

National Aeronautics and
Space Administration



Open -Source Science at NASA

Kevin Murphy
Chief Science Data Officer
October 19, 2022

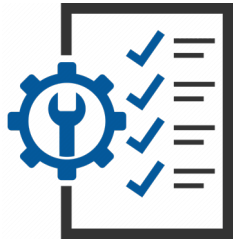


Overview

- NASA's Open Science policies and initiatives
- SMD Strategy for Data and Computing
- SPD-41
- Open Science Tools
- SMD Core Data and Computing Services
- Open Science in Earth Science
- Transform to Open Science (TOPS)

Open -Source Science Initiative

Unlocking the full potential of a more equitable, impactful, efficient, scientific future



Policy development,
education, compliance tools
Updating NASA policies on
scientific information to better
enable the activation of open
science



Core Services for Science
Discovery
Developing core data and computing
services to enable open science



ROSES Elements
Supporting open-source
software, tools, frameworks,
libraries, platforms, and training
with over \$5 million dollars in
grants

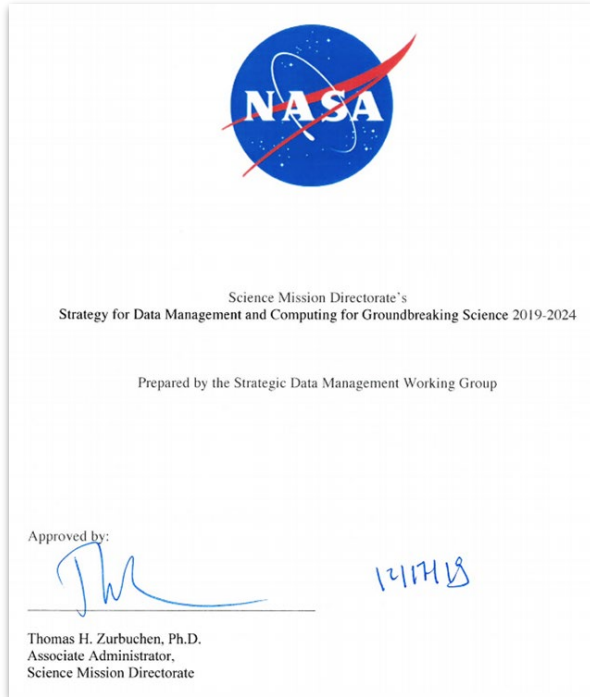


Community Building &
Partnerships - Transform to Open
Science (TOPS)
Accelerating adoption of open
science

Core Values

- As open as possible, as restricted as necessary, always secure
- Increase the accessibility, inclusion, and reproducibility of SMD scientific activities
- When possible, minimize the burden

What is the SMD Strategy for Data and Computing?



An SMD-approved strategy to enable transformational open science through continuous evolution of SMD's science data and computing systems.

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



SPD-41: Scientific Information Policy

SPD-41 was released in August 2021.

SPD-41 brings together existing NASA and Federal guidance.

- SPD-41: The Science Information Policy - <https://go.usa.gov/xtNTJ>
- Science Information Policy Website - <https://go.usa.gov/xtNTt>



SPD-41a was released in November with proposed additions. An RFI was released to the community and closed on March 4, 2022.

SPD-41a Policy (updates to SPD-41)

Data

Scientific data **should be FAIR** and shall be made publicly available with a clear, open, and accessible data license no later than the publication of the research, **and be citable** .

Mission data shall be openly available with no period of exclusive access.

Software

Research software **shall** be publicly available no later than the publication of the research, assigned a permissive software license, **and be citable** .

Mission software shall additionally be developed openly in a publicly accessible, version-controlled platform that allows for contributions and engagement from the community.

Publications

Manuscripts versions of as -accepted manuscripts shall be deposited in a NASA repository and made publicly available within 12 -months. **Publishing as open access is supported and posting preprints is encouraged.**

Mission publications shall additionally be made publicly available at the time of their publication.

Science workshops and meetings shall be open to broad participation and documented in public repositories.

SPD-41a Implementation

Implementation plan will be developed by CY2023 and include:

- Software release policy and update to NASA processes for software release
- Guidance for awards, contracts, ROSES, and Announcement of Opportunities; PIs should include these costs in proposals
- Incorporated text into AOs
- Incentives for the community to make the transition - e.g ROSES22 F8. Supplement for Open Source Software

SPD-41a is *forward looking* - it is meant to apply to work going forward. Existing missions and investigations should adopt parts of this policy consistent with available resources.

Supporting Open Source Science

Supporting open source software, open data, and open science practices in the SMD community.

- ROSES20 E.7 Open Source Tools, Frameworks, and Libraries selected 16 proposals supporting 22 different projects.



Full description of supported projects is available on [NSPIRES](#)

Open ROSES22 Elements

F.2 [Topical workshops, symposium, and conferences](#)

Events designed to expand SMD open science capabilities including open science practices; hackathons, data challenges, or un-conferences; or training related to open science. This element has a rolling deadline.

F.8 [Supplemental Open Source Software Awards](#)

Supplemental award to modernize software and release as open source. This element has a rolling deadline.

F.14 [Transform to Open Science Training](#)

Development of training material, summer schools, and virtual cohorts to advance open science literacy for TOPS. **Opened on September 9, closes Dec 8**

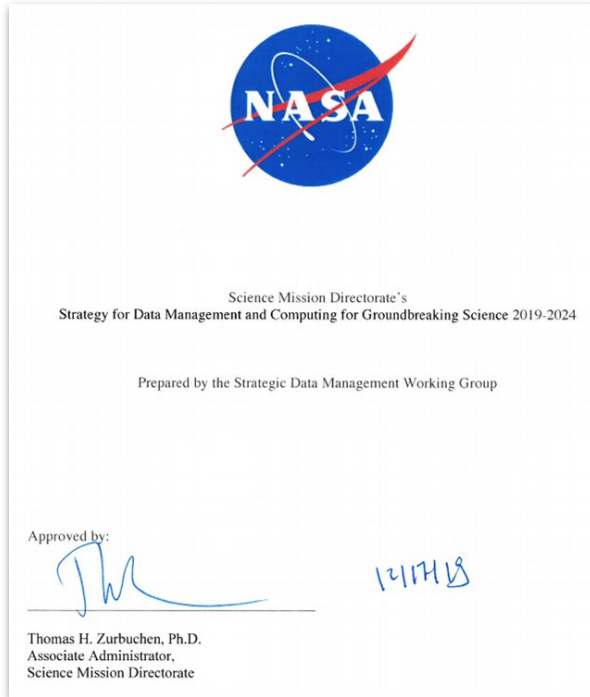
F.19 [Multidomain Reusable Artificial Intelligence Tools](#)

Enabling critically needed machine learning tools to advance Heliophysics and Earth Science research **Opened on October 4, closes Jan 13**

Coming Soon:

Solicitations to support 'F.15 High Priority Open Source Science' and 'F.16 Supplement for scientific software platforms.'

What is the SMD Strategy for Data and Computing?



An SMD-approved strategy to enable transformational open science through continuous evolution of SMD's science data and computing systems.

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



Activities for Core Services

The CSDO is conducting two activities to develop cyberinfrastructure to support the Strategy for Data Management and Computing and SPD41:

1. Defining Core Data and Computing Services Requirements

Common SMD IT policies, software and computing capabilities to support:

- Moving to hybrid cloud environments: computing, storage, cybersecurity, networking, and business processes
- Open-Source Science/SPD-41 requirements: Research Data and Software Archive, User Registration, Data Set Search, Journal Search, AI/ML models, and more

2. Data and Computing Architecture Study

- Study to evaluate architecture options for scientific data and computing elements of Core Services infrastructure.
- Produce recommendations for a Hybrid Cloud Infrastructure for SMD (mixed computing, storage, and services environment made up of on-premises infrastructure, private cloud services, high-end computing, and a public cloud)

Core Services funding initiates in FY24 and ramps up fully in FY25.

**Earth Science Division is
the Pathfinder for Open
Source Science.**



Exemplar: Open -Source Science Policy for Earth Science Missions

- 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.

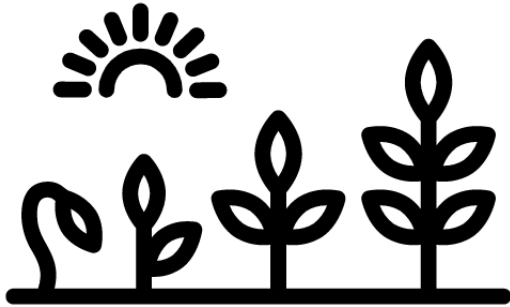
5 NASA and [Partner] software, documentation and data shall be properly marked, cited, and/or attributed . Metrics to measure and acknowledge open -source science contributions will be developed.

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

Collaborative, accessible, inclusive, transparent, and reproducible from the beginning.

Evolving **ESDS** to support Open Science in ESO Era

How can we adapt existing infrastructure to better serve users?



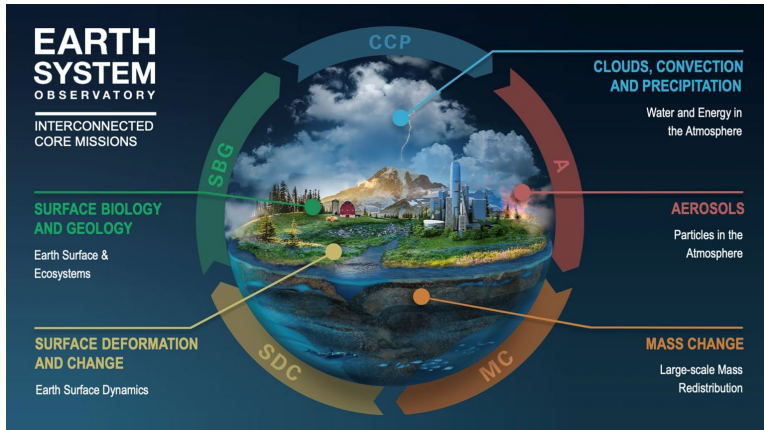
- Migrate Data to the **Cloud Environment**
- Prepare for the **ESO Era**
- **Community engagement**

Preparing for the ESO Era

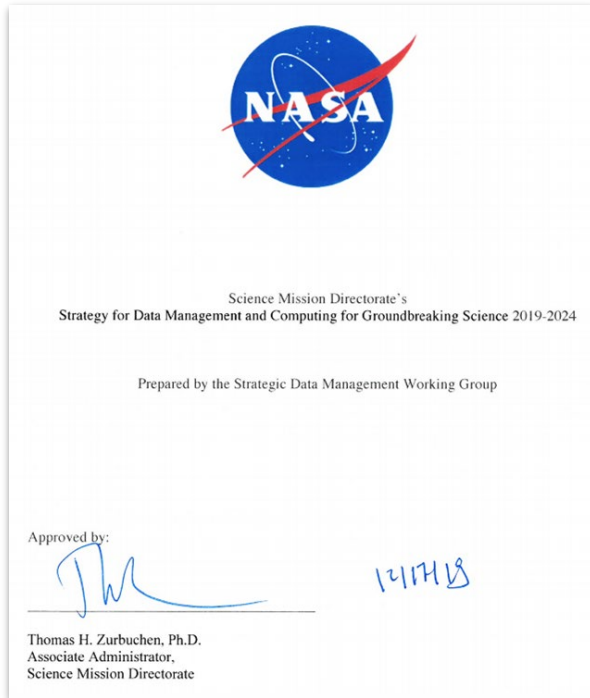
The interconnected nature of ESO missions requires a more open, collaborative, inclusive, and transparent work model end-to-end for success. DAACs will transform to Science Enabling Centers (SECs) to serve users and mission science teams in the the ESO era.

SECs will:

- Focus on end users and applications
- Have direct roles in ESO mission science objectives
- Engage with mission science teams earlier, in fusing Open Data Stewardship practices and facilitating Early adopter participation
- Lead the transformation to conducting collaborative open science in the cloud
- **The ESO Processing Workshop and Latency Study will inform this transition.**



What is the SMD Strategy for Data and Computing?



An SMD-approved strategy to enable transformational open science through continuous evolution of SMD's science data and computing systems.

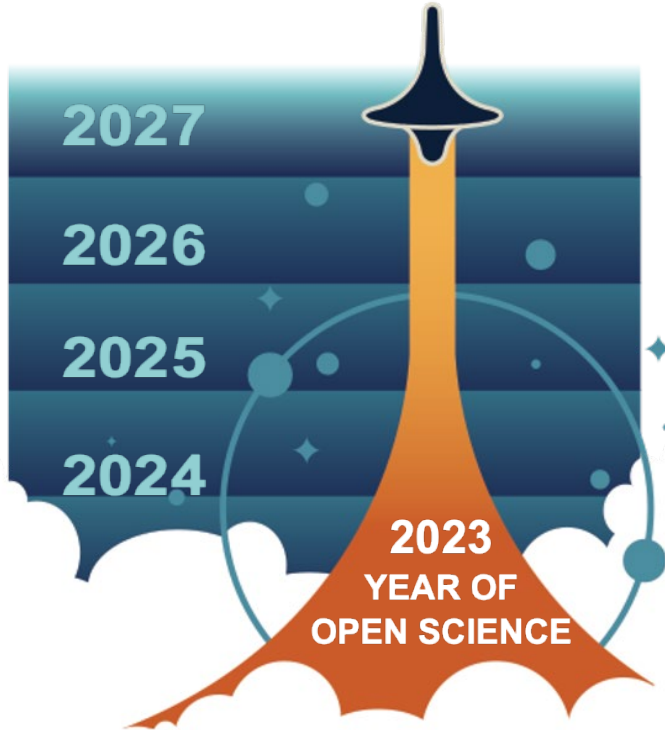
Goal 1: Develop and Implement Capabilities to Enable Open Science

Goal 2: Continuous Evolution of Data and Computing Systems

Goal 3: Harness the Community and Strategic Partnerships for Innovation



Leading the Path to Open -Source Science



Transform to Open Science (TOPS) is a \$40 million* 5 -year NASA Science Mission Directorate initiative geared towards accelerating the adoption and understanding of *open science with training and outreach*

Key Goals:

- Increase understanding & adoption of open science
- Accelerate major scientific discoveries.
- Broaden participation by historically underrepresented communities

*pending appropriations

What is TOPS doing?



- *OpenCore* is a community developed set of training materials to introduce open science and run workshops at 13 society meetings annually
- NASA has allocated \$3 million/year to fund projects related to Open Science Training via the “TOPST” ROSES 22 element.
 - Develop Discipline ScienceCore modules
 - OpenCore summer schools
 - OpenCore virtual cohorts
- Co-leading the Office of Science and Technology Policy (OSTP) Subworking group on the Year of Open Science
- Developing and maintaining a open science community on GitHub to share resources and ensure an open and transparent working environment



In Summary

We have a 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.

Backups