

# Open Science by Design

*WGISS October 19, 2021*

**Katie Baynes**

Deputy Program Executive

Earth Science Data Systems (ESDS)

NASA HQ



# Open Science

“We define open science as a **collaborative culture** enabled by technology that empowers the **open sharing of data, information, and knowledge** within the scientific community and the wider public to accelerate scientific research and understanding.”

2

# Why “Open-Source” Science?

*Builds on concepts from Open Source Software revolution that expanded participation in developing code and applies it to the scientific process to accelerate discovery by openly conducting science from project initiation through implementation.*



# WHAT IS “OPEN” ABOUT OPEN-SOURCE SCIENCE?

## The Four Meanings of “Open” in Open-source Science



### OPEN (**TRANSPARENT**) SCIENCE

scientific process and results should be visible, accessible, and understandable



### OPEN (**INCLUSIVE**) SCIENCE

process and participants should welcome participation by and collaboration with diverse people and organizations

### OPEN (**ACCESSIBLE**) SCIENCE

data, tools, software, documentation, and publications should be accessible to all (FAIR)



### OPEN (**REPRODUCIBLE**) SCIENCE

scientific process and results should be open such that they are reproducible by members of the community



# Open-Source Science Policy for Earth System Observatory

- A. All mission data, metadata, software, databases, publications, and documentation shall be available on a full, free, open, and unrestricted basis starting in Phase B with no period of exclusive access.
- B. Science workshops and meetings shall be open to broad participation and documented in public repositories.

**1** Software shall be developed openly in a publicly accessible, version-controlled platform using a permissive software license allowing for community use and contributions.

**2** Manuscripts shall be published with open access licenses; versions of as-accepted manuscripts shall be made available as open preprints and deposited in a NASA or [Partner] repository upon publication.

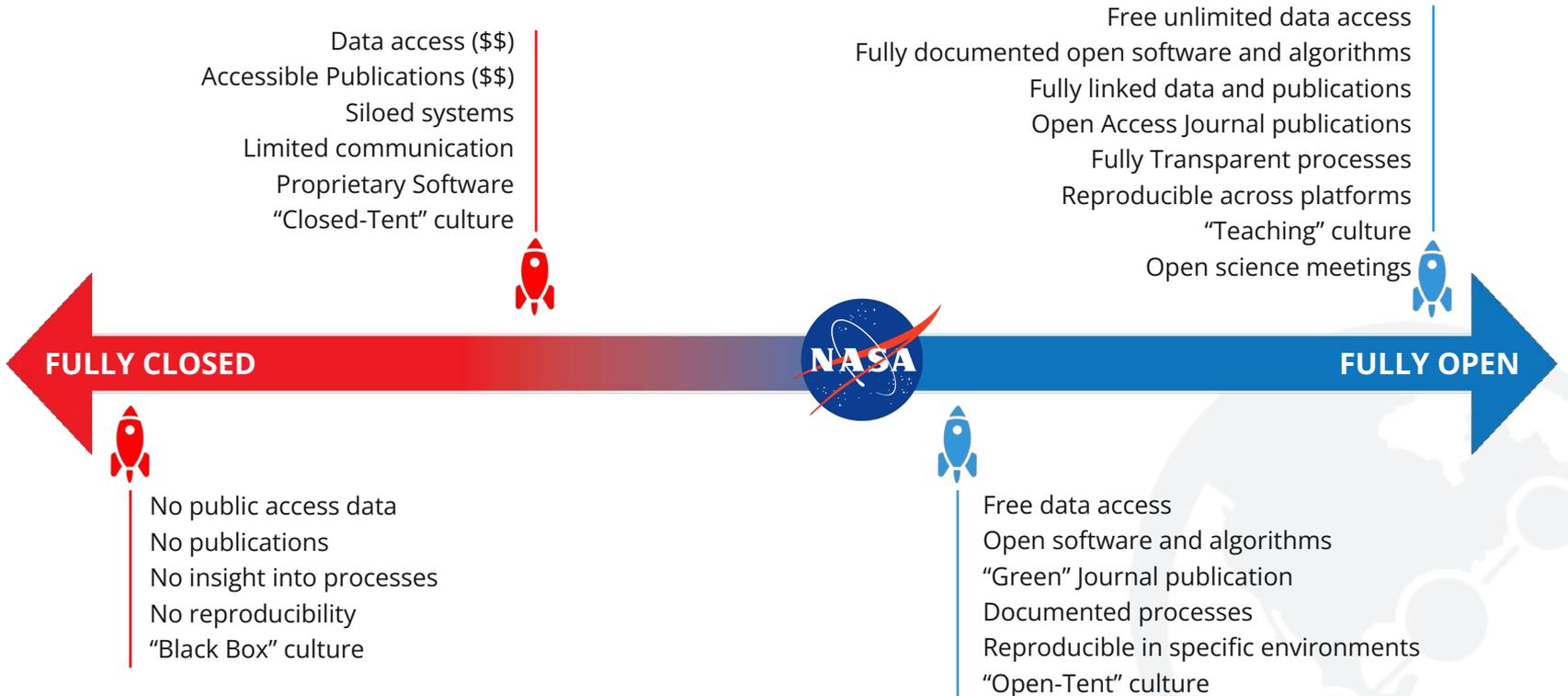
**3** All mission data, calibration information, and simulated products supporting development and validation of algorithms shall be made available without any conditions to use.

**4** Scientific data, metadata, software, publications and documentation shall be archived and made available by NASA and/or [Partner] starting in Phase B.

**5** NASA and [Partner] software, documentation and data shall be properly marked, cited, and/or attributed. Metrics to measure and acknowledge open-source science contributions will be developed.

**6** NASA and [Partner] will mutually develop an Open-Source Science Plan that specifies details of collaboration.

# A CONTINUUM OF OPEN-SOURCE SCIENCE



An aerial photograph of a forested island in a lake. The island is covered in dense green trees and is surrounded by dark water. The sky is dark and overcast.

# Building an Open Sourced Science Ecosystem

- Initiate new missions, research and applied activities as open science projects.
- Implement clear policies for software, publications and data.
- Integrate and improve existing capabilities to support data management, access, computing, analytics and collaboration.
- Build the community through training, workshops, competitions and incentives.

**Multi-Mission Algorithm and Analysis Platform (MAAP)**

Science-focused, cloud-based environment to discover, process, analyze, and share NASA and ESA data

**Interactive Storytelling Platform**

Interactive exploration, kick-started by NASA's COVID-19 dashboard activities

**Common Metadata Repository**

Open source, cloud-native, super fast Earth Science catalog and discovery

**Earthdata Publication Tool**

Centralized authoring and management of NASA Earth Science new products

**Distribution and Analysis Platform(s)**

Algorithm development and data production

Models and Model Processing

**Policies**

Open source, open data, articles

**Algorithm Publication Tool**

Centralized authoring and publishing and discovery tool for NASA Earth Science ATBDs

**Advanced Metrics Collection**

Cloud-based configurable ingest and archive metrics tracking

**Multi-mission Data Lake**

**Cumulus**

Open source, cloud-native, reusable Ingest and Archive Workflow System

**NGAP** Security, backup, cost controls, scalability

Operational

Active Development

Formulation



# **Beyond Technology and towards Community Engagement**



**Transform to OPen Science (TOPS)** is a 5-year effort focused on capacity building, partner engagement, and incentives to help accelerate scientific discovery through open science.

## TOPS Focus Areas

### Public Engagement

- **Designate 2023 as Year of Open Science**
- Partnering with professional orgs., publishing TOPS articles in high-impact journals
- Engage early with historically excluded communities
- TOPS GitHub - <https://github.com/nasa/Transform-to-Open-Science>

### Capacity Building

- Create FAIR - Analysis-Ready Cloud-Optimized (ARCO) data
- Develop learning resources
- TOPS JupyterHub
- Host and sponsor events (summer schools, multi-day trainings, massive open online courses)

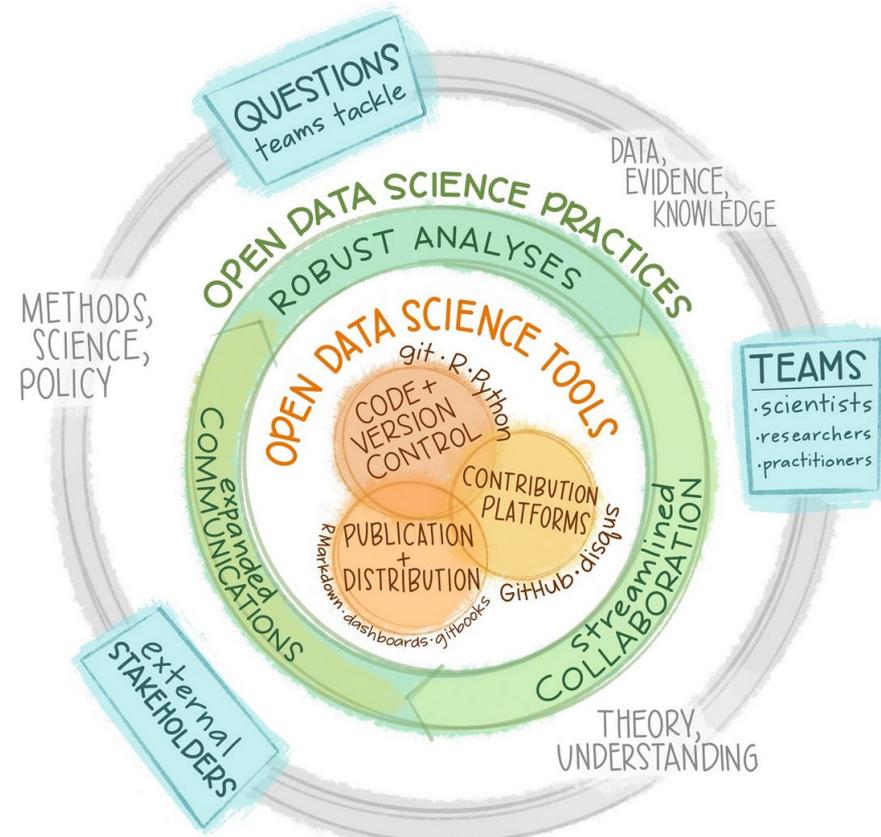
### Incentives

- Develop NASA Open Source Science Awards program
- Leverage prizes and challenges and cross-division science use cases
- Increased citizen science activities

# Building an Open Science Aware Community

## Building Internal Capacity

- Openscapes initiative to “train the trainers”:  
<https://nasa-openscapes.github.io/>
- Examining policy to encourage publication in Green/Gold Journals





## In Summary

We have a strong vision for the future and have been making strides towards open source science within the ESDS program and NASA Science more broadly.

For upcoming missions, we will build-in open science principles at project initiation to tackle common challenges.

We are excited to continue to **build a community** dedicated to transparency, inclusivity, accessibility, and reproducibility.