NASA's SMD and Open Source Science

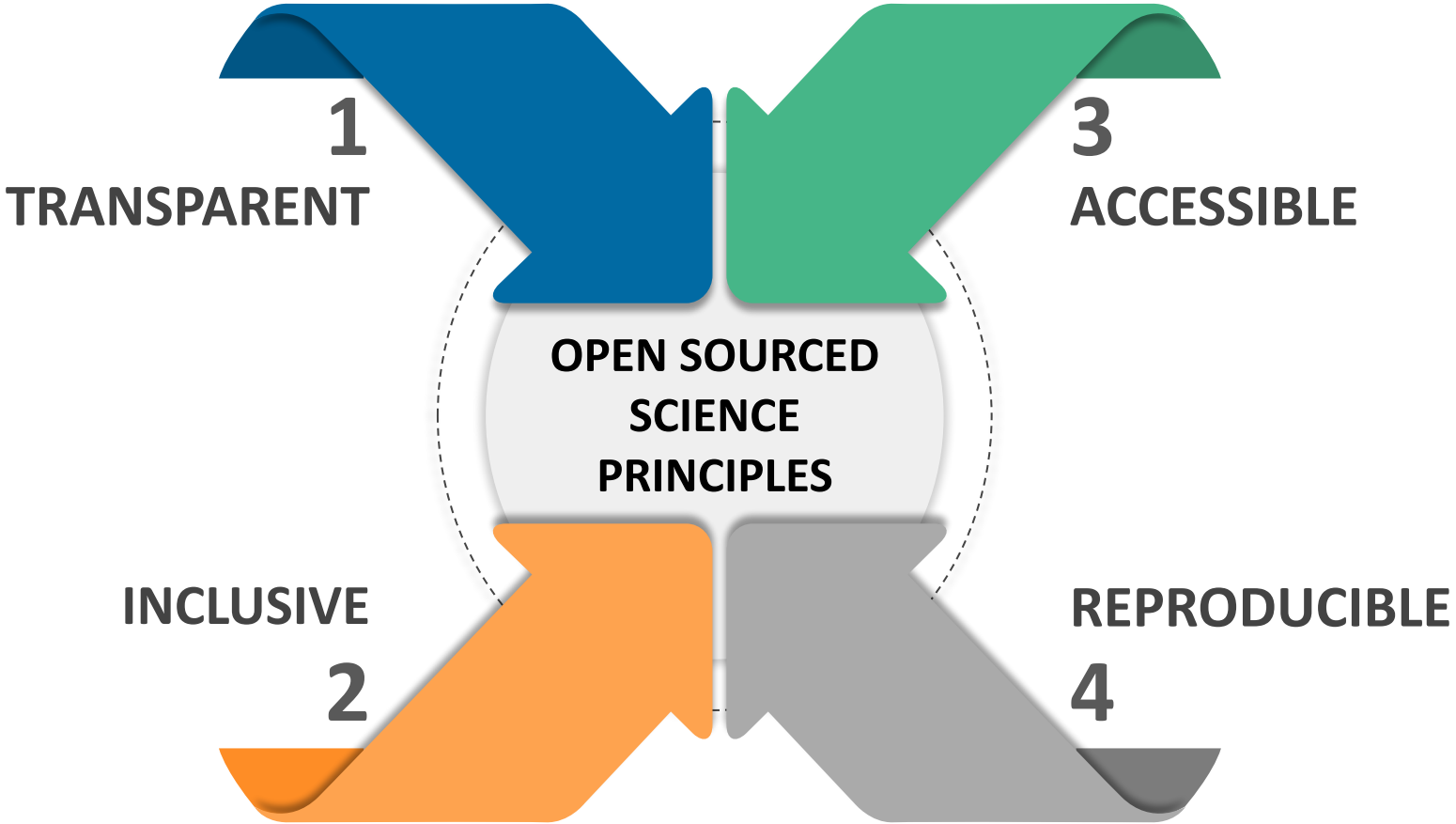
Katie Baynes
kathleen.baynes@nasa.gov
Deputy Program Manager
Earth Science Data Systems
May 18, 2021

https://unsplash.com/photos/6-jTZysYY_U

A Vision for Open Science

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

https://unsplash.com/photos/6-jTZysYY_U


1
TRANSPARENT
3
ACCESSIBLE
OPEN SOURCED
SCIENCE
PRINCIPLES
INCLUSIVE
2
REPRODUCIBLE
4

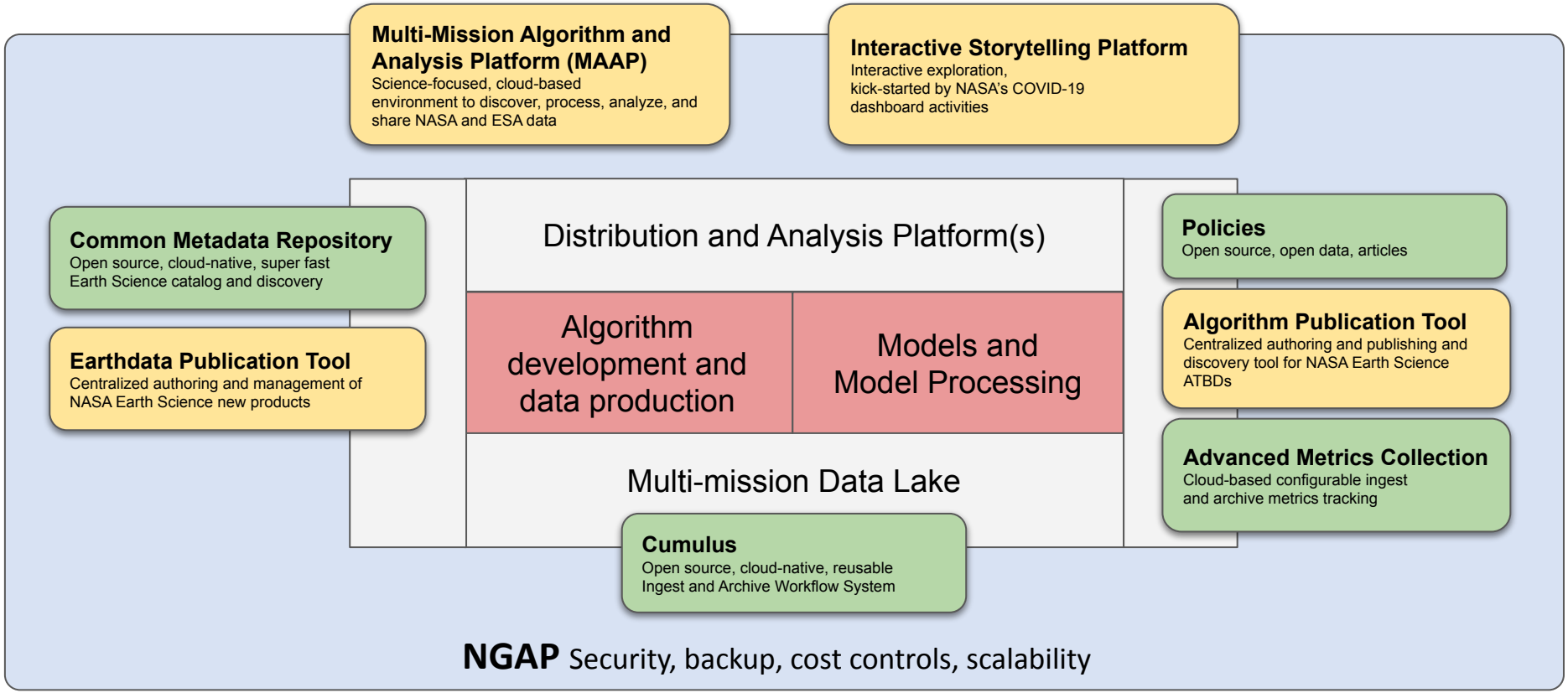
A Continuum of Open Sourced Science

- 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

- 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



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)

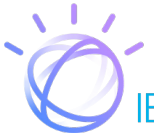
Interagency collaborations

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



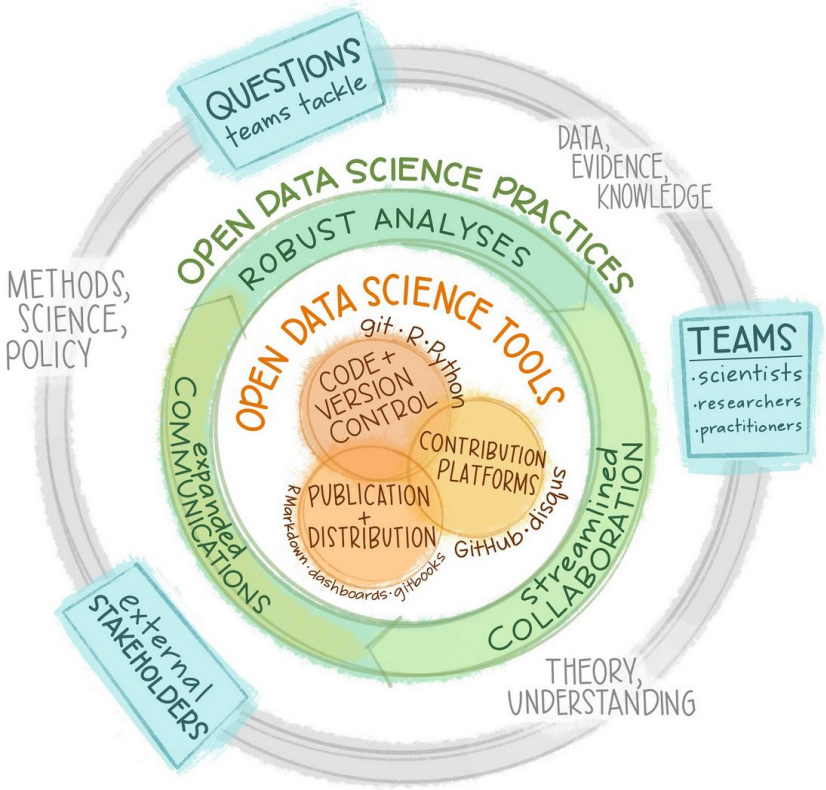

Building an Open Science Aware Community

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

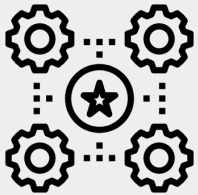
- Openscapes: utilizes “train the trainers” paradigm to teach open science principles and techniques
- Examining policy to encourage publication in Green/Gold Journals



https://unsplash.com/photos/6-jTZysYY_U

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

<u>Approach</u>
Initiate an architecture concepts and prototyping study co-lead by JPL and GSFC with support from DO teams.

<u>Expected DO Commitment</u>
.5 WYE and 3 workshops over 12 month period

<u>Deliverable</u>
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.

https://unsplash.com/photos/6-jTZysYY_U