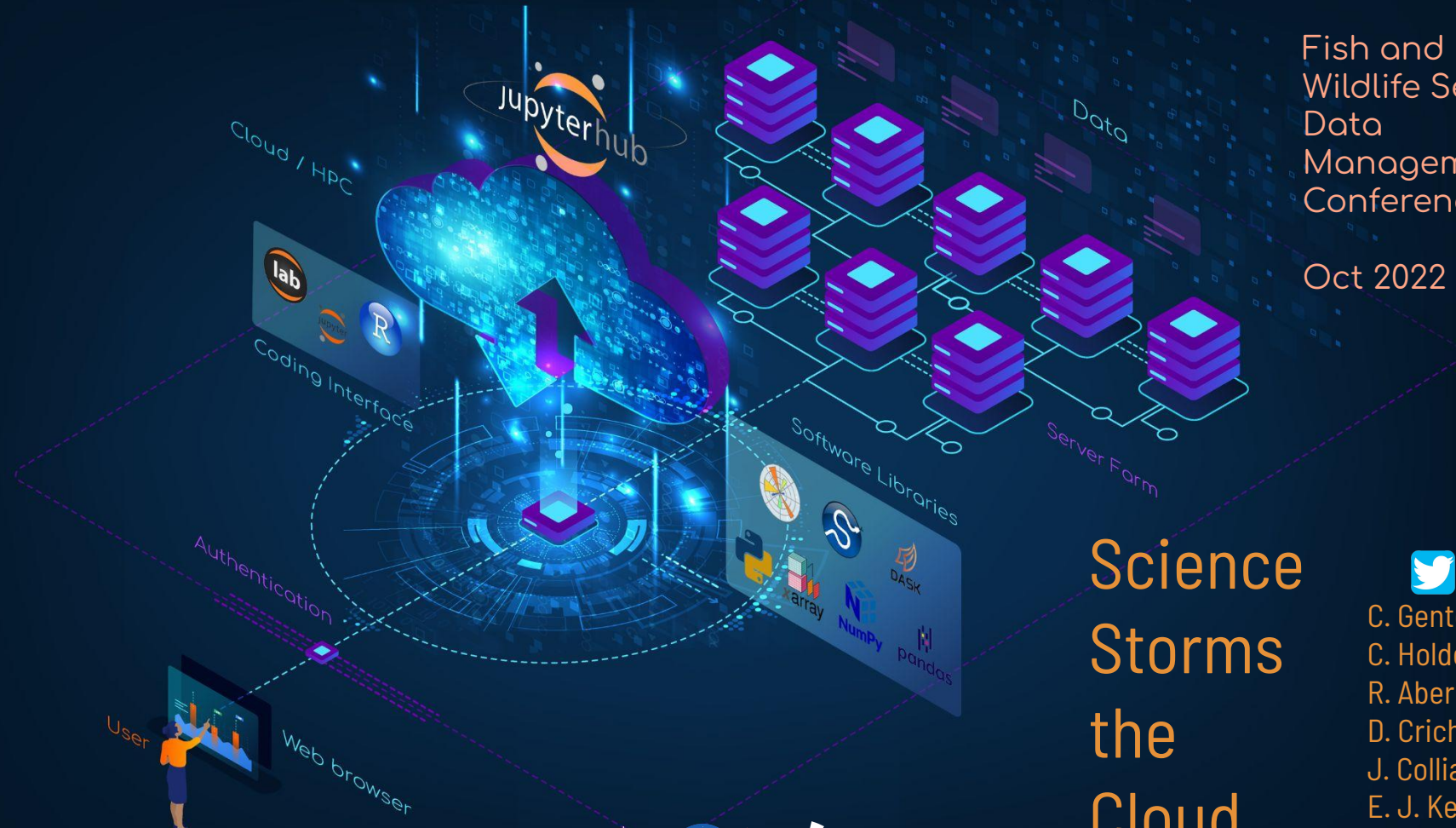


Oct 2022



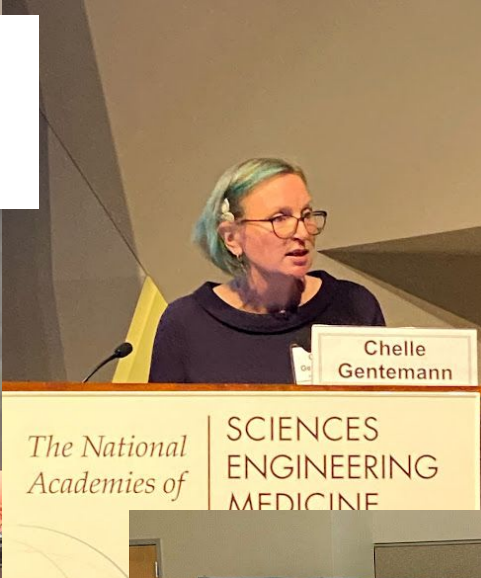
Science Storms the Cloud



- C. Gentemann
- C. Holdgraf
- R. Abernathy
- D. Crichton
- J. Colliander
- E. J. Kearns
- Y. Panda
- R. P. Signell

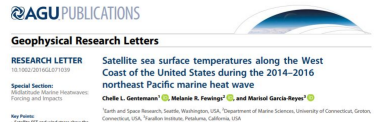


Who am I? Dr. Chelle Gentemann
Why am I here talking to you?
More: [@ChelleGentemann](https://twitter.com/ChelleGentemann) 



Left Commercial Company

Lost access to data and compute infrastructure.



Last Matlab/Fortran Paper

Started learning python. Not sure it is a big deal at first. October 2017 Tubbs fire - lost all data.



OceanHackWeek/Tutorials

Started helping lots of other people learn python/cloud computing.

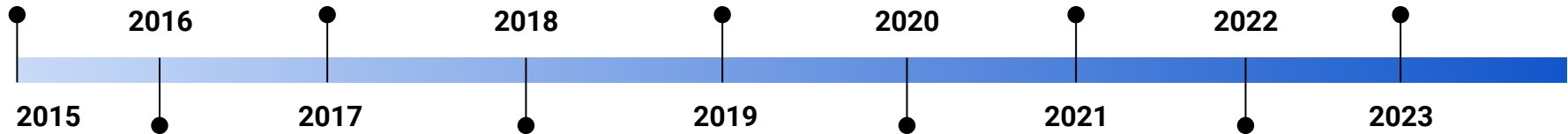


NASA TOPS Mission

Idea for Transforming to Open Science.

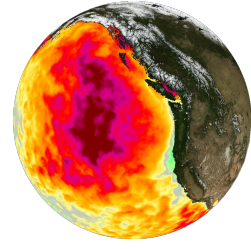
NASA Year of Open Science

What will you do?



Downloaded 8TBs datasets

Analysis of global 1km daily MUR SSTs. 23 days to subset; 3 months total churn on data for analysis; submitted paper August 2016



Python & Cloud Computing

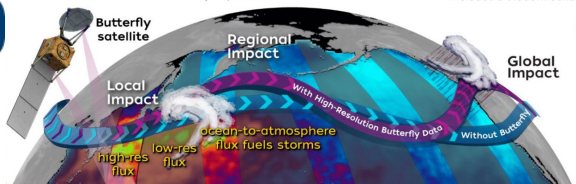
Python ecosystem too powerful to ignore any longer. Combining with cloud computing for reproducible science.

Learning, Frustration, Tears, Joy, Success, Tears, Joy....

PANGEO

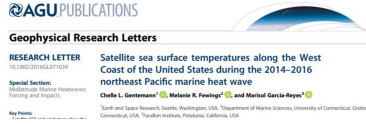
Earth Venture Mission - 3 - NASA competition - extended COVID version

PI on new proposed science mission for NASA - [Butterfly](#), completely cloud-based



Berkeley Stats 159

Using cloud-optimized version of MUR SST, entire class redid all figures and results in 2017 paper. [Notebook](#) is ~100 lines of code & runs in <10 minutes.



Left Commercial Company

Lost access to data and compute infrastructure.

Last Matlab/Fortran Paper

Started learning python. Not sure it is a big deal at first. October 2017 Tubbs fire - lost all data.

OceanHackWeek/Tutorials

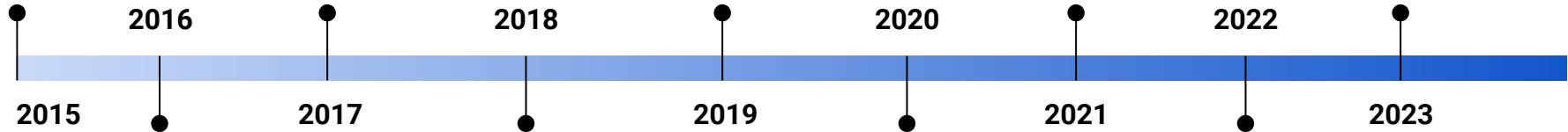
Started helping lots of other people learn python/cloud computing.

NASA TOPS Mission

Idea for Transforming to Open Science.

NASA Year of Open Science

What will you do?



2016

Downloaded 8TBs datasets

Analysis of global 1km daily MUR SSTs. 23 days to subset; 3 months total churn on data for analysis; submitted paper August 2016

degrees Celsius

2018

Python & Cloud Computing

Python ecosystem too powerful to ignore any longer. Combing with cloud computing for reproducible science.

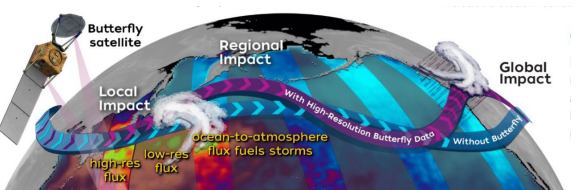
Learning, Frustration, Tears, Joy, Success, Tears, Joy....

PANGEO

2020

Earth Venture Mission - 3 - NASA competition - extended COVID version

PI on new proposed science mission for NASA - [Butterfly](#), **completely cloud-based**

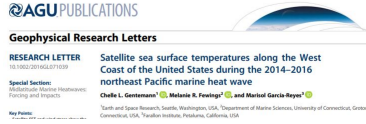


2022

Berkeley Stats 159

Using cloud-optimized version of MUR SST, entire class redid all figures and results in 2017 paper. [Notebook](#) is ~100 lines of code & runs in <10 minutes.





Left Commercial Company

Lost access to data and compute infrastructure.

Last Matlab/Fortran Paper

Started learning python. Not sure it is a big deal at first. October 2017 Tubbs fire - lost all data.

OceanHackWeek/Tutorials

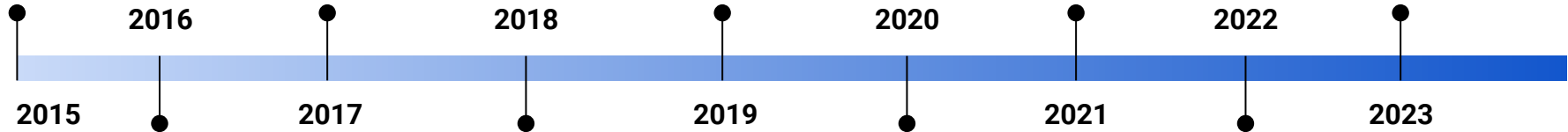
Started helping lots of other people learn python/cloud computing.

NASA TOPS Mission

Idea for Transforming to Open Science.

NASA Year of Open Science

What will you do?



2015

2016

2017

2018

2019

2020

2021

2022

2023

Downloaded 8TBs datasets

Analysis of global 1km daily MUR SSTs. 23 days to subset; 3 months total churn on data for analysis; submitted paper August 2016

Python & Cloud Computing

Python ecosystem too powerful to ignore any longer. Combining with cloud computing for reproducible science.

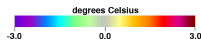
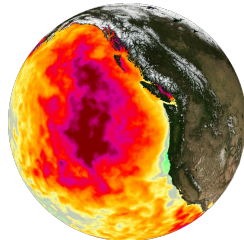
Earth Venture Mission - 3 - NASA competition - extended COVID version

PI on new proposed science mission for NASA - [Butterfly](#), completely cloud-based

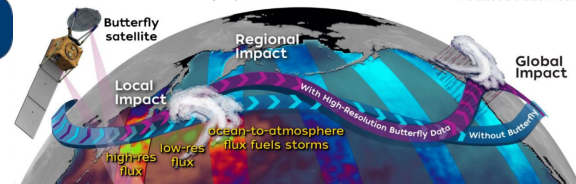
Berkeley Stats 159

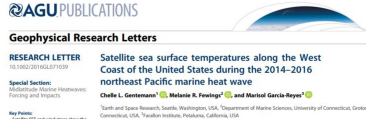
Using cloud-optimized version of MUR SST, entire class redid all figures and results in 2017 paper. [Notebook](#) is ~100 lines of code & runs in <10 minutes.

Learning, Frustration, Tears, Joy, Success, Tears, Joy....



PANGEO





Left Commercial Company

Lost access to data and compute infrastructure.

Last Matlab/Fortran Paper

Started learning python. Not sure it is a big deal at first. October 2017 Tubbs fire - lost all data.

OceanHackWeek/Tutorials

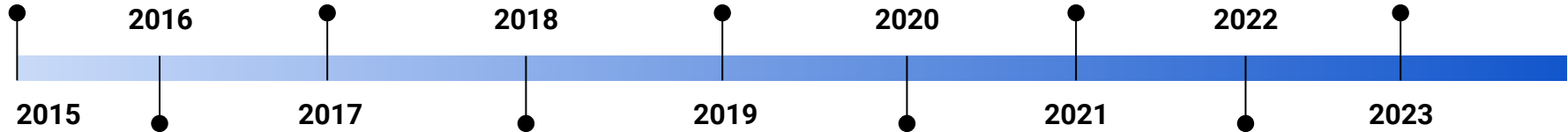
Started helping lots of other people learn python/cloud computing.

NASA TOPS Mission

Idea for Transforming to Open Science.

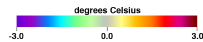
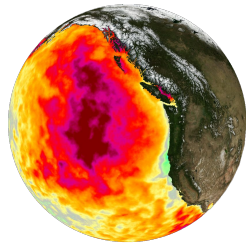
NASA Year of Open Science

What will you do?



Downloaded 8TBs datasets

Analysis of global 1km daily MUR SSTs. 23 days to subset; 3 months total churn on data for analysis; submitted paper August 2016



Python & Cloud Computing

Python ecosystem too powerful to ignore any longer. Combing with cloud computing for reproducible science.

Learning, Frustration, Tears, Joy, Success, Tears, Joy....

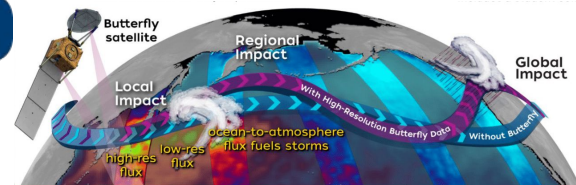
PANGEO

Earth Venture Mission - 3 - NASA competition - extended COVID version

PI on new proposed science mission for NASA - [Butterfly](#), completely cloud-based

Berkeley Stats 159

Using cloud-optimized version of MUR SST, entire class redid all figures and results in 2017 paper. [Notebook](#) is ~100 lines of code & runs in <10 minutes.



Data

Bring the data to you

Months to download

New version? Download again

Storage and maintenance costs

Discourages external collaborations

Reinforces institutional advantages

Closed Software

A black and white photograph of a cemetery. In the foreground, there is a large, ornate cross. To its left is a dark, rectangular tombstone with the inscription "Ps. 62. 2. Meine Seele ist stille zu Gott, der mir hilft." Other various styles of tombstones and a large, pointed Gothic-style monument are visible in the background. The ground is overgrown with grass and weeds.

Redundant efforts

Error prone

Impedes advancements

Difficult to share, version, etc.

Reinforces closed science

Local Infrastructure



Compiler-specific OS

Unique environment

Software restrictions

Reinforces institutional advantages

Pay-wall Publishing



Restricts access to knowledge

Perpetuates exclusionary practices

Reinforces institutional advantages

How does who participates in science affect solutions?

Is a solution the *best* solution?

Heart valves and seat belts are made that only fit men's bodies (significantly increasing mortality rates for women)

AI cropping algorithms for Twitter/Zoom have racial biases

Voice-recognition software only recognizes the voices of men



Hacking Diversity with Inclusive Decision Making

New research reveals how inclusive decision making activates diversity for better business performance and a decisive competitive advantage.

RESEARCH ARTICLE | SOCIAL SCIENCES

Gender-diverse teams produce more novel and higher-impact scientific ideas

Yang Yang, Tanja Y. Tiao, Teresa K. Woodruff, and Brian Uzor

Edited by Susan Fiske, Princeton University, Princeton, NJ; received January 14, 2022; accepted July 24, 2022

August 29, 2022 | 119 (38): e2200841119 | <https://doi.org/10.1073/pnas.2200841119>

96767

Significance

Science teams made up of men and women produce papers that are more novel and highly cited than those of all-men or all-women teams. These performance advantages increase the greater the team's gender balance and appear nearly universal. On average, they hold for small and large teams, the 45 subfields of medicine, and women- or men-led teams and generalize to published papers in all science fields over the last 20 y. Notwithstanding these benefits, gender-diverse teams remain underrepresented in science when compared to what is expected if the teams in the data had been formed without regard to gender. These findings reveal potentially new gender- and teamwork synergies that correlate with scientific discoveries and inform diversity, equity, and inclusion (DEI) initiatives.

Abstract

Science's changing demographics raise new questions about research team diversity and

nature communications

Explore content | About the journal | Publish with us

nature > nature communications > articles > article

Article | Open Access | Published: 04 December 2018

The preeminence of ethnic diversity in scientific collaboration

Rodnor K. Alshbali, Jalal Bahwan, & Wei Lee Woon

Nature Communications 9 Article number: 5163 (2018) | [Cite this article](#)

26k Accesses | 147 Citations | 618 Altmetric | [Metrics](#)

and economic benefits of diversity, we analyze over 9 million papers to study the relationship between research impact and five classes of discipline, gender, affiliation, and academic age. Using randomized establish the presence of homophily in ethnicity, gender and affiliation. ffect of diversity on scientific impact, as reflected in citations. :lasses considered, ethnic diversity had the strongest correlation with :urther isolate the effects of ethnic diversity, we used randomized I again found a clear link between diversity and impact. To further tgs, we use coarsened exact matching to compare the scientific impact of pters and scientists with closely-matched control groups. Here, we find :resulted in an impact gain of 10.63% for papers, and 47.67% for scientists.

Murphy, M. C., Mejia, A. F., Mejia, J., Yan, X., Cheryan, S., Dasgupta, N., et al. (2020). Open science, communal culture, and women's participation in the movement to improve science. *Proceedings of the National Academy of Sciences*, 117(39), 24154-24164. <https://doi.org/10.1073/pnas.1921320117>

Technology is increasing our ability to build together

Remove barriers

Address identified needs collaboratively



Collaborations
accelerate the
speed of science
and amplifies
impacts

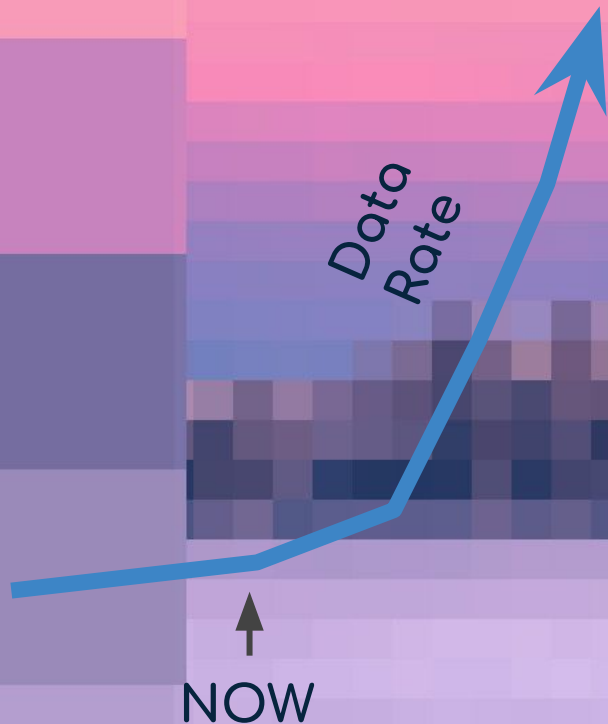
Open project include more people

Better results





Data



Cloud-based data

Easier to collaborate

Easier to reproduce and build on

Access not bandwidth-limited

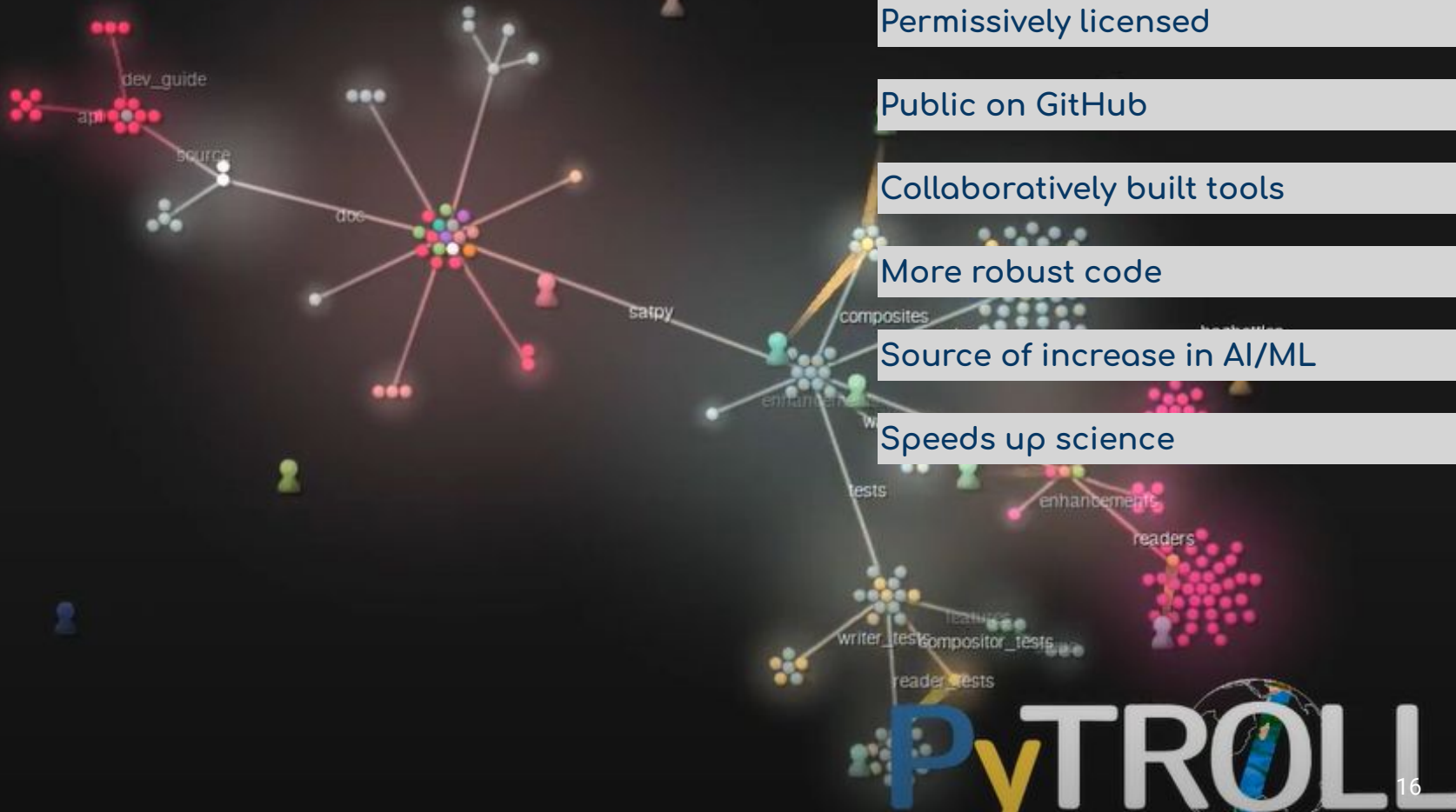
More Interdisciplinary research

Broadens participation



Open Software

2018-Apr-22

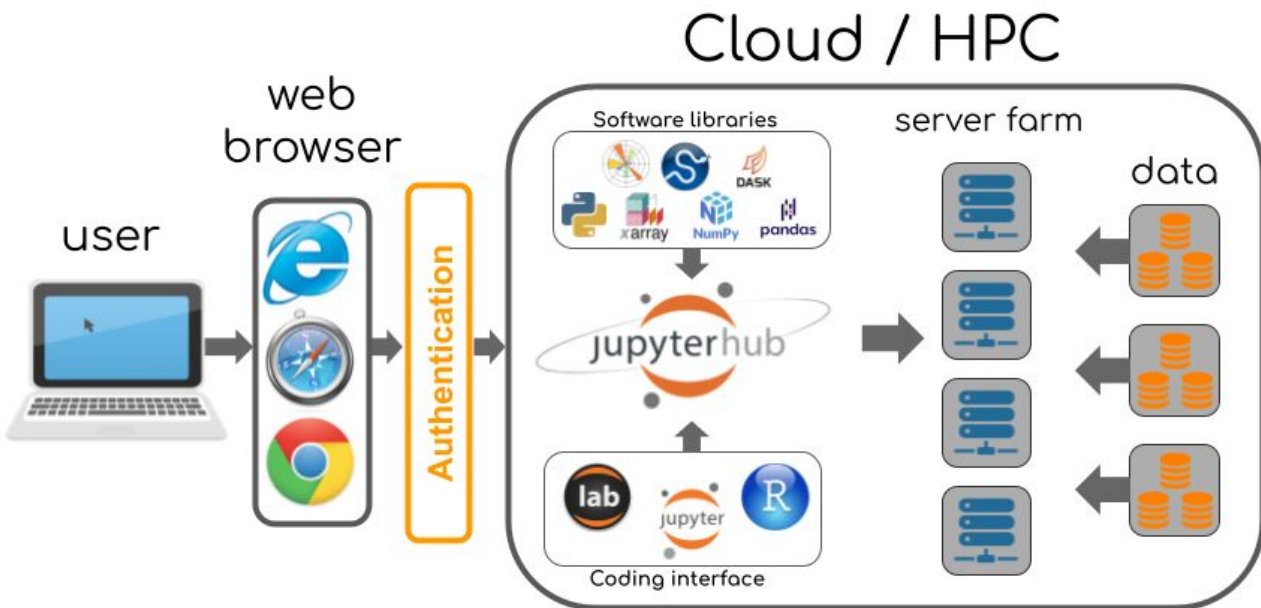


Open Cyberinfrastructure - Science Data Platforms

Platform agnostic, open source infrastructure solutions

Developed by the community

Rapidly becoming default for science

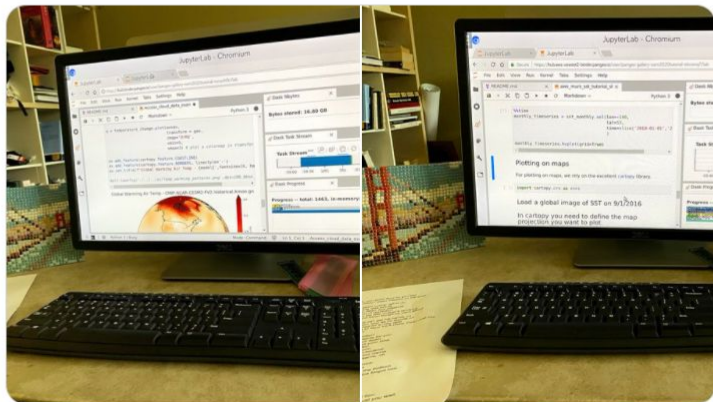


A super computer behind every device



Chelle-ter in Place
@ChelleGentemann

On my kids @Raspberry_Pi running @TeamKano #OpenSource OS I'm analyzing @GCPcloud #cmip6 climate data and @awscloud MUR SST from @podaac. A \$36 computer running processes on both AWS and GCP with over 80 workers and 245GB. #openscience! @NASAEarth



1:18 PM · Mar 1, 2020 · Twitter Web App

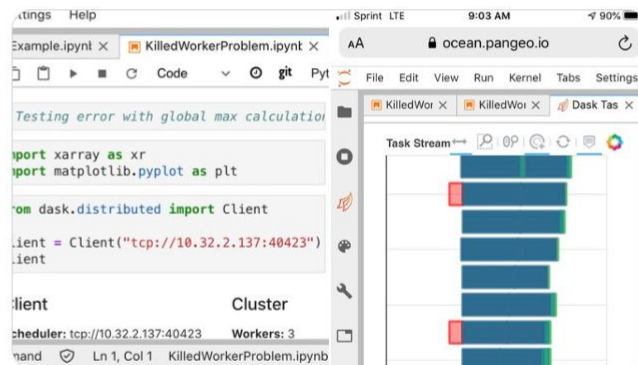


Julius Busecke @JuliusBusecke · Feb 6

I am analyzing #CMIP6 on the train on MY PHONE!



Goddamn it, @pangeo_data is amazing! This has literally been the only time I have wanted a bigger phone screen 😂



2 13 66

Replies



Thomas Moore @SurfTasmania · Feb 6

Replying to @JuliusBusecke and @pangeo_data
Nice! That's way cooler than spinning up a @pangeo_data HPC cluster on your laptop from 35,000 feet up. 🚀

Some of the biggest recent breakthroughs were enabled by open science

First image of a black hole

“We’re deeply grateful to all the open source contributors who made our work possible.” –Dr. Katie Bouman

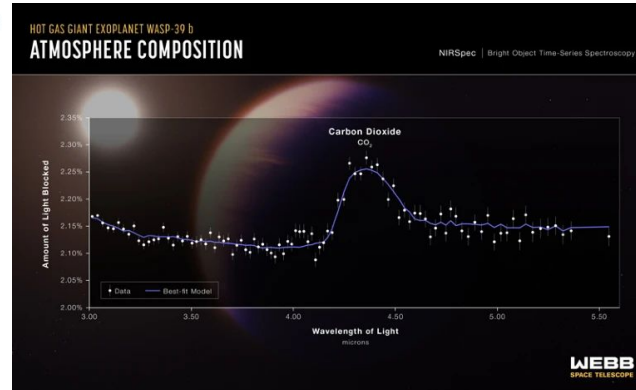


“with the open source projects in NumFOCUS, we were able to iterate our algorithms so fast that they enabled us to finish our work in two years”

we greatly improve[d] our own work by adopting well-tested community packages that contain the collected wisdom of many other projects.” –Dr. Lindy Blackburn

“The open source community is very important for scientists; imagine if we had to do everything from scratch every single time.” –Dr. Chi-Kwan Chan

Astronomers see CO₂ on exoplanet for first time



“NASA’s open science guiding principles are centered in our Early Release Science work, supporting an inclusive, transparent, and collaborative scientific process.”
- co-author Dr. Natasha Batalha

- Open collaboration that was advertised in the years leading up to data collection with open Slack community (341 people and counting!!)
- JWST data was made public immediately upon collection
- All data reduction and scientific interpretation can be reproduced through open software and data archived via Zenodo Community Collection
- Open preprint server (Arxiv)
- Published Nature Open Access

Leading the Path to Open-Source Science



NASA's Transform to Open Science (TOPS) is a \$40 million 5-year mission geared towards accelerating the adoption & understanding of open science

Key Goals:

- 20,000 earn open science certification
- 2x participation by historically underrepresented communities
- 5+ major discoveries

2023 is NASA's Year of Open Science

NASA Science has designated 2023 as the Year of Open Science. Throughout the year NASA will be energizing and uplifting open science across the scientific community through:



Visibility

Open Science everywhere: Articles, announcements, Twitter Spaces, conferences

2023 Big annual meetings
Open Science Themes, integrated into society comms



Capacity Sharing Resources

Online, free, Open Science curriculum on Open edX

Workshops, events, virtual cohorts, science team meetings, hackathons

Many paths to Open Science



Incentives

Open Science Badge/Certification

High profile prizes and challenges

High profile awards in support of open science research



Changing the Game

Require open data, open software, open access

Funding decisions consider open science activities

Awards, promotions, evaluations consider Open Science activities and teams as well as individuals

How YOU can Get Involved:

To implement a cultural shift, we need community engagement from the broad spectrum across the scientific community!

We are looking for community partners to co-develop YOOS activities

- Develop open science action plans
- Share your data, software, publications
- Nominate science teams for summer schools
- Organize events
- Join TOPS email list!

Learn more and collaborate with us - we're working on GitHub!



TOPS Email List



TOPS Website



Q&A

Learn more and
collaborate with us!



TOPS Email List



TOPS Website