



# NASA's SMD and Open Source Science

Katie Baynes

[kathleen.baynes@nasa.gov](mailto:kathleen.baynes@nasa.gov)

Deputy Program Manager  
Earth Science Data Systems

May 18, 2021

# A Vision for Open Science

*Expand participation, improve reproducibility, and accelerate scientific discovery for societal benefit.*

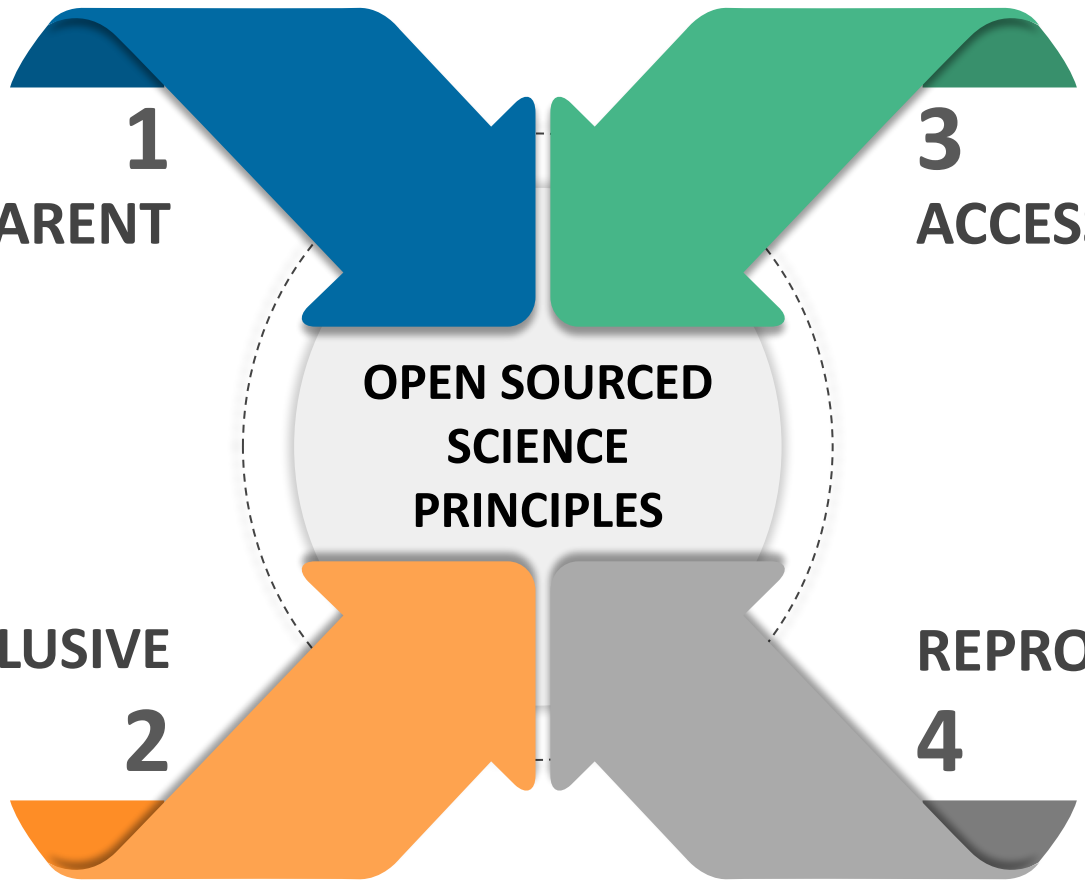
**1**  
**TRANSPARENT**

**3**  
**ACCESSIBLE**

**OPEN SOURCED  
SCIENCE  
PRINCIPLES**

**INCLUSIVE**  
**2**

**REPRODUCIBLE**  
**4**



# A Continuum of Open Sourced Science

- Data access (\$\$)
- Accessible publications (\$\$)
- Siloed systems
- Limited Communication
- Proprietary Software
- “Closed-tent” culture

- Free unlimited data access
- Fully documented open software and algorithms
- Fully linked data and publications
- Open Access Journal Publications
- Fully Transparent Processes
- Reproducible across platforms
- “Teaching and Learning” Culture



**Fully Closed**

**Fully Open**

- No public data access
- No publications
- No insight into processes
- No reproducibility
- “Black Box” Culture

- Free data access
- Open software and algorithms
- “Green” Journal Publication
- Documented Processes
- Reproducible in specific environments
- “Open-Tent” Culture

### 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

*Starting with S6 Michael Freilich, all future data access and storage will be cloud-native, including SWOT and NISAR and will have access to these capabilities*

# Strategic Partnerships to further Open Science Goal

## Open Science Drivers

Accessibility to the scientific process for non-scientists  
Broader, and new perspectives and techniques to tackle challenging problems

## Activities

### Space Act Agreements

Amazon Web Services (AWS) and Google (active)  
IBM, Microsoft and NVIDIA (in works)

### Interagency collaborations

## Recent Successes

### AWS Data Sandbox

cloud credits for open science workshops to support the research and applications community

### Formulation of SpaceML Collaboration

machine learning operations (MLOps) components and workflows that can be utilized by Earth science tools

### NOAA

prototyping NASA data system software and collaboration on joint data lake



Google Cloud Platform



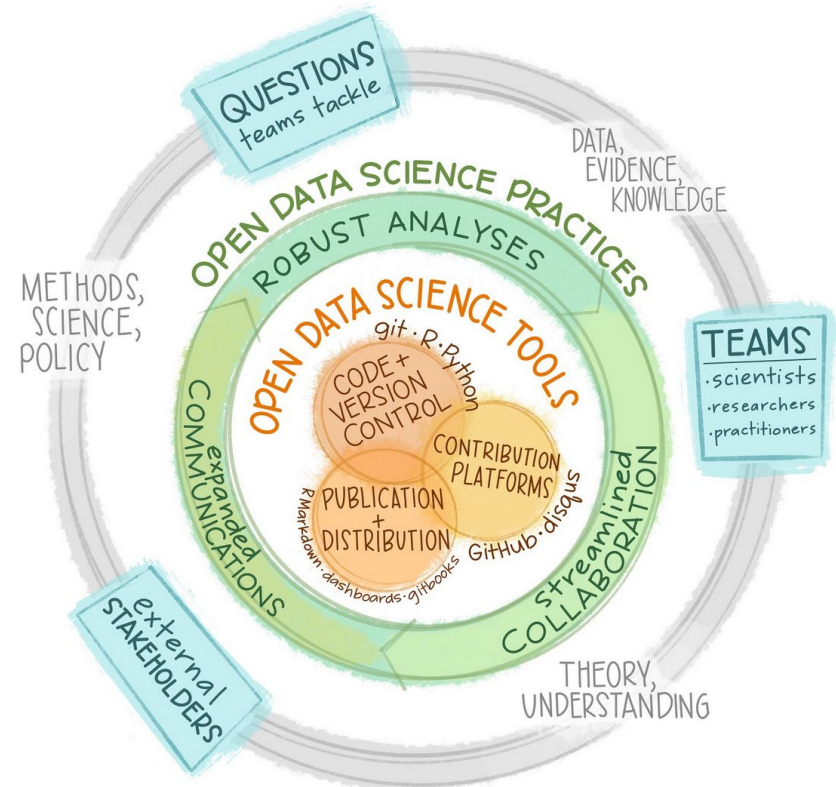
# Building an Open Science Aware Community

## Externally Focused Investments

- Targeted Workshops and Hackathons
- Engaging next generation of Data and Earth Scientists (FDL, SpaceML, Radiant Earth)
- Publishing about Open Science in the Community
- Updating messaging from communications

## Building Internal Capacity

- Openscapes: utilizes “train the trainers” paradigm to teach open science principles and techniques
- Examining policy to encourage publication in Green/Gold Journals



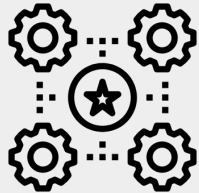


# Open Science for the Next Generation of NASA Missions



## Cross DO Study: Data Processing

Developing concepts for a common science data processing system for product generation (L1 - 4).



Study facilitated and funded by Data Systems Program

### Approach

Initiate an architecture concepts and prototyping study co-lead by JPL and GSFC with support from DO teams.

### Expected DO Commitment

.5 WYE and 3 workshops over 12 month period

### Deliverable

Architectural options for an open common science data processing system for L1 - 4 products that identifies risks, incentives and potential partnerships.

Study facilitated and funded by ESDS program

### Approach

Initiate hardware and ground system architecture study to minimize data latency and support cross-DO science. Emphasis on identification of approaches for downlink, networking, and L0 generation using in-house and commercial capabilities.

### Expected DO Commitment

.25 WYE

### Deliverable

Architectural options including costs and risks to reduce data product latency.

## Cross DO Study: Data Latency

Evaluating flight hardware and ground system architectures to minimize product latency and support cross-DO science product generation.



An aerial photograph of a coastline. The left side shows brown, textured land with intricate patterns of waterways and inlets. The right side shows a vast expanse of dark blue ocean. The text "Thank You" is centered in the middle of the image.

Thank You