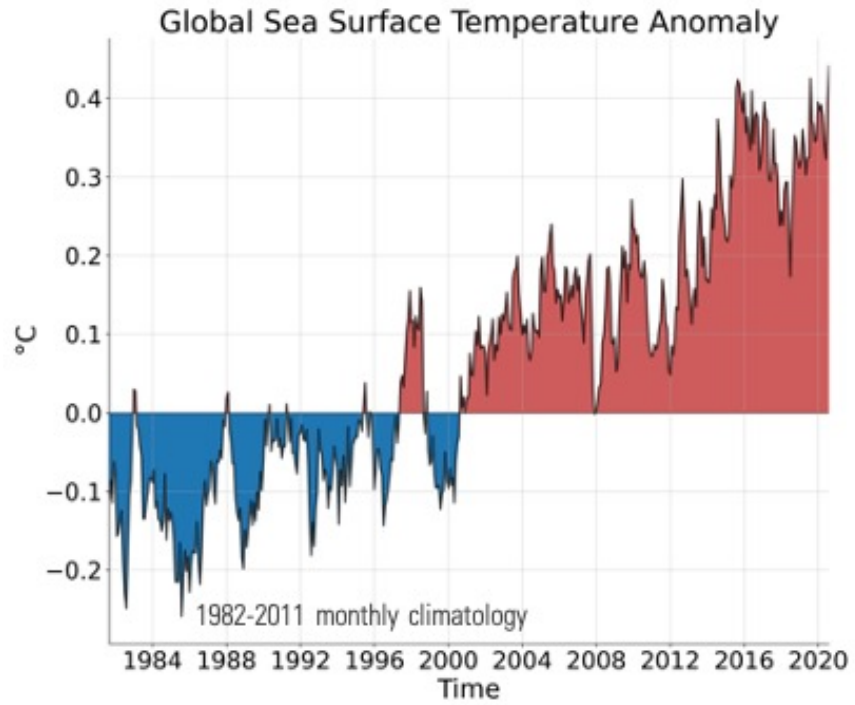
Satellites around for
>4 decades now

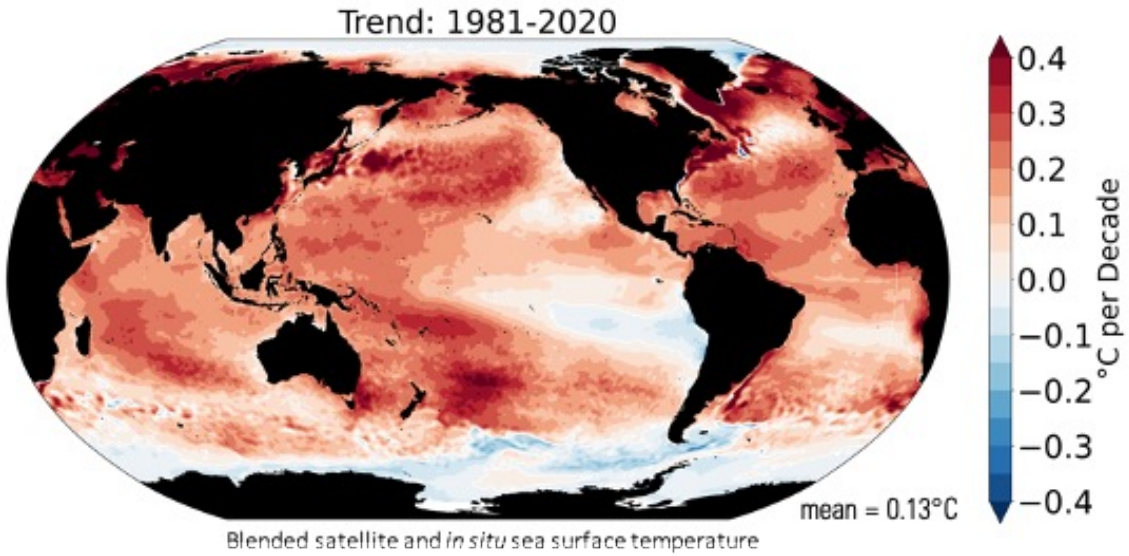
**Great for Climate Studies!**



FARALLON INSTITUTE

# CLIMATE RESEARCH



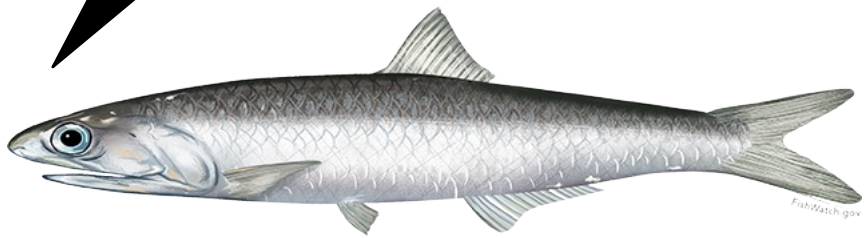




https://www.hillaryscannell.com/research/



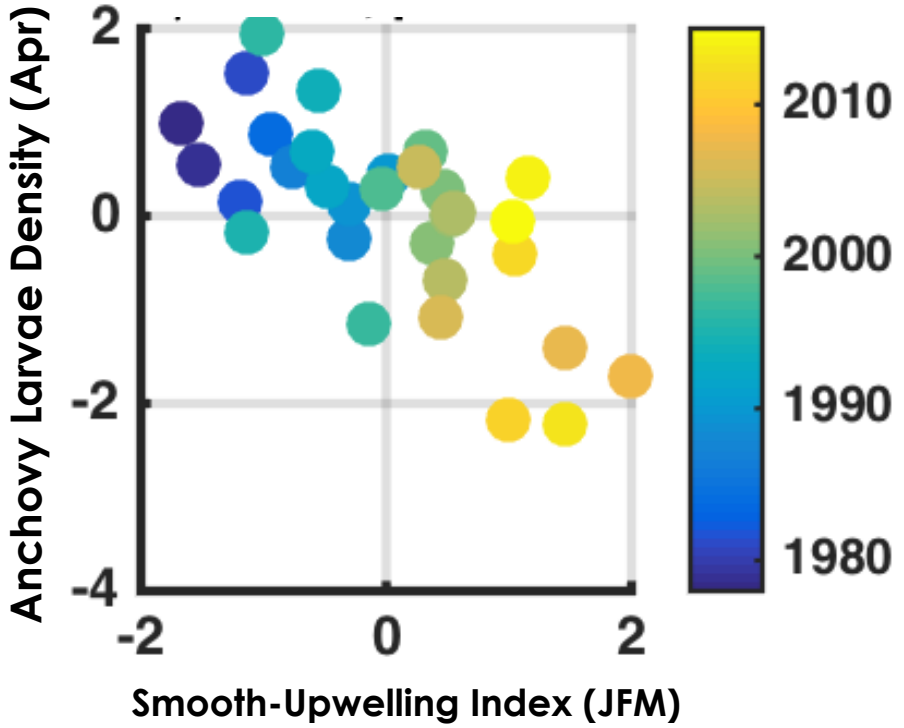
FARALLON INSTITUTE

I'm an anchovy



# ECOSYSTEMS RESEARCH
## (OR ANY OTHER FIELD)

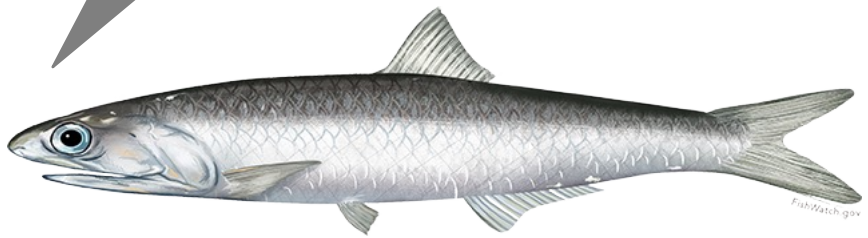
Biological indicator



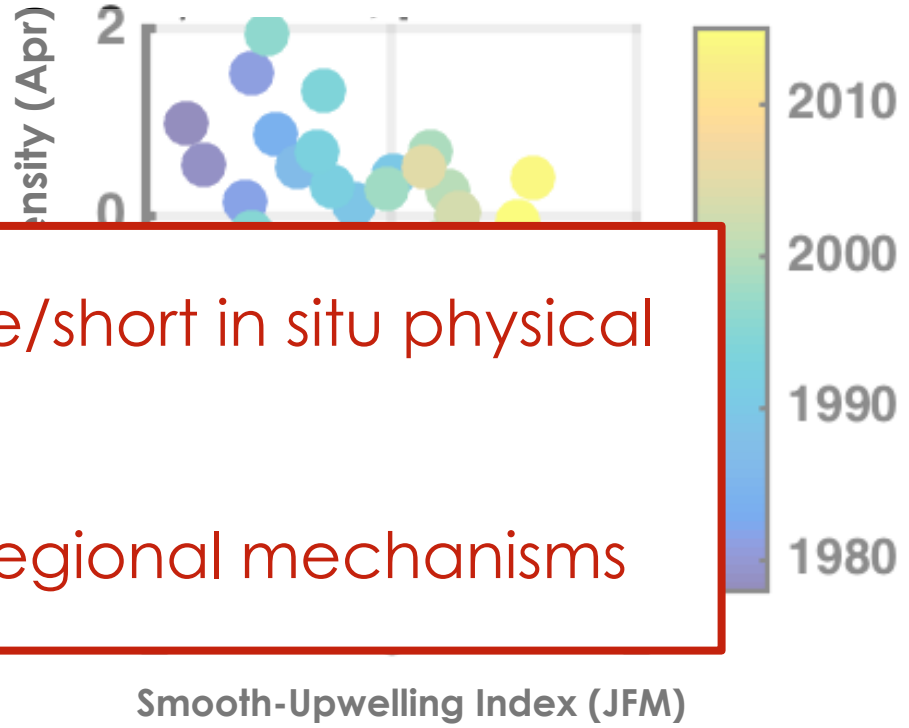

Climate indicator

FARALLON INSTITUTE

I'm an anchovy

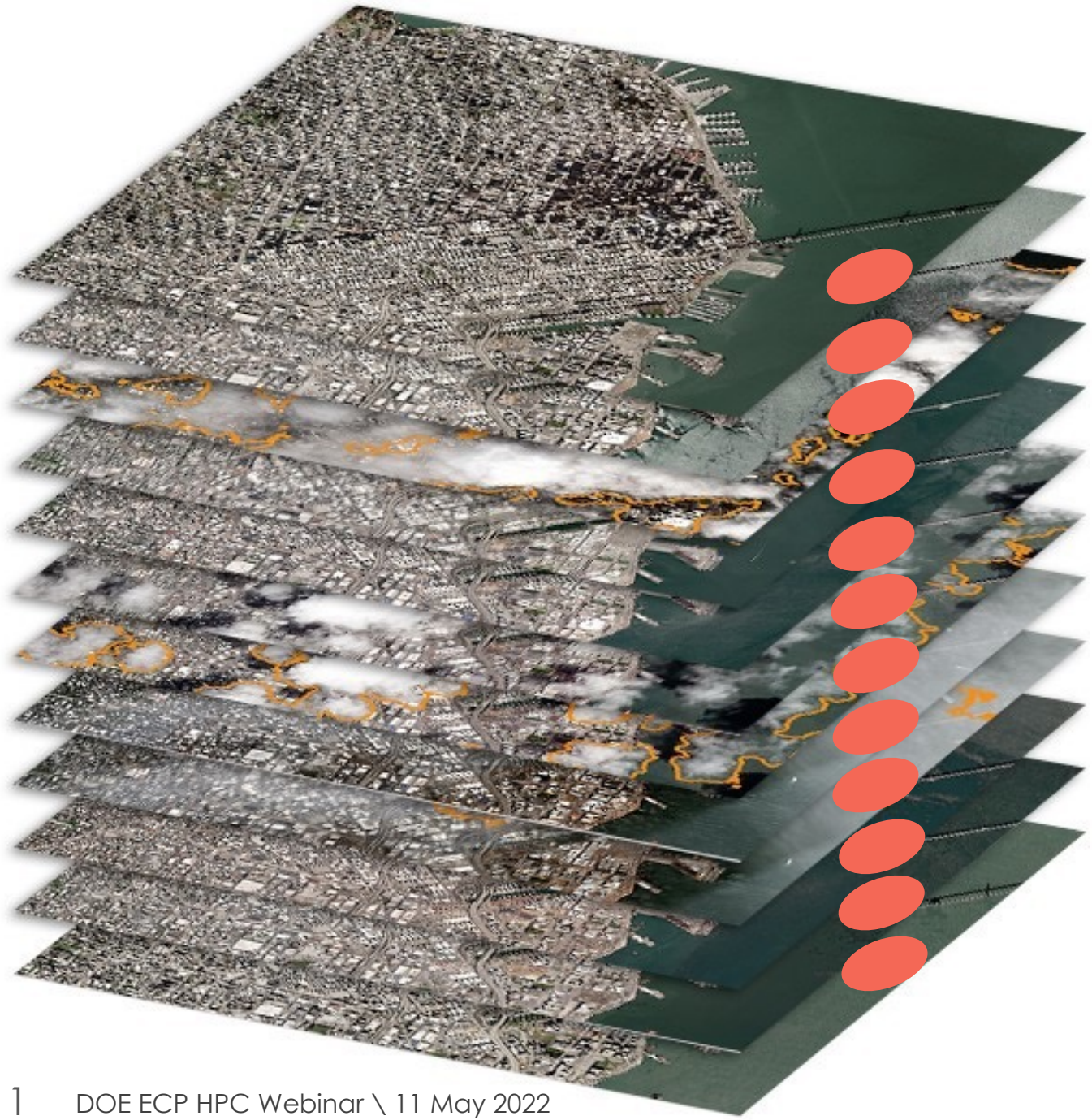


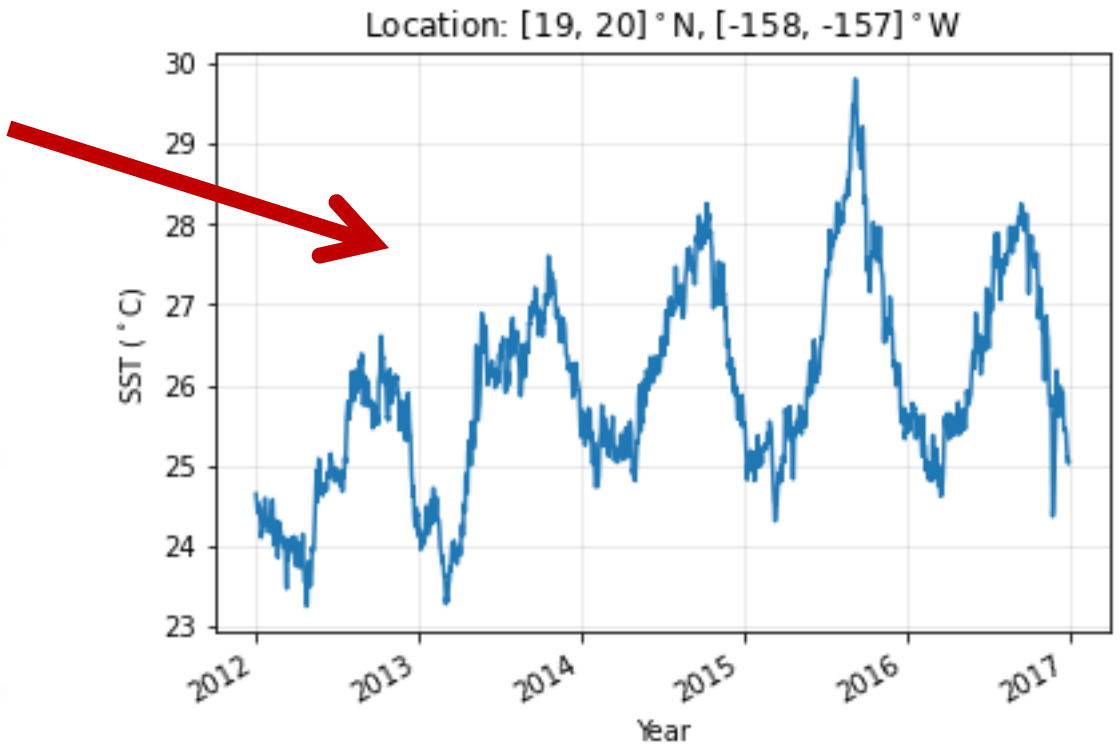
# ECOSYSTEMS RESEARCH
## (OR ANY OTHER FIELD)




- Analysis where little/short in situ physical data is available
- Investigate local/regional mechanisms

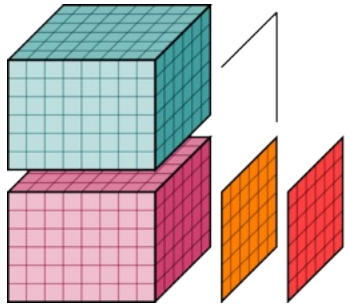
FARALLON INSTITUTE






FARALLON INSTITUTE

python™

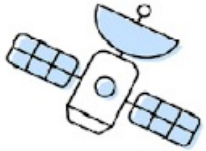


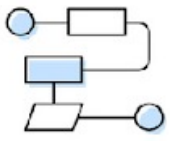
xarray

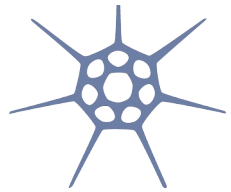


Google Earth Engine











FARALLON INSTITUTE

# But knowing where to start & discern the large (& changing) information available online can be intimidating & discouraging

This tutorial aims to provide an overview of what is possible, bypassing many of the background struggles we all had to surmount:

- Installing and updating python
- Having the correct libraries
- Learning the basic to arrive to the cool stuff
- Locating the data
- Learning how to get the data

# BUILD CAPACITY IN USE OF SATELLITE DATA & CLOUD RESOURCES FOR CLIMATE STUDIES


**Scientists**

- with limited expertise: *satellite data & computing*
- with limited resources: *storage, computing power & fast internet*
- need an update on tools and/or skills



FARALLON INSTITUTE

# ROADMAP

- Motivation
- The tutorial
- Lessons learned

FARALLON INSTITUTE

github.com/marisolgr/python_sat_tutorials

Search or jump to... / Pull requests Issues Marketplace Explore

marisolgr / python_sat_tutorials Public

Unwatch 1 Star 0 Fork 1

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

main 1 branch 0 tags Go to file Add file Code

marisolgr last details for pangeo binder 36647fe 4 days ago 110 commits

| | | |
|---|---|---|
| .ipynb_checkpoints | last details for pangeo binder | 4 days ago |
| data | Chapter 4 update, now chapter 4a | 7 months ago |

About

Tutorials to access and process satellite data in the cloud

Readme

GPL-3.0 License

# Satellite Data Python Tutorials launch binder

Tutorial to learn how to access and process Satellite Data using Python and JupyterLab in the Cloud

## Objective

This tutorial aims to provide scientists who want to use satellite data with the necessary tools for obtaining, temporally analyzing, and visualizing these data using the Cloud. **Note:** This in **not** a tutorial on Python per se - there are a myriad of resources for that. The purpose of this tutorial is to learn, through **examples**, only the necessary Python code and tools required to do simple temperal analysis of satellite data. We want you to get your toes wet, get to see and use the power of Python, and then maybe you will want to learn more. For that, we encourage you to visit the links on the **Resources** section at the end of each chapter.

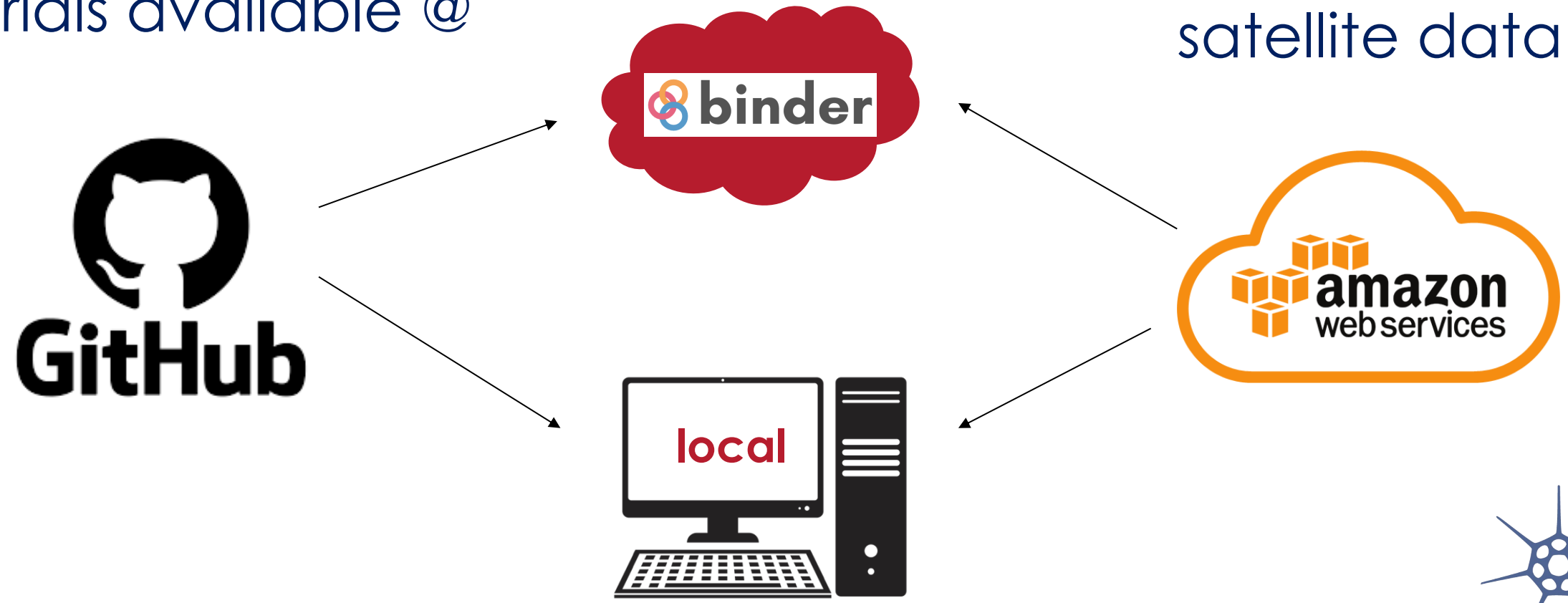
This project, supported by the Better Scientific Software foundation, and originally by NASA, aims to increase accessibility of satellite data and cloud technologies to a broad scientific community through easy-to-follow Python examples.



FARALLON INSTITUTE

# where to run it

tutorials available @




satellite data @



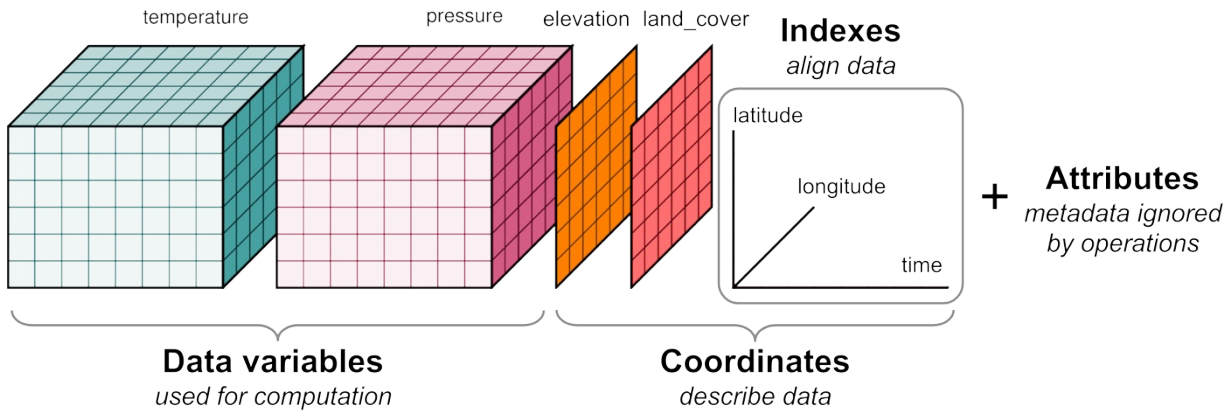
FARALLON INSTITUTE

# INTRODUCTORY CHAPTERS

- Chapter 1 – What's Python
- Chapter 2 – Jupyter Notebooks
- Chapter 3 – Basics of Python: Data structures
- Chapter 4a – xarray
- Chapter 4b – matplotlib
- Chapter 5 – Satellite data in the cloud

xarray

python™

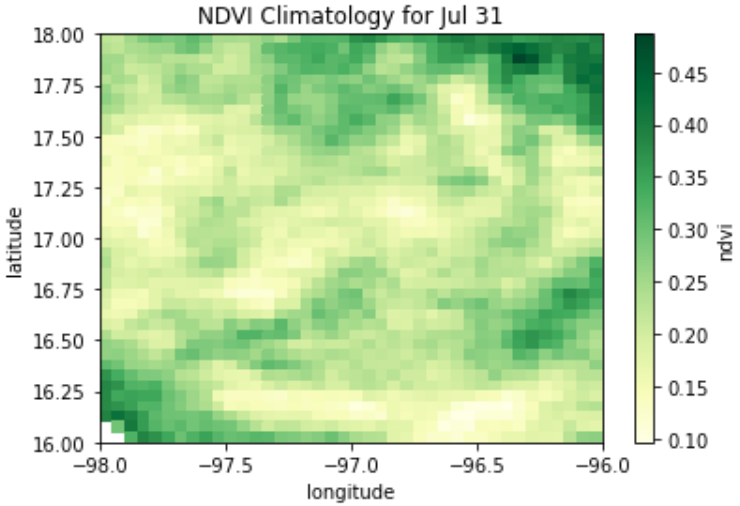


FARALLON INSTITUTE

# EXAMPLE CHAPTERS

- Chapter 6 – Ocean: SST
- Chapter 7 – Atmosphere: winds
- Chapter 8 – Land: Vegetation Index

- **Plot data, basic time series analysis**



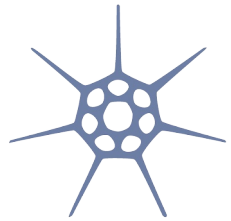


**Data:**
- Satellite
- Satellite-based

**Acquisition:**
- Cloud
- Online

# Tutorial for Acquisition & Analysis of **Time Series of Satellite** *Earth* **Data** in Python on the Cloud

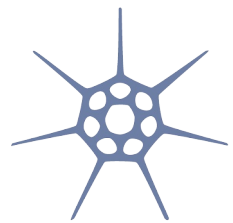
github.com/marisolgr/python_sat_tutorials

FARALLON INSTITUTE

# ROADMAP

- Motivation
- The tutorial
- Lessons learned



FARALLON INSTITUTE

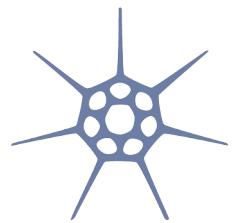
# PERSONAL PERSPECTIVE

- I am not an expert on technology or coding
- I am a user
  - one step removed from the end user who has limited coding expertise

FARALLON INSTITUTE

# LESSONS LEARNED

- Lessons learned:
  - Making the tutorials
  - Teaching the tutorials



FARALLON INSTITUTE

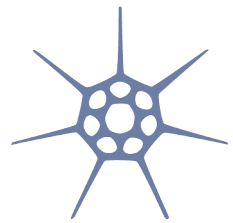
# TEACHING THE TUTORIAL

- Time series earth data acquisition is a challenge on its own
  - Gridded data formats vs. Time-step 'image' files
  - Generalizations are not simple because of diverse data types, but problems are similar
- Some coding expertise is necessary
- People that tried (and failed many times) to access similar data, find it most useful

FARALLON INSTITUTE

# MAKING THE TUTORIALS

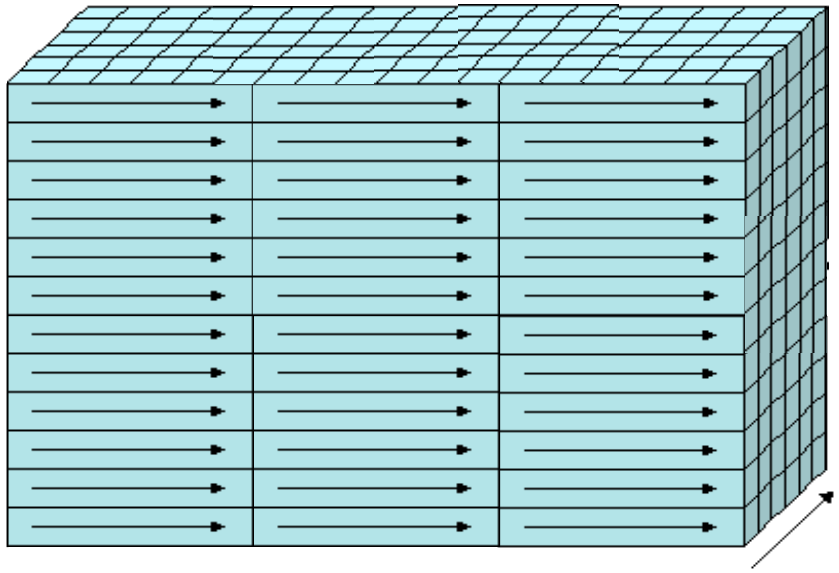
- Challenges:
  - Data
  - Software
  - Bringing it all together



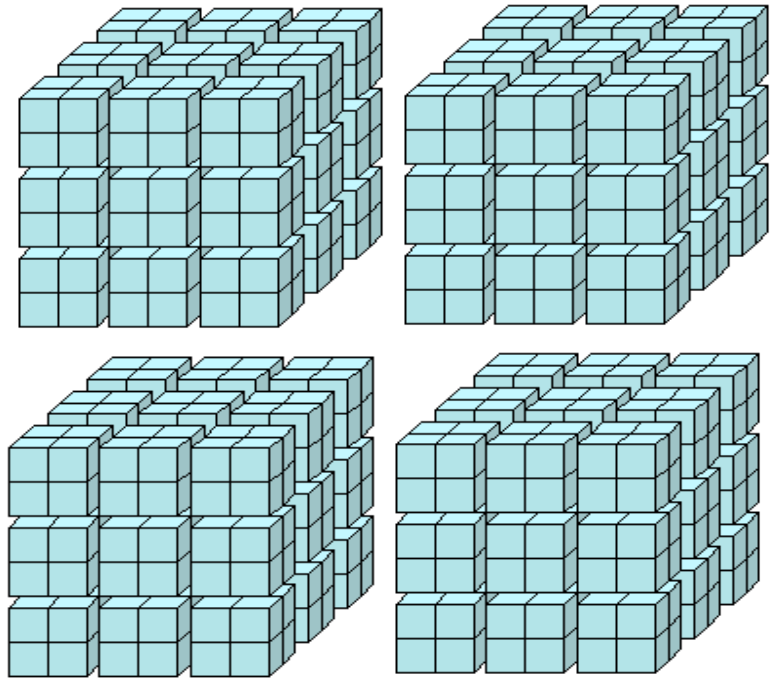
FARALLON INSTITUTE

# CHALLENGES WITH DATA

- Formats
- Storage
- Availability



index order



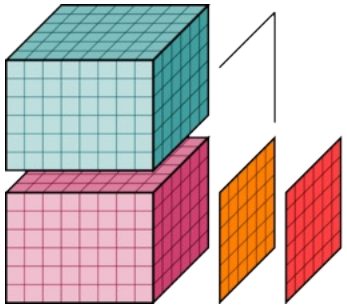
chunked



FARALLON INSTITUTE

# CHALLENGES WITH SOFTWARE

- New coding language, tools, libraries
  - Updates and versions
  - Availability and stability of tools
  - Access/functions differ for each data/format
- Coding expertise
  - New tools/methods can be difficult to grasp quickly

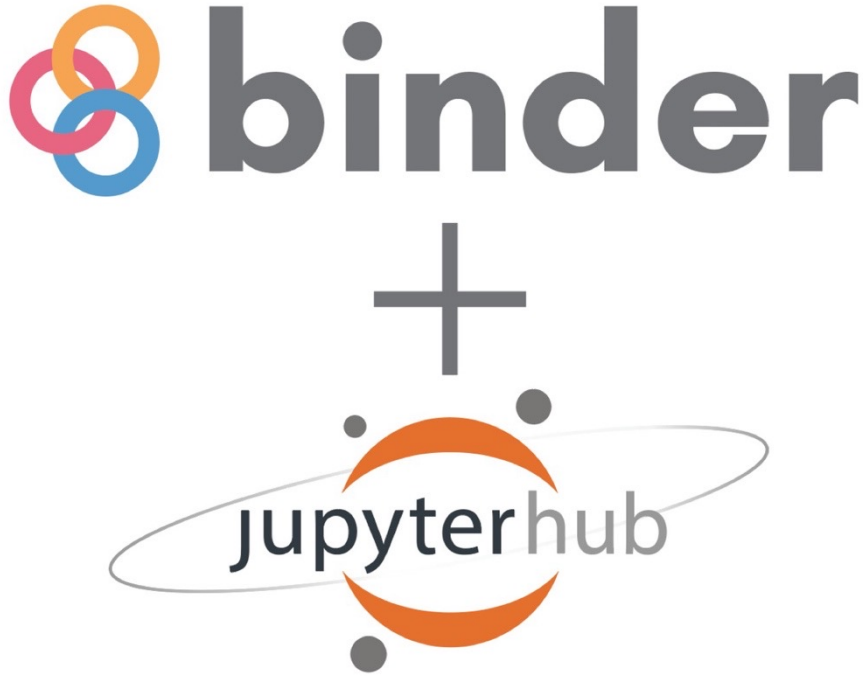


xarray

DASK

# BRINGING IT ALL TOGETHER

- Tools
  - Guidance for which one to use
  - Versions



# BRINGING IT ALL TOGETHER

- Tools
  - Guidance for which one to use
  - Versions
- Coding expertise
  - Troubleshooting work flow problems
  - Steep & long learning curve

FARALLON INSTITUTE

# ONE EXAMPLE

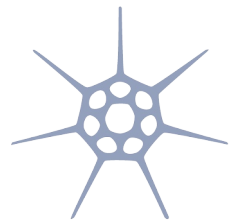
- Started with mybinder. Worked great!
- Few updates in the libraries were challenging when building in mybinder, and resulted in no access to data
- Since I couldn't fix access to AWS data, I switched to pangeo binder (another platform)
- A change in the Pangeo server configurations and images led to loose access to AWS data (libraries problems that the user – me - couldn't fix)
- Tried mybinder again, and worked!


# ONE EXAMPLE

- Started with mybinder. Worked great!
- Few updates in the libraries were challenging when building in mybinder, and resulted in no access to data
- Since I couldn't fix ac
  pangeo binder (anoth
- A change in the Pangeo server configurations and images led to loose access to AWS data (libraries problems that the user – me - couldn't fix)
- Tried mybinder again, and worked!

I had to get help from an expert!

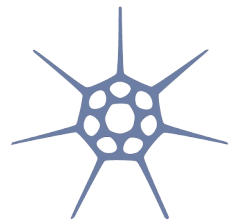




FARALLON INSTITUTE

# THE GOOD PARTS

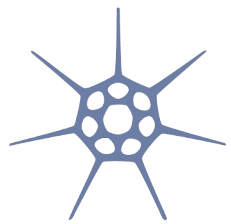
- Accessing the data for free, from a free online binder, was not possible when I started. Now you can access it from anywhere.
- New clever and promising ways to store and access data are emerging
- Things are changing, and fast, but seems like in a good direction
- Many fields have similar challenges, opportunities for collaborative work



FARALLON INSTITUTE

# WISH LIST

- ✓ Standard data formats that facilitate access & preprocessing in any axes
  - ✓ zarr or similar chunking/mapping techniques
- ✓ Maturity and stability of data & software
  - ✓ new and upgraded versions at a fast pace is not conducive to broader use
- ✓ More time to get there & to learn cloud computing



FARALLON INSTITUTE

# TAKE AWAY MESSAGES:

- Broad use of cloud data & resources - we are close but not there yet
- Access of data is not just availability of data – accessibility is important
- The computing expertise needed to access & acquire cloud data is very very underestimated

FARALLON INSTITUTE