



Transforming to Open Science: perspectives on how to best support open science

Paige E. Martin
Program Scientist
NASA Chief Science Data Office

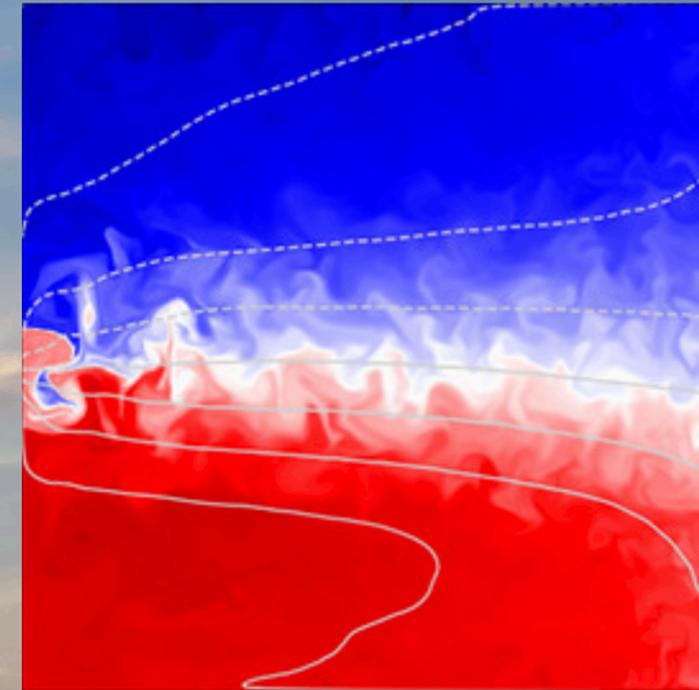


MY BACKGROUND

PHYSICAL OCEANOGRAPHY



PhD @ University of Michigan



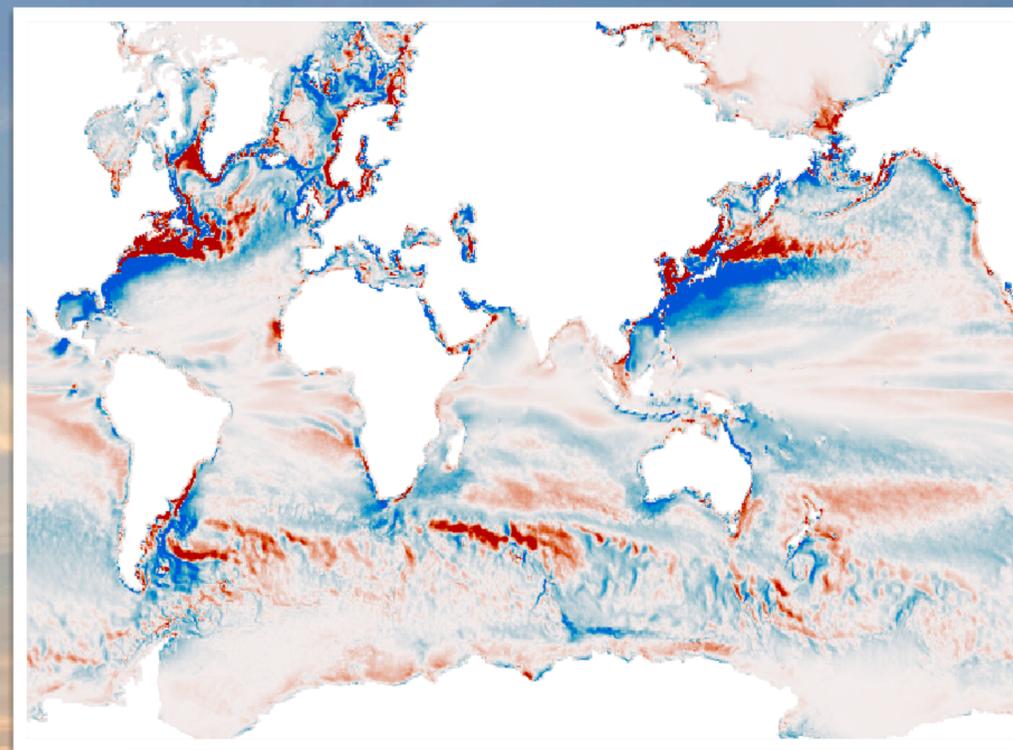
Energy budgets and temperature variance in idealized model

MY BACKGROUND

CLIMATE DATA SCIENCE



Postdoc @ Lamont Doherty
Earth Observatory



Air-sea interaction in CESM

CURRENTLY

NASA'S CHIEF SCIENCE DATA OFFICE



Program Scientist



NASA's Transform to Open
Science (TOPS) & Open Source
Science Initiative



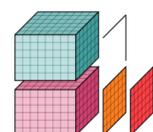
A community for big data geoscience

WHAT IS PANGEO?

- Community obsessed with efficient data processing.

Founded in 2016. Scientists and software developers coming together. <http://pangeo.io/>
Weekly meeting / seminar. Discourse Forum. Annual meeting. Workshops at AGU / AMS / etc.

- Interoperable Software



xarray



Foundation in Scientific Python: Jupyter, Xarray, Dask, Zarr. Broad ecosystem of interoperable packages for analysis, visualization, and machine learning.

- Data and Computing Infrastructure

Deployment recipes for cloud and HPC. Open, public, cloud-based JupyterHubs and Binders for Data-proximate computing. PB of analysis-ready, cloud-optimized data stored in public cloud (GCS, AWS) and OpenStorageNetwork.

Open Source Science

Climate Science

Life Science

Materials Science

Reproducibility

Map of Science

A community that brings together
scientific users and software
engineers to accelerate science

Pathways that guided me to open science

1. Accessibility: I lost access to Matlab 😞

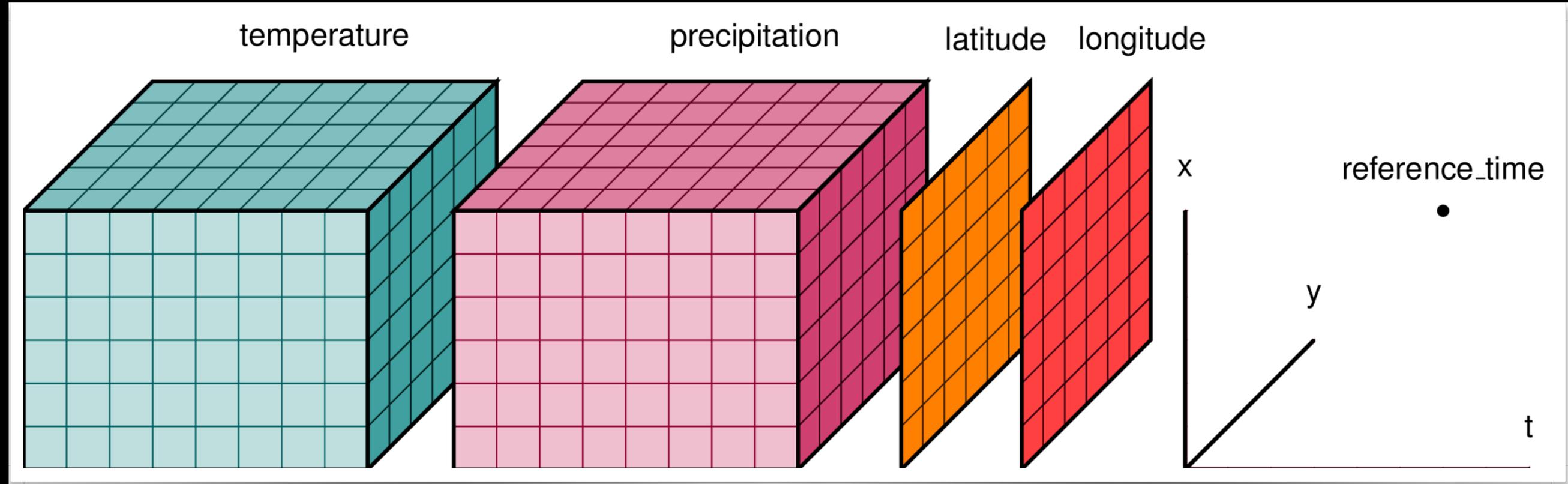


Pathways that guided me to open science

1. **Accessibility**: I lost access to Matlab 😞
2. **Scalability**: open source software allowed for handling of my large, complex datasets 🤸



SCALABILITY



$$\begin{aligned}
 0 = 2 \int \int & \left(\underbrace{-\text{Re} \left[\widehat{oT_m}^* (\widehat{o u_m oT_m})_x \right] - \text{Re} \left[\widehat{oT_m}^* (\widehat{o v_m oT_m})_y \right]}_{\text{horizontal advection}} + \underbrace{\frac{1}{oH_m} \text{Re} \left[\widehat{oT_m}^* \widehat{o w_{ek} oT_m} \right]}_{\text{vertical advection}} + \underbrace{\frac{1}{o\rho oC_p oH_m} \text{Re} \left[\widehat{oT_m}^* \widehat{F_m^{e+}} \right]}_{\text{entrainment heat flux}} \right. \\
 & - \underbrace{\frac{1}{o\rho oC_p oH_m} \text{Re} \left[\widehat{oT_m}^* \widehat{F_\lambda} \right]}_{\text{sensible/latent heat flux}} - \underbrace{\frac{1}{o\rho oC_p oH_m} \text{Re} \left[\widehat{oT_m}^* (\widehat{F_0^\uparrow} + \widehat{F_m^\downarrow}) \right]}_{\text{radiative heat flux}} \\
 & \left. + \underbrace{\text{Re} \left[\widehat{oT_m}^* oK_2 \widehat{\nabla_H^2 oT_m} \right] - \text{Re} \left[\widehat{oT_m}^* oK_4 \widehat{\nabla_H^4 oT_m} \right]}_{\text{diffusion}} \right) dx dy,
 \end{aligned}$$

Pathways that guided me to open science

1. **Accessibility**: I lost access to Matlab 😞
2. **Scalability**: open source software allowed for handling of my large, complex datasets 🦒
3. **Inclusion**: open source (and free) software removed barriers to doing science 🎉





Mumin Olatunji Oladipo

Communication Physics, Nigeria
PhD Student (University of Ilorin)
Assistant Lecturer (KolaDaisi University, Ibadan)
muminoladipo@gmail.com



coessing.org



Mumin Olatunji Oladipo

Communication Physics, Nigeria
PhD Student (University of Ilorin)
Assistant Lecturer (KolaDaisi University, Ibadan)
muminoladipo@gmail.com



“**Instrumentation is a challenge in developing countries** because of poor power supply, technical knowhow and access to sensors and other components. Hence, **research is being impeded** by no availability of data due to these factors.

Upon learning python programing at COESSING, I was able to **construct a reliable VLF [very low frequency] receiver to collect [solar flare] data for my PhD work.** ...Thanks to COESSING, Python language was useful while setting up the device. The device has been taking data for almost a year (still ongoing).

Hitherto, I was using a desktop for the data collection, but the desktop consumes about 300W which was **beyond my budget for power backup.** Consequently, the Raspberry pi setup solves the problem of power- it consumes less than 10W - therefore, our budget on power backup was sufficient. **The project was a success** and I had configured a similar system for Kebbi State University of Science and Technology, Aliero- Nigeria and I am currently assisting to setup a similar system at University of Ilorin- Nigeria.”

```

Jupyter gbz_file_1 Last Checkpoint: Last Saturday at 10:56 (autosaved)
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3
+ -> Run Code
In [5]: #DAA = data_raw_vlf.Loc[:, ('DATE', 'Daily Average', 'Day Average', 'Night Average') ]
#DAT = data_temp.Loc[:, ('DATE', 'average-o', 'average-w', 'average-c', 'average-h', 'average-l', 'average-s', 'average-e', 'average-i', 'average-a', 'average-b', 'average-ab') ]
#DAA.to_csv('VLF_RAW.csv')
#DAT.to_csv('TEMP_AVE.csv')

datadir = '/Users/SAMSUNGPC/'
data = pd.read_excel('VLF_RAW.xlsx', index_col = 0, sheet_name='VLF_RAW')
data = data.loc['2019-08-01':'2019-08-31', :]
print (data)

In [6]: corr = data.corr()
corr

Out[6]:
           Daily Average  Day Average  Night Average  average-o  average-w  average-c  average-h  average-l  average-s  average-e  average-i  average-a  average-b  average-ab
Daily Average  1.000000  0.972648  0.953500  0.074800 -0.026797  0.219549 -0.320528 -0.587462 -0.382693 -0.330761 -0.563809 -0.227503 -0.227503  0.101721
Day Average    0.972648  1.000000  0.894270  0.022815 -0.079514  0.197691 -0.346919 -0.545342 -0.404048 -0.302806 -0.567913 -0.124391 -0.124391  0.123350
Night Average  0.953500  0.894270  1.000000  0.102800 -0.011966  0.205962 -0.326352 -0.574592 -0.306539 -0.296215 -0.505550 -0.209371 -0.209371 -0.013157
average-o      0.074800  0.022815  0.102800  1.000000  0.582877  0.665004  0.415745  0.042398  0.021848  0.411342  0.004581  0.133546  0.133546  0.056057
average-w     -0.026797 -0.079514 -0.011966  0.582877  1.000000  0.410417  0.479483  0.194342  0.329175  0.289502  0.061506 -0.023869 -0.023869  0.155984

```



National Aeronautics and
Space Administration

NASA's Open-Source Science Strategy

Supporting a more equitable, impactful, and
efficient scientific future

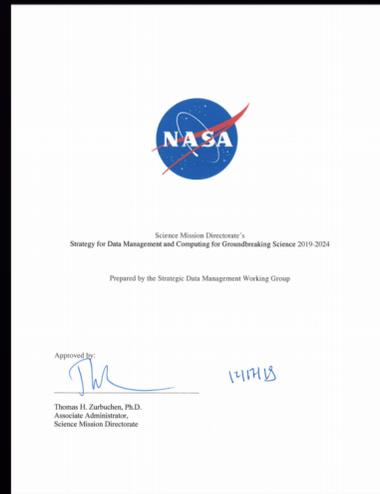


Open Science

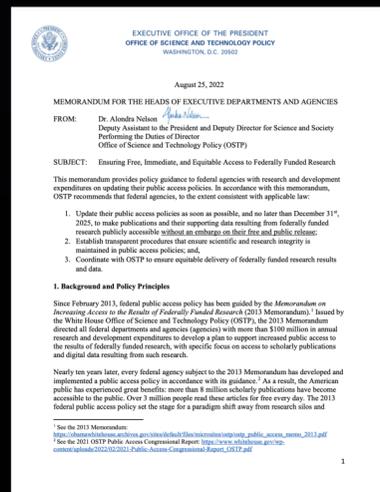
is the principle and practice of making research products and processes available to all, while respecting diverse cultures, maintaining security and privacy, and fostering collaborations, reproducibility and equity.



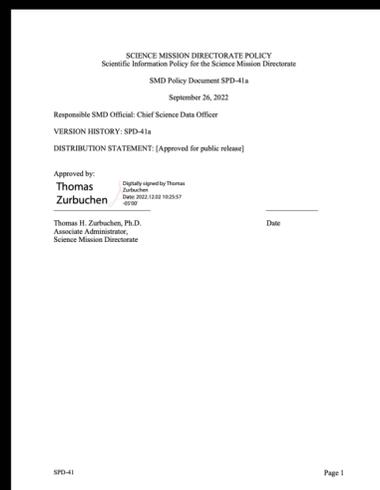
Background



[2019 - 2024]
SMD Strategy for Data Management and Computing for
Groundbreaking Science



[August, 2022]
OSTP Memo: Ensuring Free, Immediate, and Equitable
Access to Federally Funded Research



[September, 2022]
Scientific Information Policy for the Science Mission
Directorate: *SPD-41a*

Chief Science Data Office

GOAL 1

Develop and Implement Capabilities to Enable Open Science

GOAL 2

Continuous Evolution of Data and Computing Systems

GOAL 3

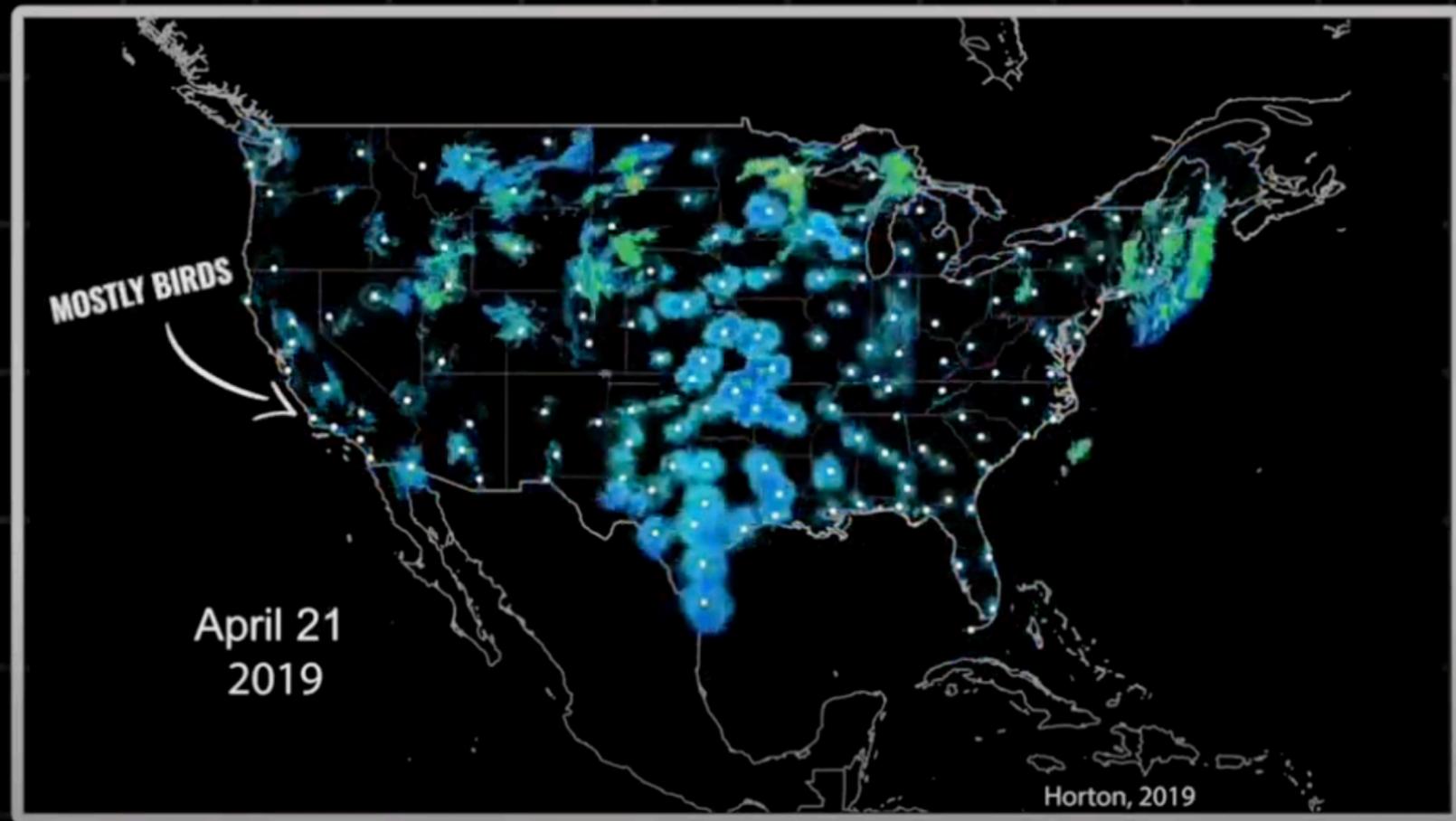
Harness the Community and Strategic Partnerships for Innovation



“Unity is a strength...when there is teamwork and collaboration, wonderful things can be achieved.”

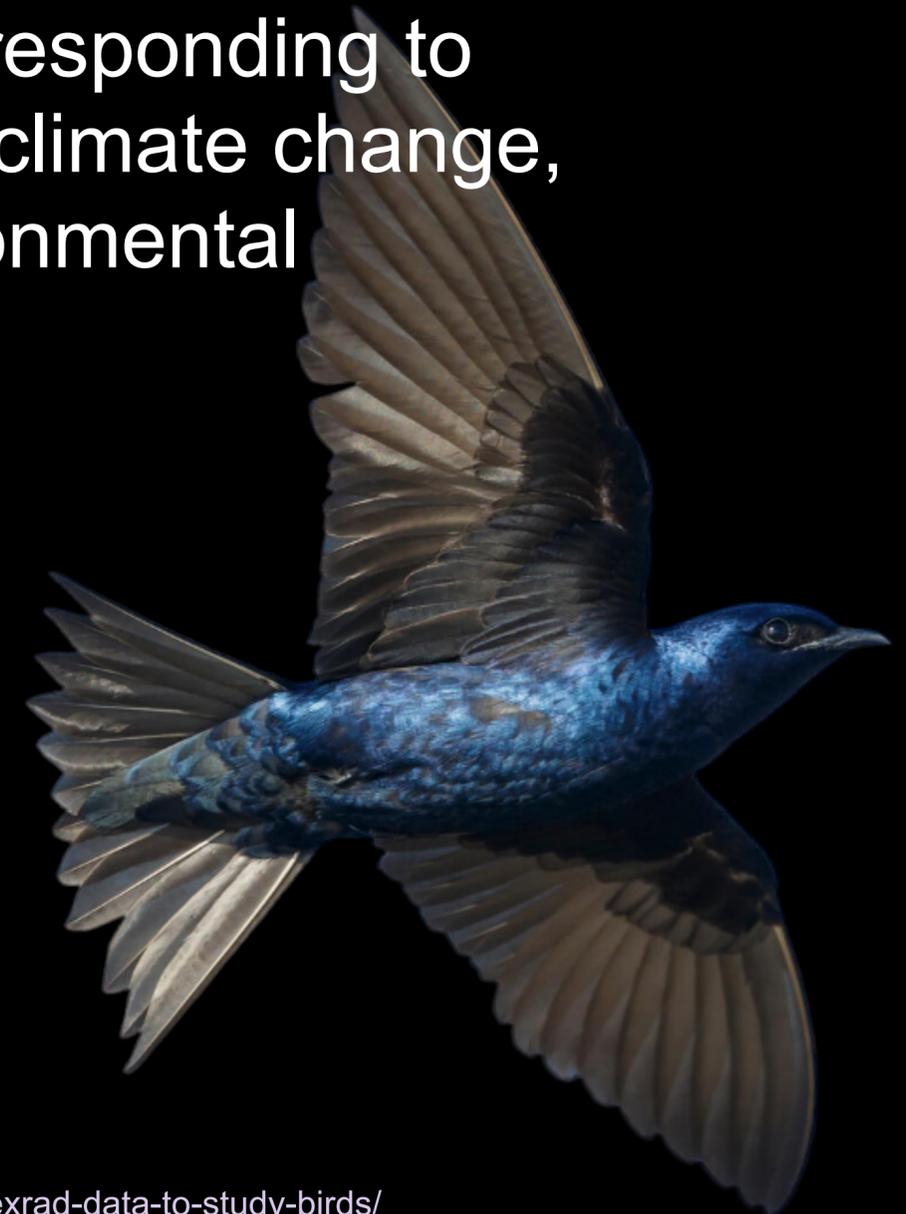
Mattie Stepanek

Open Data Enables New Science



Representation

Birds are responding to droughts, climate change, and environmental changes



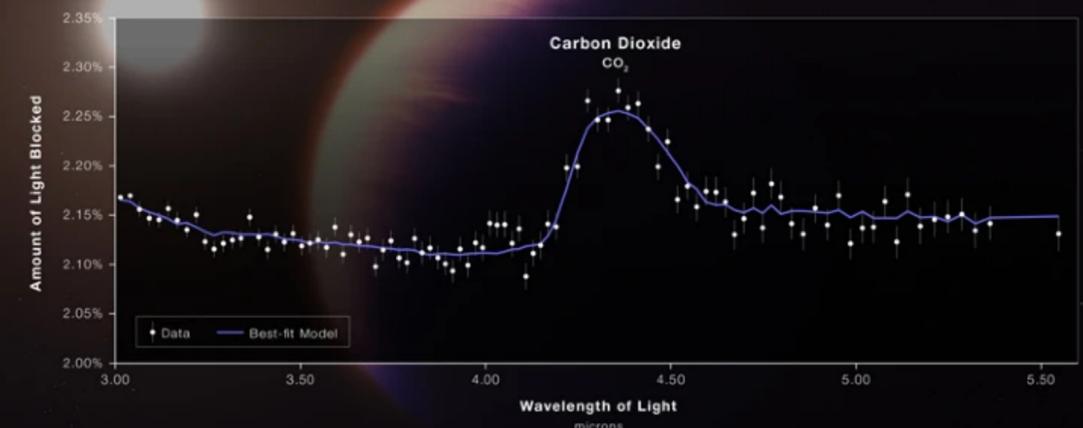
First step towards discovering life on another planet

Early Release science data leads to major discovery & 40+ publications

<https://www.nature.com/articles/d41586-022-02350-2>

HOT GAS GIANT EXOPLANET WASP-39 b ATMOSPHERE COMPOSITION

NIRSpec | Bright Object Time-Series Spectroscopy





NASA's

Open-Source Science Initiative

NASA's approach for putting Open Science into practice.



Policy and Governance

Implement policies that advance open science, support SMD working groups and meetings, and develop standards and governance.

Core Data and Computing Services

Develop SMD-wide data and computing infrastructure (Cloud and HEC), provide tools for discovery of NASA's scientific information, reduce burden of SPD-41a, and support the adoption of advanced technologies (AI/ML).

NASA's Open-Source Science Initiative

Open Science Incentives

Grants, prizes and challenges to enable groundbreaking scientific discoveries using open science principles and tools.

Community Engagement



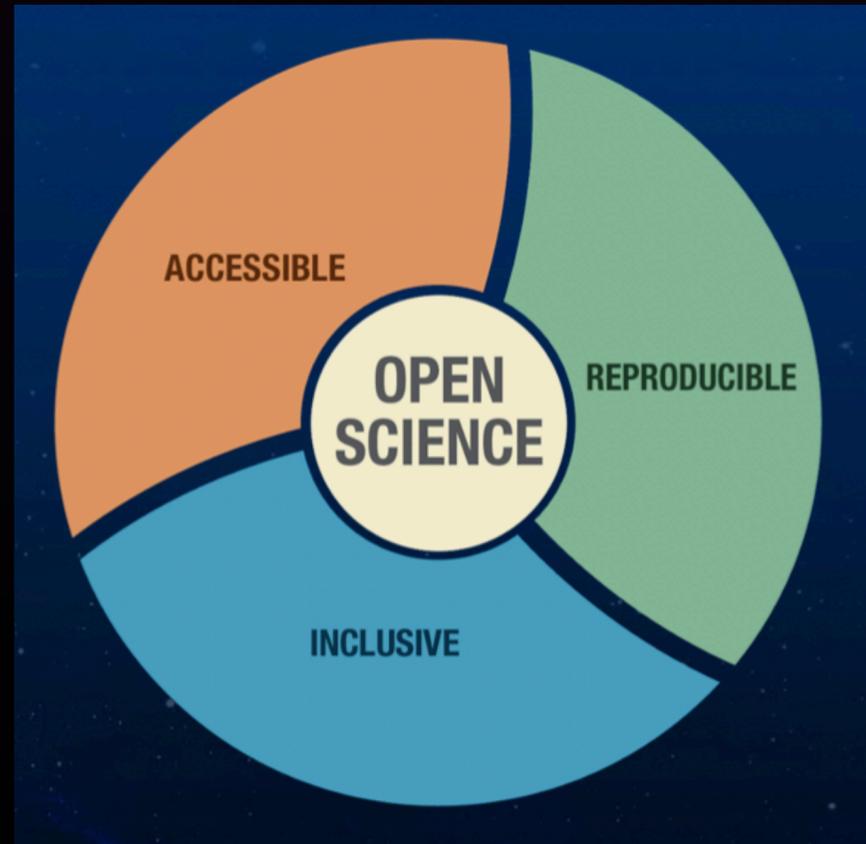
Advance open science practices in the SMD community and build strategic partnerships for innovation in open science.

SMD = Science Mission Directorate

Policy & Governance



NASA policy “SPD-41a”



SPD-41a is built on the Open-Source Science Principles of Accessibility, Reproducibility, and Inclusion

- covers the open sharing of publications, data, and software created in the pursuit of scientific knowledge
- aims to make NASA science as open as possible, as restricted as required, and always secure.
- looks to maximize openness while minimizing the burden on researchers.

[Link to full SPD-41a policy](#)
[Links to policy guidance \(PDF & GitHub\)](#)



SPD-41a is SMD's updated Scientific Information Policy

SPD-41a is *forward looking* and will apply to all future SMD-funded scientific activities

Major Policy Updates

- Peer-reviewed publications are made openly available with no embargo period.
- Research data and software are shared at the time of publication or the end of the funding award.
- Mission data are released as soon as possible and unrestricted mission software is developed openly.
- Science workshops and meetings are held openly to enable broad participation.

SMD = Science Mission Directorate

Core Data and Computing Services



Core Data and Computing Services Program

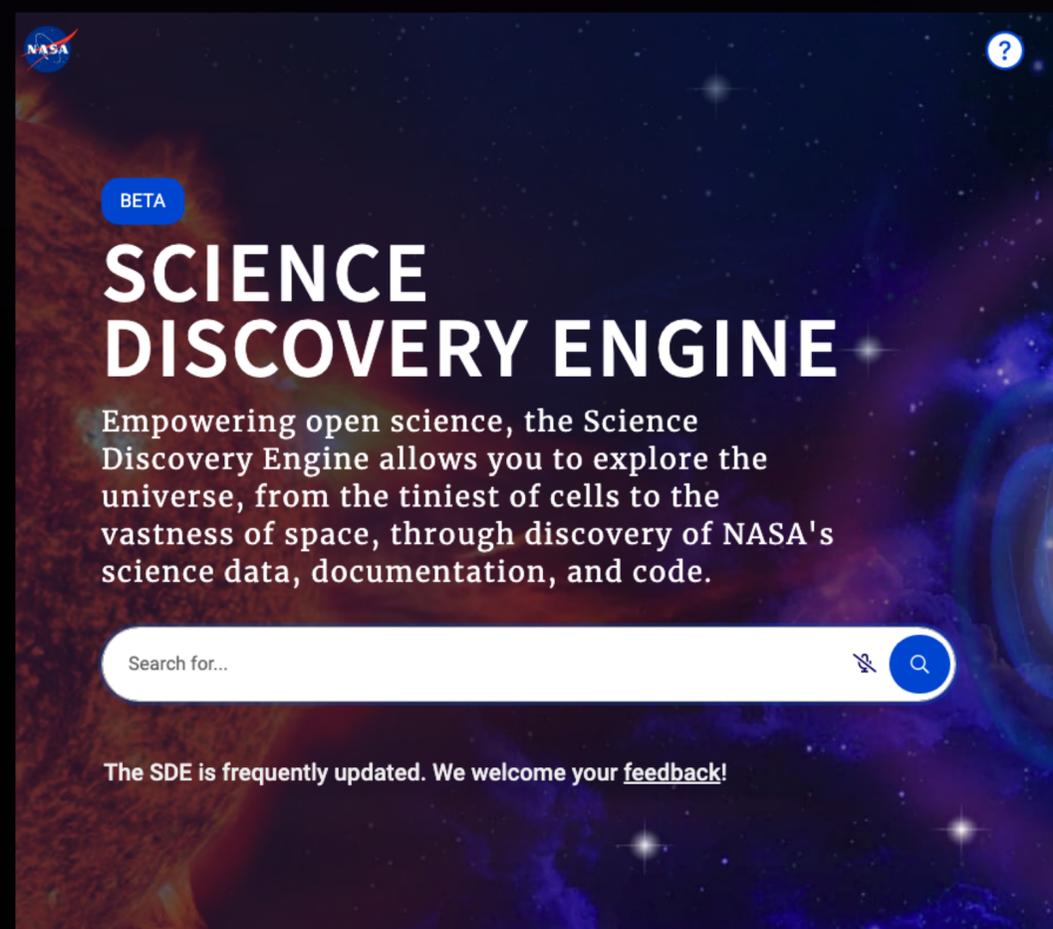
The **Core Data and Computing Services Program** will provide a layered architecture on which SMD science Divisions can seamlessly and efficiently integrate their discipline-specific services such as data archives.

SMD Core Data and Computing Services will:

- Develop SMD-wide data and computing infrastructure to support Open Science
- Develop services to support the adoption of SPD-41a by SMD Divisions

Infrastructure: Core Services

Making NASA data Findable



Science Discovery Engine

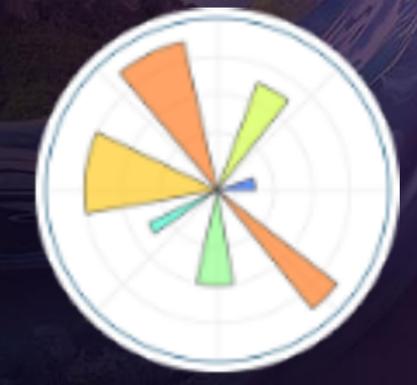
SMD data catalog to support discovery and access to complex scientific data across Divisions

Open Science Incentives



Open Science Incentives

NASA provides funding for open science through a solicitations and opportunities. This includes supporting innovative open science projects and open source tools, frameworks, and libraries.



NASA Funding for Open-Source Science in 2023

F.15	High Priority Open-Source Science	Innovative open-source tools, software, frameworks, data formats, and libraries that will have a significant impact on the SMD science community (<i>OPEN NOW! Rolling deadline in ROSES-23</i>)
F.7	Support for Open Source Tools, Frameworks, and Libraries	Improve and sustain open source tools, frameworks, and libraries that are significantly used by the SMD community (<i>OPEN SOON! ROSES-23 dates TBD</i>)
F.8	Supplemental Open Source Science Awards	Supplemental award to support open science including the conversion of legacy software to open source (<i>OPEN NOW! Rolling deadline in ROSES-23</i>)

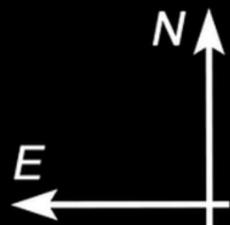
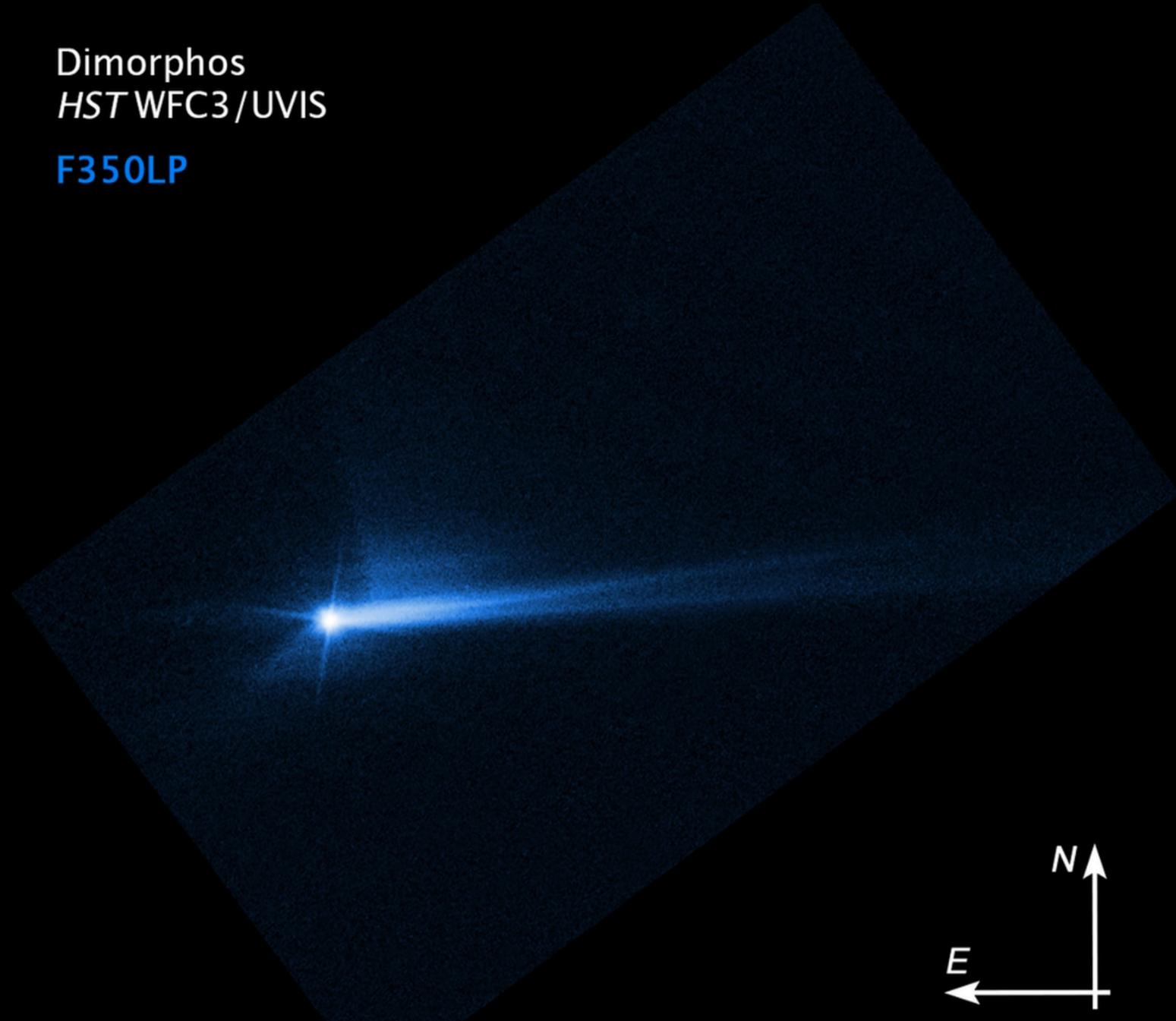
Community Engagement





NASA is looking ahead at really big challenges

Dimorphos
HST WFC3/UVIS
F350LP

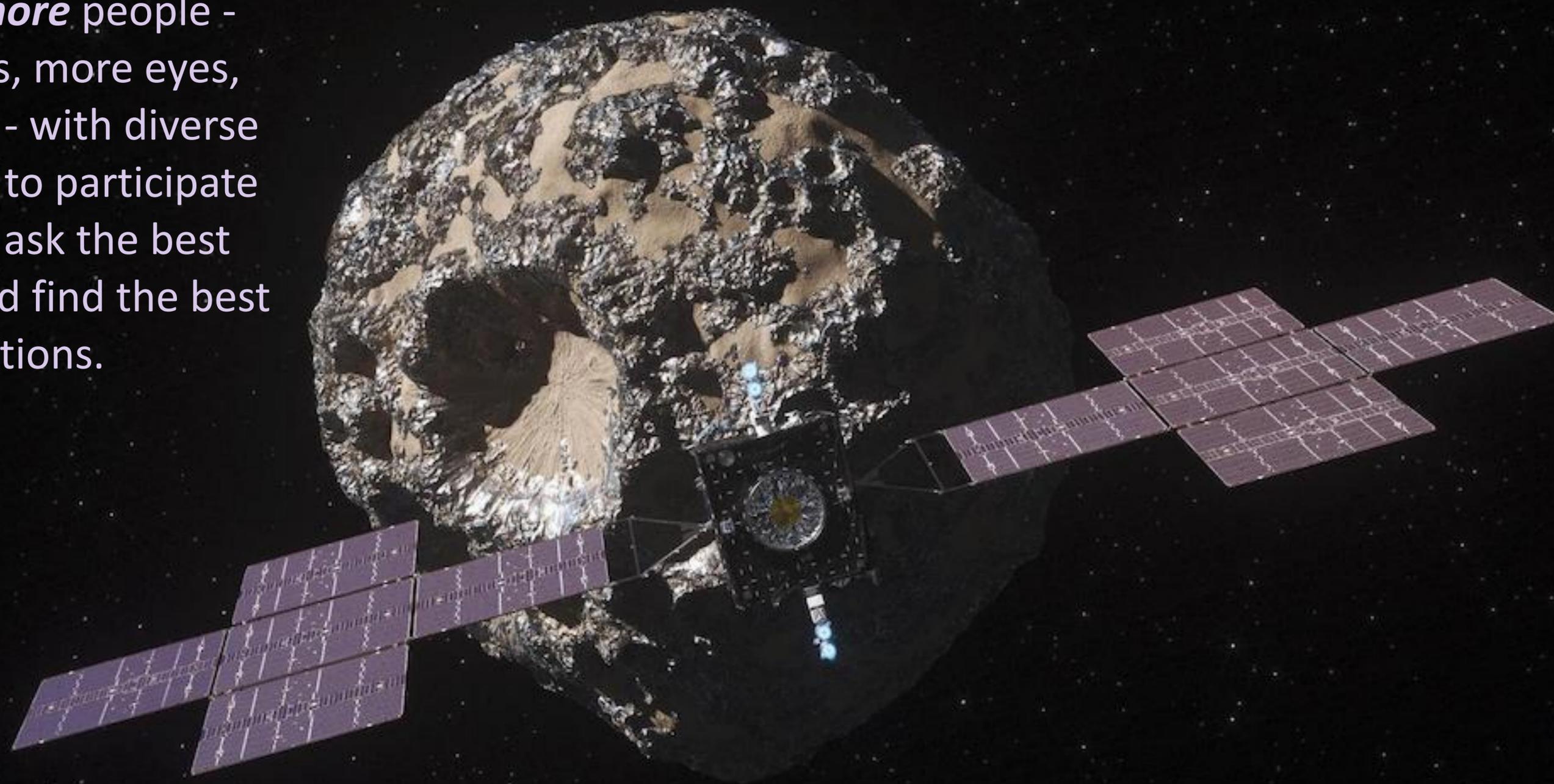


We need *more* WE science rather than ME science¹— openly sharing data, software, & results.





We need *more* people -
more hands, more eyes,
more brains - with diverse
experiences to participate
so that we ask the best
questions and find the best
solutions.



NASA's Transform to Open Science (TOPS)

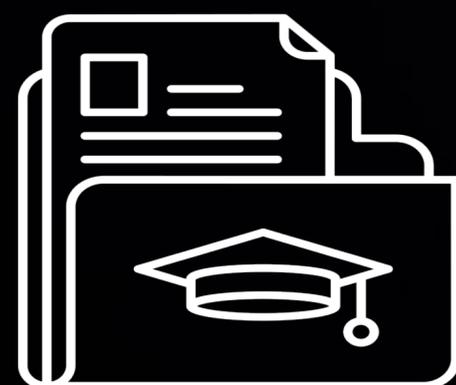
a 5-year mission to accelerate adoption of open science

Goals:

- Increase understanding and adoption of open science principles and techniques
- Broaden participation by historically excluded communities
- Accelerate major scientific discoveries



Engagement



Capacity Sharing



Incentives



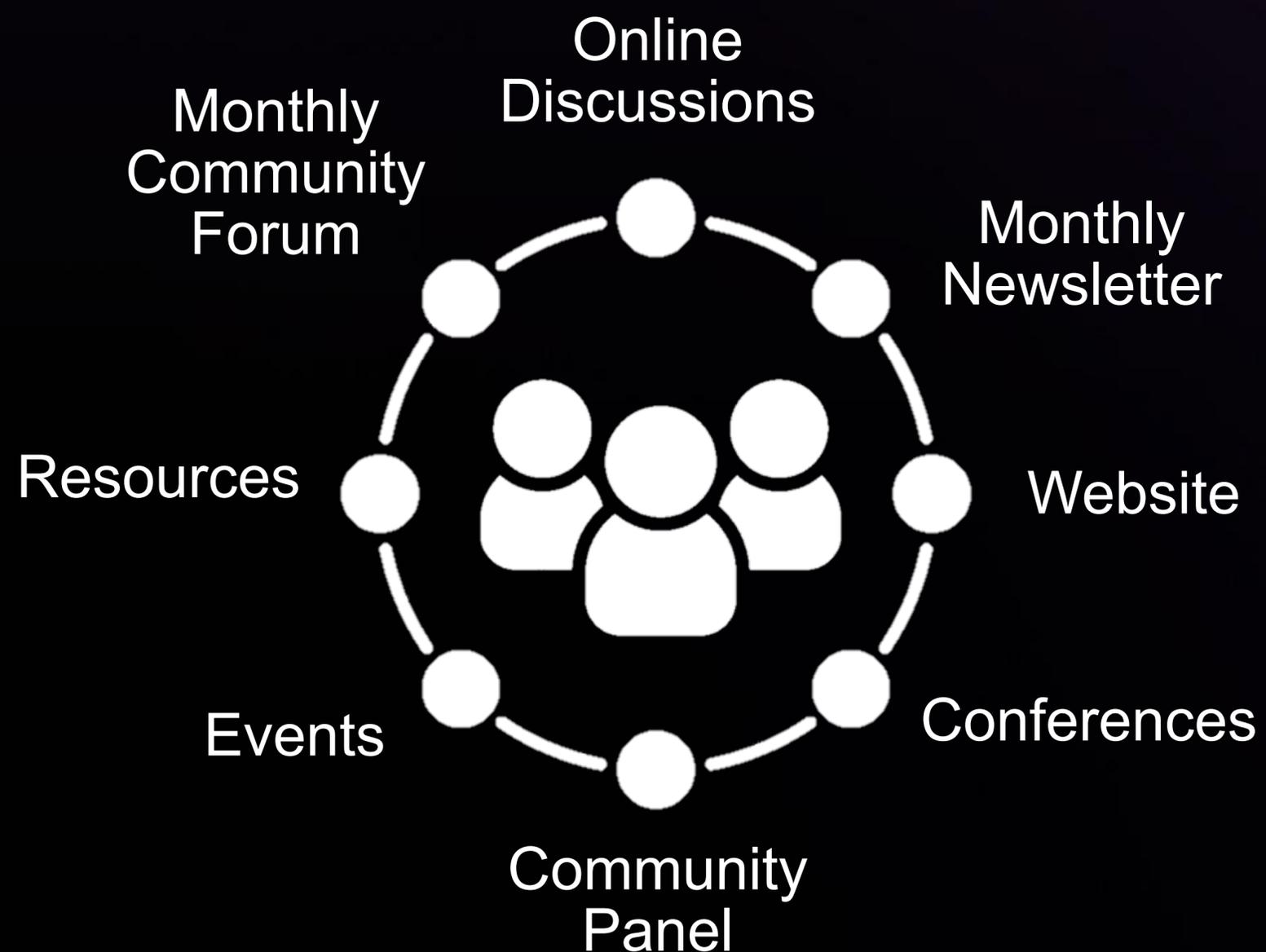
Coordination



TOPS Community Engagement

Community participation is the foundation of an open scientific process.

Listening, Learning, Collaborating, & Engaging



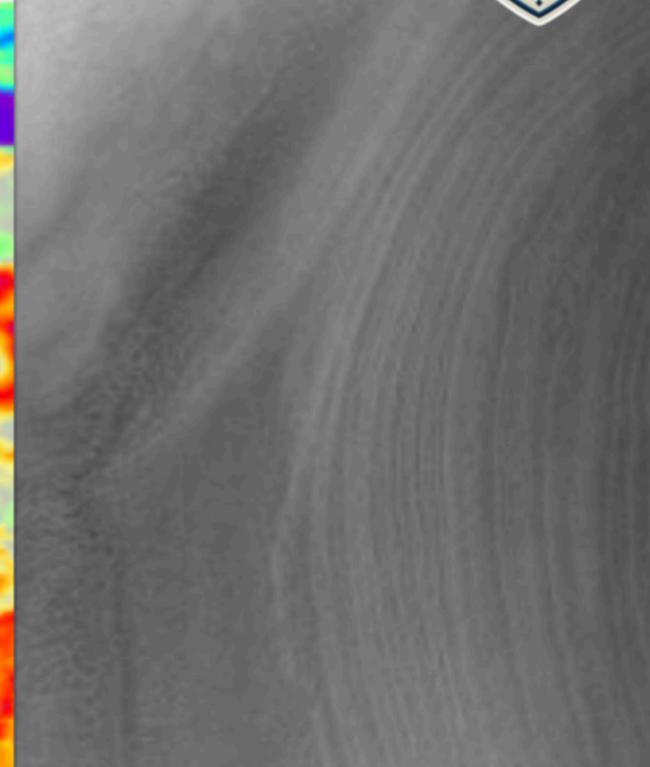
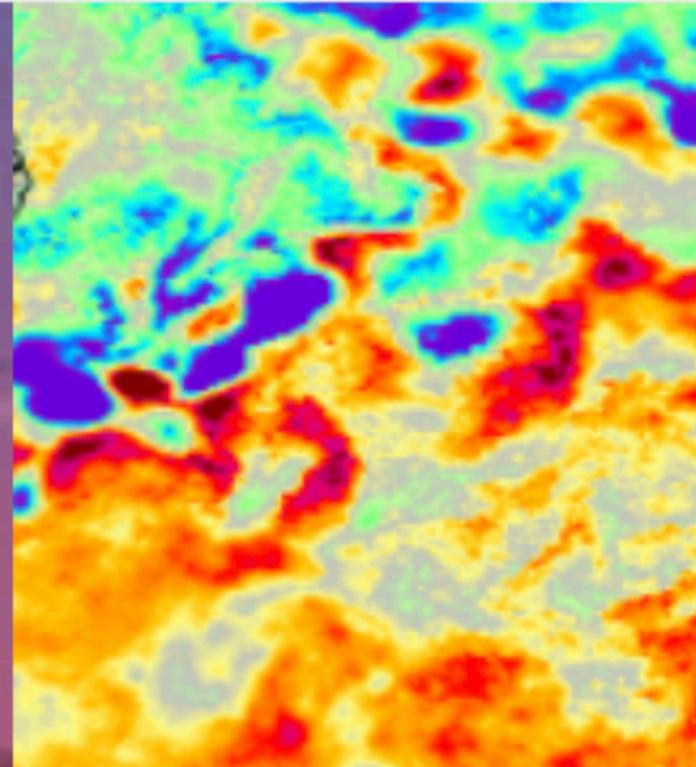
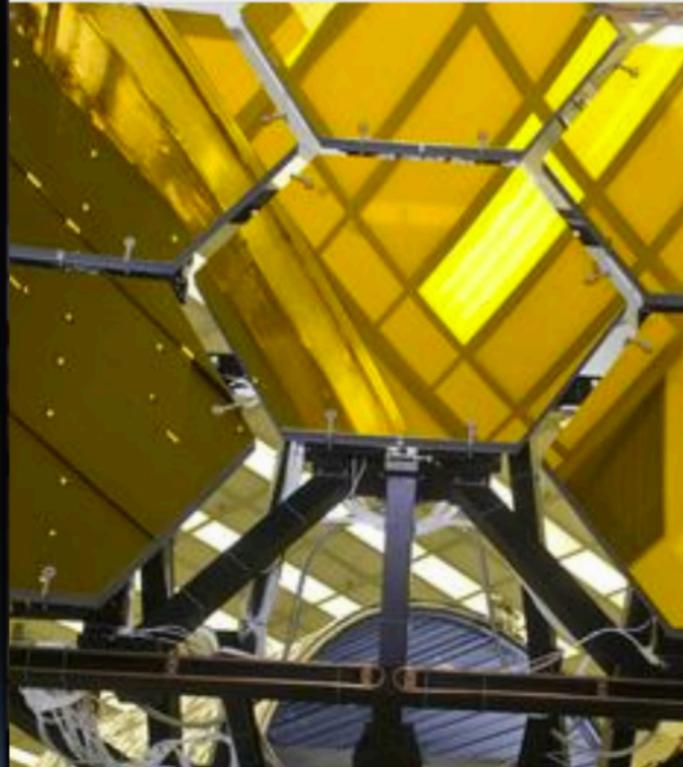
Open Science
Success Stories:



<https://zenodo.org/record/6994587#.ZG0IUOzMJoZ>



Open Science 101



**Ethos
of Open
Science**

**Open
Tools**

**Open
Software**

**Open
Data**

**Open
Results**

Images: JSC, JPL, & MSFC





Why take OS101?

Open Science 101: A **community-developed** introduction to open science with inclusivity, accessibility, and diversity at the forefront.

Designed to provide researchers with **core open science skills**, for example:

- How to write data/software management plans
- Know about open science tools and best practices (e.g. ORCID)
- Grow connections across a community of open science practitioners

Obtain a NASA Open Science badge!



Enroll now!

<https://bit.ly/EnrollOS101>



TOPS Incentives

It is important to recognize and incentivize the transition to Open Science



Awards

Create Open
Science Awards



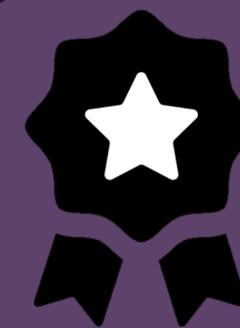
Badges

Recognize open
science skills with
digital badges



Prizes & Challenges

Science competition
prizes and
challenges that
incorporate open
science best
practices



Recognition

Open science
activities
recognized in
evaluations

Coordination



The White House announces 2023 A Year of Open Science



CDC ♦ DOA ♦ DOC ♦ DOE ♦ DOS ♦ DOT ♦ NASA ♦ NEH ♦ NIH ♦ NIST ♦ NOAA ♦ NSF ♦ SI ♦ USDA ♦ USGS

A multi-agency (15) initiative across the US Federal Government to spark change and inspire open science engagement through events and activities that will advance adoption of open science.

Website: <https://open.science.gov/>



THE OPEN SCIENCE LANDSCAPE (MY PERSPECTIVE)



1. How can an NSF FFRDC like NCAR participate in and support broader community open science efforts?
2. How can NCAR's HPC environment complement activities that are evolving on the commercial cloud?

WHAT IS NCAR? (MY PERSPECTIVE)

- Institute that has been and continues to be a world leader in the atmospheric and climate science communities
 - Supercomputing facilities
 - Develops and runs climate models
 - Employs many scientists (researchers, software engineers, etc.)

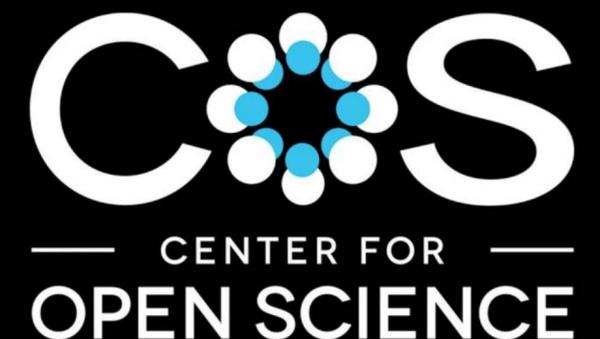
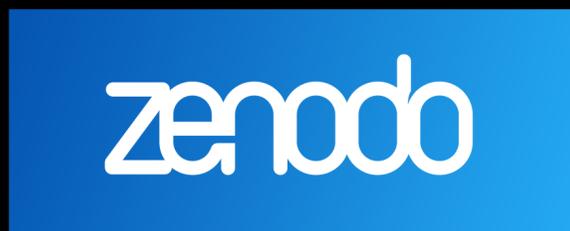


PUSH TOWARD OPEN SCIENCE

2023: A Year of Open Science

AGU23
WIDE. OPEN. SCIENCE.

Open
Source
Science



OPEN SCIENCE IS WHAT SCIENCE SET OUT TO BE,
BUT PAST SCIENCE CONSTRAINED

OPEN SCIENCE IS WHAT SCIENCE SET OUT TO BE, BUT PAST SCIENCE CONSTRAINED

Computing resources were few and far
between



<https://www2.cisl.ucar.edu/ncar-supercomputing-history>

1960 NCAR opens <https://ncar.ucar.edu/who-we-are/history>

1963 1st NCAR supercomputer - CDC 3600

OPEN SCIENCE IS WHAT SCIENCE SET OUT TO BE, BUT PAST SCIENCE CONSTRAINED

Computing resources were few and far
between

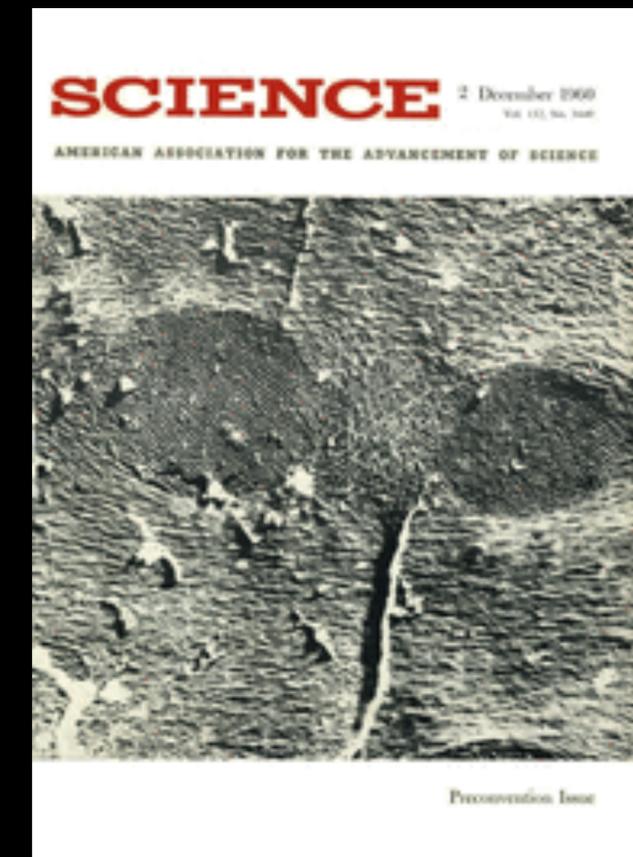
Scientific knowledge more
difficult to share



<https://www2.cisl.ucar.edu/ncar-supercomputing-history>

1960 NCAR opens <https://ncar.ucar.edu/who-we-are/history>

1963 1st NCAR supercomputer - CDC 3600



<https://www.science.org/loi/science/group/d1960.y1960>

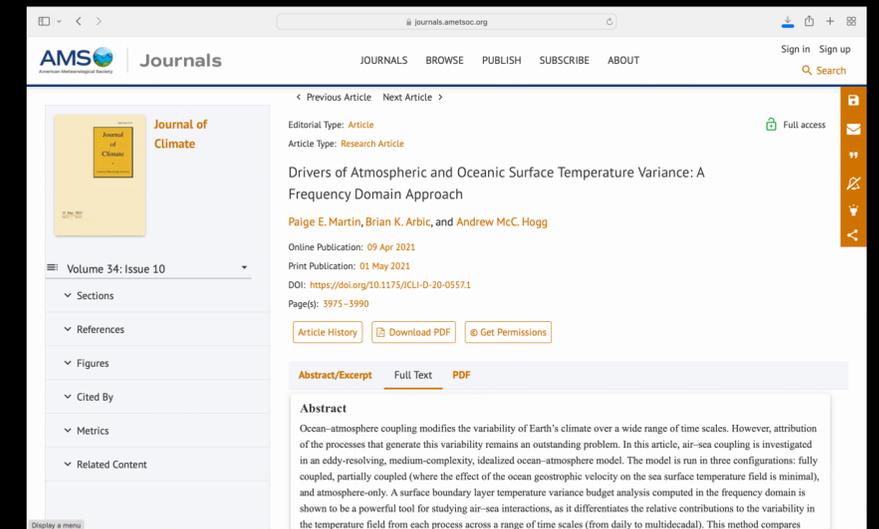
OPEN SCIENCE TODAY

Digital world!



<https://www.simplilearn.com/what-is-internet-article>

Less reliance on institutes



More computational science

Huge datasets

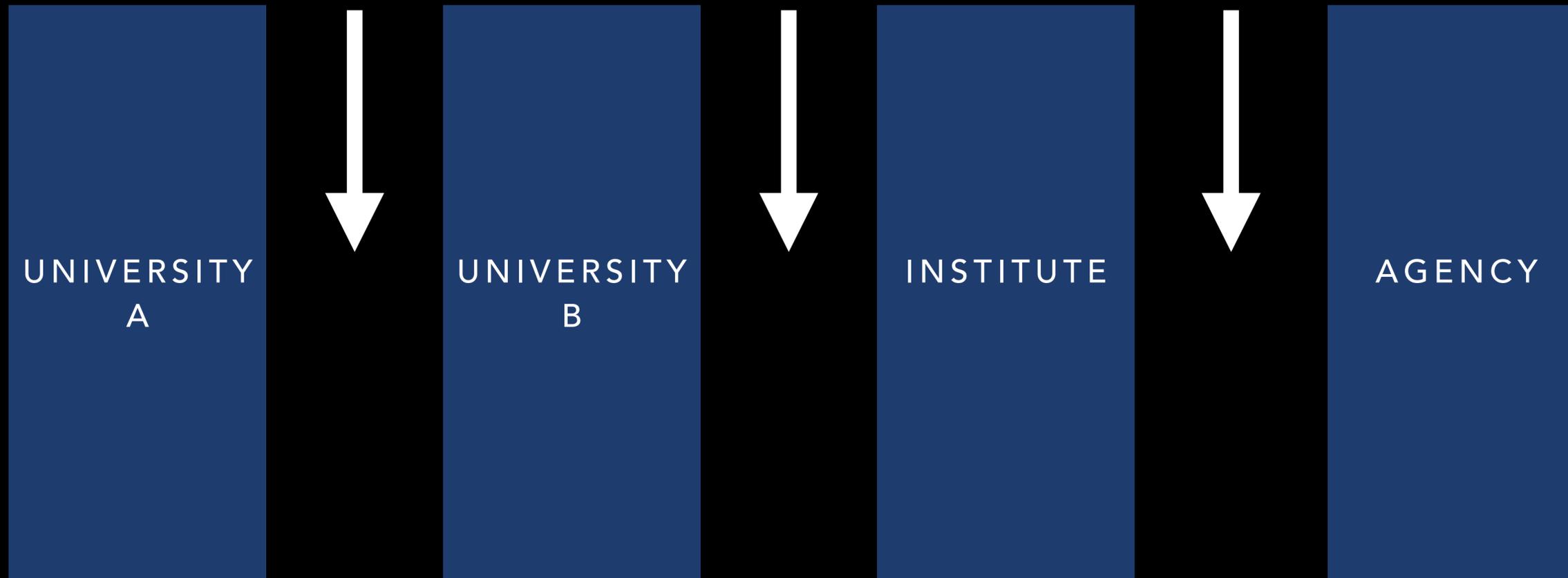


We are changing how we do science

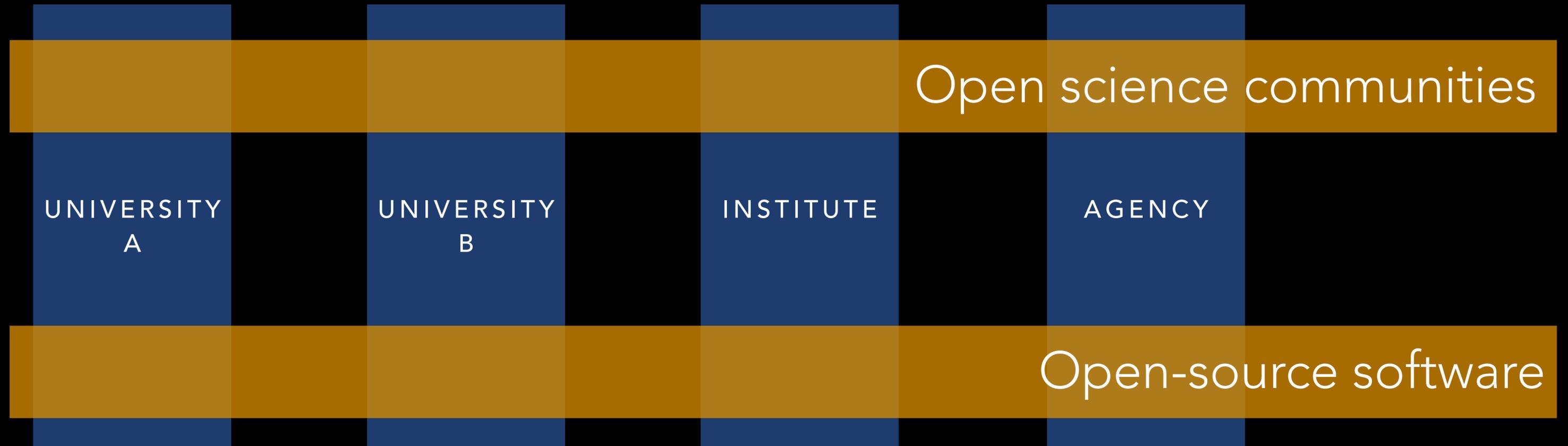


PARADIGM SHIFT IN HOW WE DO SCIENCE

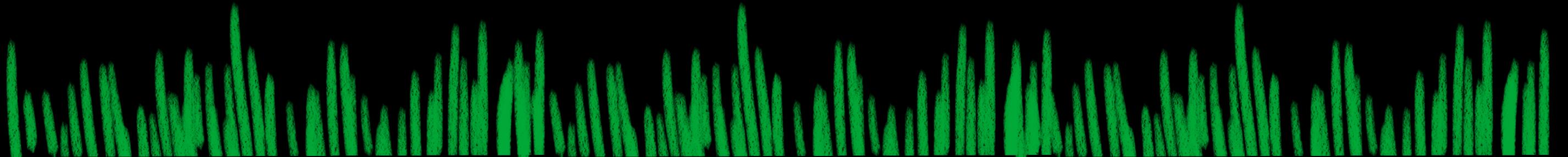
PARADIGM SHIFT IN HOW WE DO SCIENCE



PARADIGM SHIFT IN HOW WE DO SCIENCE



Grouped around: scientific discipline, structure of data, types of analysis



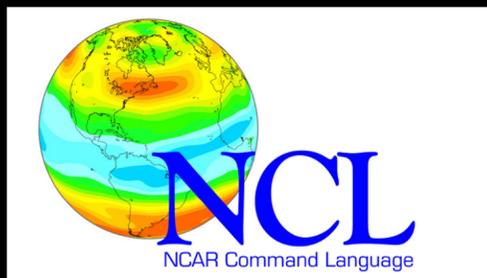
HOW CAN INSTITUTES SUPPORT THIS PARADIGM
SHIFT TOWARD OPEN SCIENCE?



It's hard!

HOW CAN INSTITUTES SUPPORT THIS PARADIGM SHIFT TOWARD OPEN SCIENCE?

NCAR ALREADY SUPPORTS
OPEN SCIENCE!



Climate Data Gateway
at NCAR

NCAR is sponsored by
National Science Foundation



HOW CAN INSTITUTES SUPPORT THIS PARADIGM
SHIFT TOWARD OPEN SCIENCE?

**PAY SOFTWARE ENGINEERS TO
CONTRIBUTE TO COMMUNITY TOOLS**

HOW CAN INSTITUTES SUPPORT THIS PARADIGM SHIFT TOWARD OPEN SCIENCE?

PAY SOFTWARE ENGINEERS TO CONTRIBUTE TO COMMUNITY TOOLS

- ✓ Direct positive impact on the scientific community
- ✓ Ensures long-term maintenance of community tools
- ✓ Builds expertise in community tools at the institute
- ✓ Encourages collaboration, rather than creation of new tools

HOW CAN INSTITUTES SUPPORT THIS PARADIGM
SHIFT TOWARD OPEN SCIENCE?

ACCOUNT FOR OPEN SCIENCE ACTIVITIES IN
REVIEWS AND EVALUATIONS
(E.G. HIRING, PROMOTION, AWARDS, ETC.)

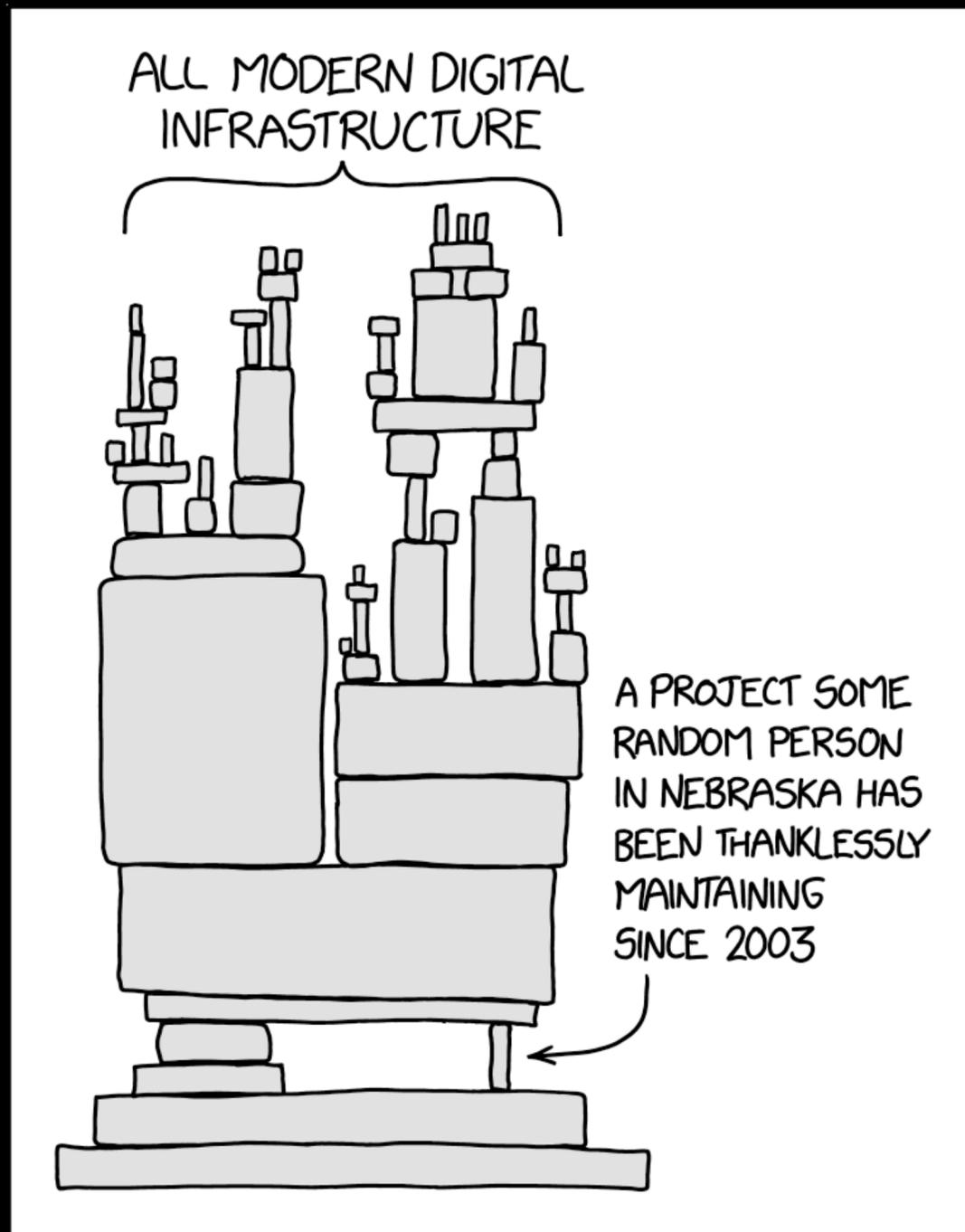
HOW CAN INSTITUTES SUPPORT THIS PARADIGM SHIFT TOWARD OPEN SCIENCE?

ACCOUNT FOR OPEN SCIENCE ACTIVITIES IN
REVIEWS AND EVALUATIONS
(E.G. HIRING, PROMOTION, AWARDS, ETC.)

- ✓ Attending community meetings
 - ✓ Taking the time to add metadata before sharing data
 - ✓ Documenting your code
 - ✓ Fixing bugs in a widely used software package

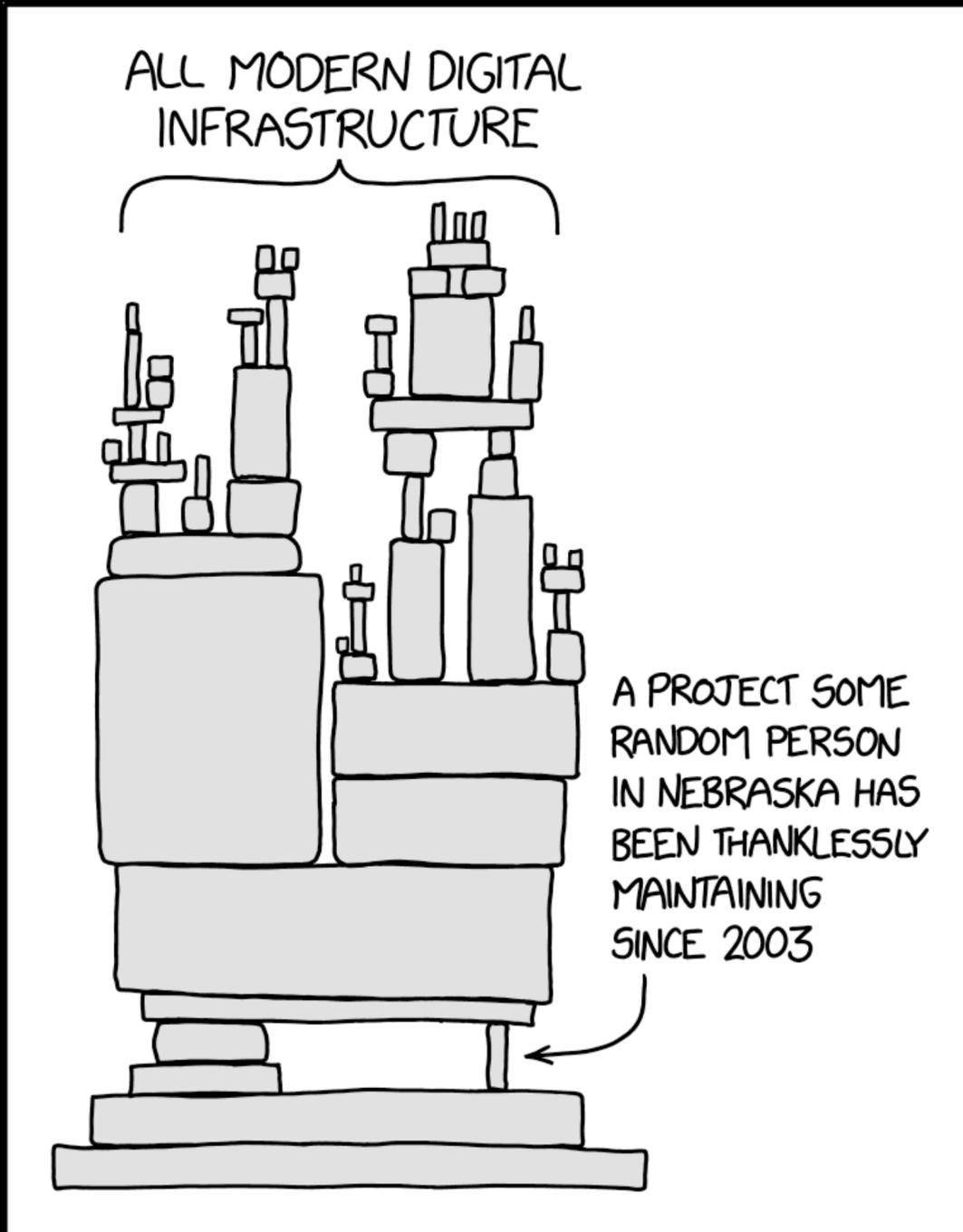
CONCERNS & CHALLENGES FOR INVESTING IN COMMUNITY SOFTWARE

CONCERNS & CHALLENGES FOR INVESTING IN COMMUNITY SOFTWARE



IT'S TOO RISKY - SUSTAINABILITY OF OPEN-SOURCE SOFTWARE IS OFTEN AN OPEN QUESTION

CONCERNS & CHALLENGES FOR INVESTING IN COMMUNITY SOFTWARE



IT'S TOO RISKY - SUSTAINABILITY OF OPEN-SOURCE SOFTWARE IS OFTEN AN OPEN QUESTION

Institutes can play a crucial role in "de-risking" by contributing sustained resources to community-accepted tools

CONCERNS & CHALLENGES FOR INVESTING IN COMMUNITY SOFTWARE

COMMUNITY TOOLS MAY NOT BE OPTIMIZED OR BEST-SUITED
FOR AN INSTITUTE'S INFRASTRUCTURE/DATASETS

CONCERNS & CHALLENGES FOR INVESTING IN COMMUNITY SOFTWARE

COMMUNITY TOOLS MAY NOT BE OPTIMIZED OR BEST-SUITED
FOR AN INSTITUTE'S INFRASTRUCTURE/DATASETS

INTEROPERABILITY IS KEY

CAN I ADD A NEW FEATURE TO AN
EXISTING TOOL?

WHAT TOOL CAN I BUILD THAT BEST
INTEGRATES WITH THE COMMUNITY-
ADOPTED SOFTWARE STACK?

CONCERNS & CHALLENGES FOR INVESTING IN COMMUNITY SOFTWARE

FUNDING VALUES NOVELTY

Newly
developed tool



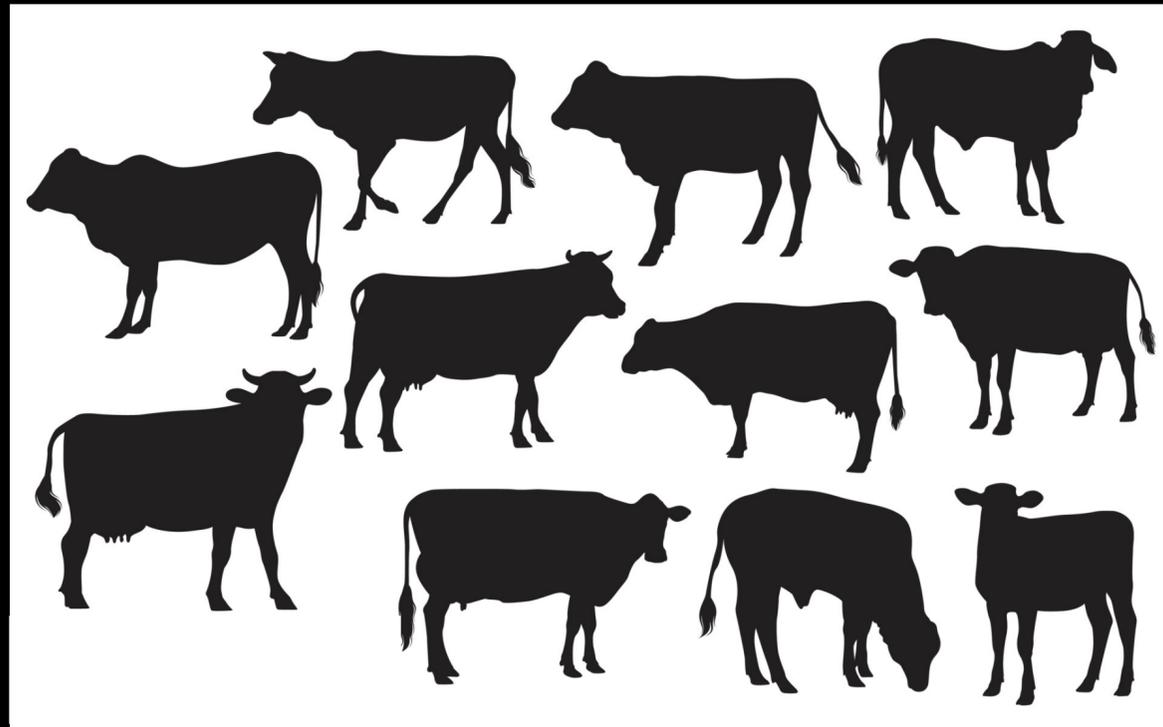
Contributions to
existing tools



HOW CAN NCAR'S HPC SYSTEM SUPPORT THE
MIGRATION OF SCIENCE TOWARD THE CLOUD?

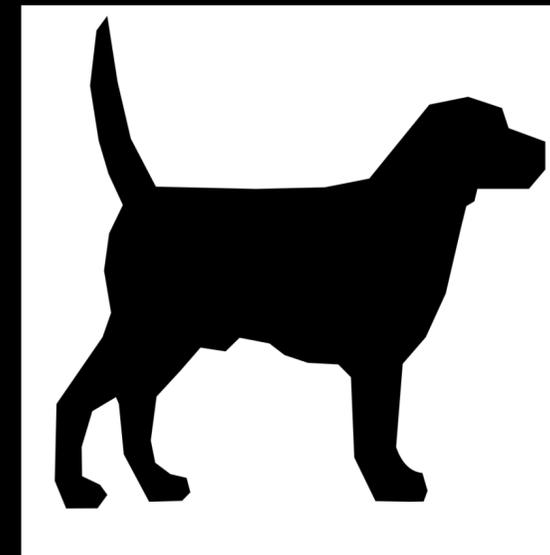
HOW CAN NCAR'S HPC ENVIRONMENT COMPLEMENT ACTIVITIES THAT ARE EVOLVING ON THE CLOUD?

Cloud



[vecteezy.com](https://www.vecteezy.com)

HPC



[https://commons.wikimedia.org/wiki/
File:Dog_silhouette.svg](https://commons.wikimedia.org/wiki/File:Dog_silhouette.svg)

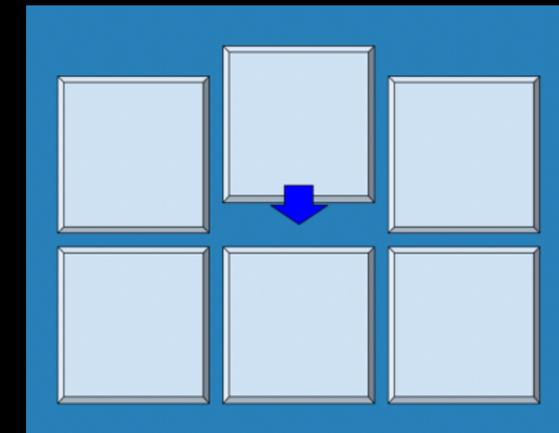
HOW CAN NCAR'S HPC ENVIRONMENT COMPLEMENT ACTIVITIES THAT ARE EVOLVING ON THE CLOUD?

MAKE DATA ACCESSIBLE FROM THE CLOUD



A distributed storage cloud for the research community

- Stores many Pangeo datasets at a single facility
- Performant for Cloud-based workflows



kerchunk

- Allows chunked data to act like cloud-optimized Zarr stores
- Does not alter the underlying files

HOW CAN NCAR'S HPC ENVIRONMENT COMPLEMENT ACTIVITIES THAT ARE EVOLVING ON THE CLOUD?

LEVERAGE BENEFITS OF HPC

More robust and tailored environment

- Good for running large models
- Good for running specialized workflows

Lower cost

- Good for very long runs
- Good for workflows that don't need to scale up and down



Keep contributing, building,
and sustaining your code.



Thank you for your
contributions.



National Aeronautics and
Space Administration

Thank You!

Thank you for the discussions: Tom Nicholas,
Deepak Cherian, Demitri Muna, JL Ganache, Julius
Busecke, Ryan Abernathey

PAIGE E. MARTIN
paige.e.martin@nasa.gov
paigem.github.io