Data Management and Computing for SMD

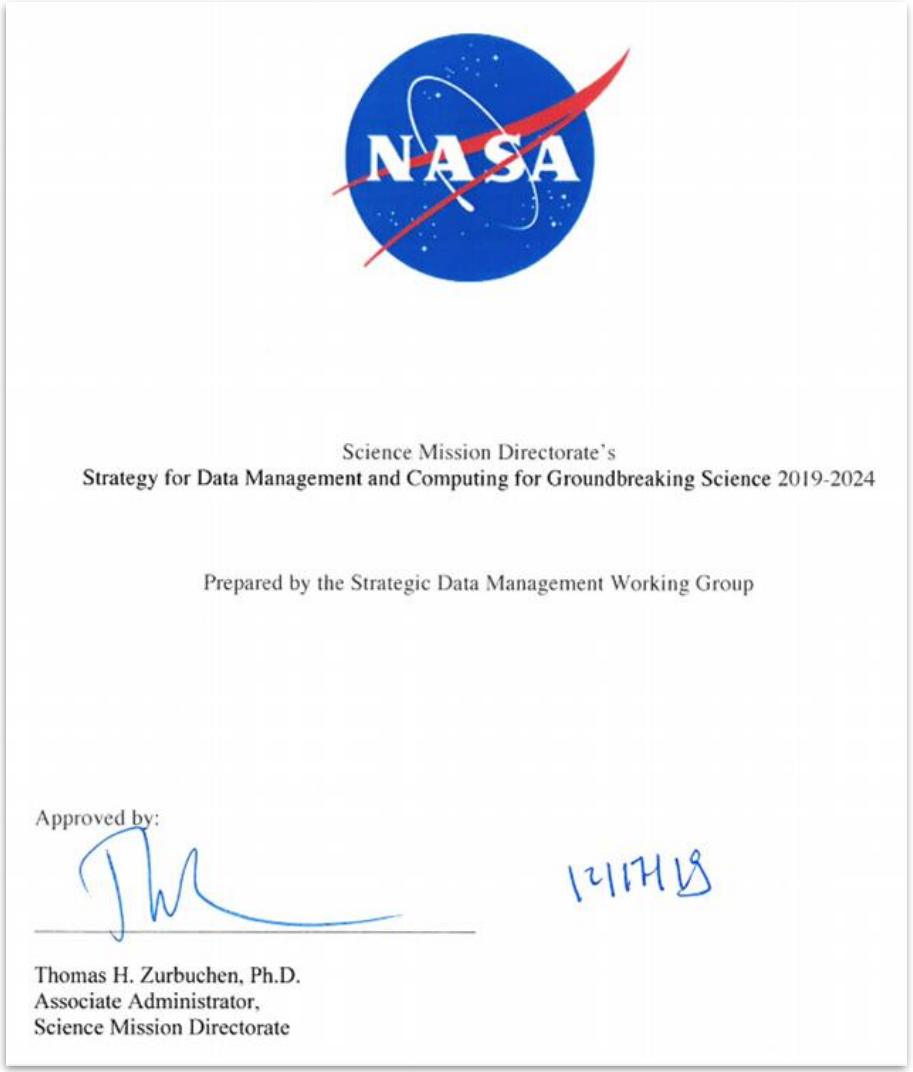
Expanding participation, improving reproducibility, and accelerating scientific discovery for societal benefit.

October 14, 2021

Kevin Murphy

CC BY

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



Science Mission Directorate's
Strategy for Data Management and Computing for Groundbreaking Science 2019-2024

Prepared by the Strategic Data Management Working Group

Approved by:

12/17/19

Thomas H. Zurbuchen, Ph.D.
Associate Administrator,
Science Mission Directorate

Vision: To enable **transformational open science** through continuous evolution of science data and computing systems for NASA's Science Mission Directorate.

Mission:

- Lead an i**nnovative and sustainable program** supporting NASA's unique science missions with academic, international and commercial partners to **enable groundbreaking discoveries with open science data**.

- **Continually evolve systems** to ensure they are usable and support the latest analysis techniques while protecting scientific integrity.

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

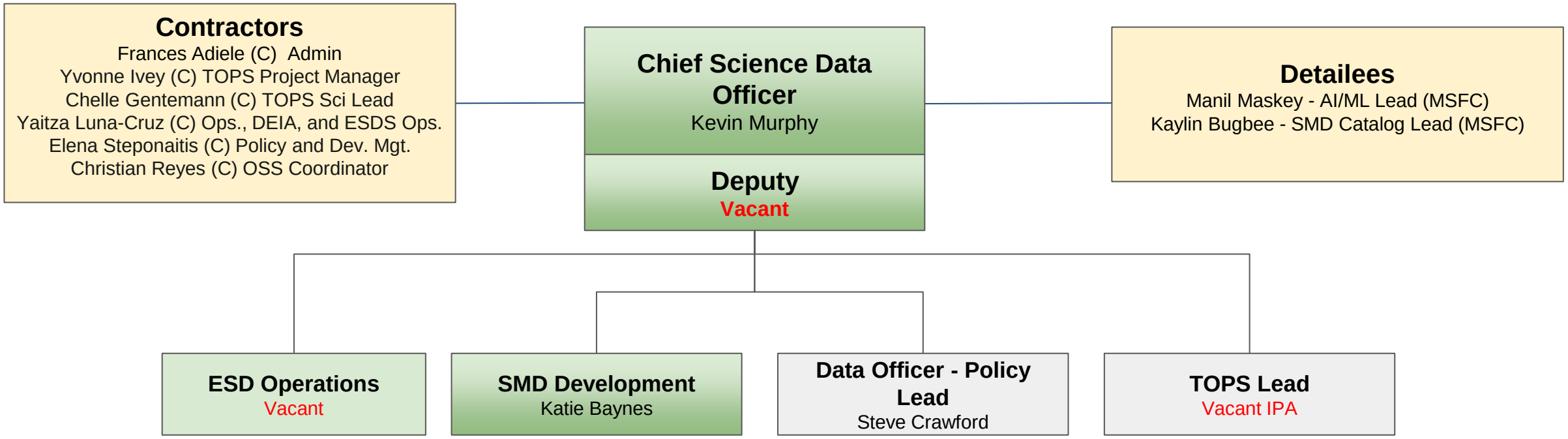
CC BY

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

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	
1.1	Develop and implement a **consistent open data and software policy** tailored for SMD	**2.1**	Establish **standardized approaches for all new missions** and sponsored research that encourage the adoption of advanced techniques	**3.1**	Develop **community of practice and standards group**
1.2	Upgrade capabilities at existing archives to **support machine readable data access using open formats and data services**	**2.2**	Integrate investment decisions in High-End Computing with the strategic needs of the research communities	**3.2**	Partner with a**cademic, commercial, governmental and international organizations**
1.3	Develop and implement a SMD data catalog to support discovery and access to complex scientific data across divisions	**2.3**	Invest in capabilities to use commercial cloud environments for open science	**3.3**	Promote opportunities for continuous learning as the field evolves through collaboration
1.4	Increase transparency into how science data are being used through a free and open unified journal server	**2.4**	Invest in the tools and training necessary to enable breakthrough science through application of AI/ML		

CC BY

Chief Science Data Office



CC BY

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."

Ramachandran, R., Bugbee, K., & Murphy, K. J. Moving from Open Data to Open Science. Earth and Space Science, Wiley Publication https://doi.org/10.1029/2020EA001562

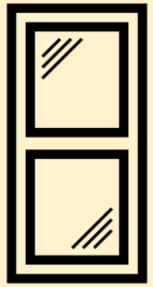
CC BY

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.

CC BY

Building **trust in the scientific process** through **transparency, accessibility, inclusivity, and reproducibility**

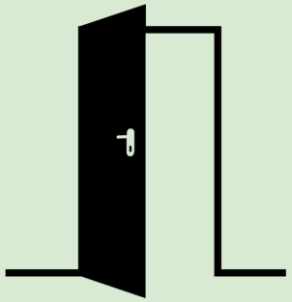


Open (**Transparent**) Science

Both the scientific process and results should be visible, accessible and understandable.

Open (**Accessible**) Science

Data, tools, software, documentation, publications should be accessible to all (FAIR).



Open (**Inclusive**) Science

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

Created by Gregor Cresnar
from Noun Project

Open (**Reproducible**) Science

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

CC BY

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.

* Projects should release all information with open licenses unless exceptions are granted based on laws or regulations, including classified, ITAR, EAR and CUI restrictions. CSDO evaluates and s or declines deviation requests by projects for NASA.

CC BY

Open-Source Science

Initiates **Transform to OPen Science (TOPS),** a 5-year program to increase understanding and adoption of open science principles and techniques
Designates **2023 as Year of Open Science**

Continues **investments in open-source science digital infrastructure**, **cross-divisional AI capabilities and Digital Transformation activities**. (ROSES elements, data catalog, open journal database)

Prototype **common data catalog** by FY22Q4, ADS expansion

Initial investments in cross-division **open scientific cloud environments and data analysis platform prototypes**.

Divisional investments in Open-Source Science are **aligned** with this program.

Fiscal year	OSS Total ($K)
FY21	$8,000
FY22	$21,000
FY23	$20,000
FY24	$20,000
FY25	$20,000
FY26	$20,000
FY27	$20,000

CC BY

Open-Source Science Accomplishments

✓	Initiated the **common SMD data catalog project** to enable cross-divisional data search and discovery.
✓	**Expanded access to free and open journals** Astrophysics Data System (ADS) journal database to include Planetary Science and Heliophysics
✓	Awarded **5 cross-divisional AI projects and selected 8 proposals** to provide support to **high value Open-Source Tools, Frameworks, and Libraries**.
✓	Co-leading the **Agency Digital Transformation (DT) Data and Knowledge Hub** activity to enable easy access to data information NASA-wide.
✓	**Released SMD's scientific information policy (SPD-41)** to support open science by requiring missions to release scientific data, publications and software openly.
✓	Completed **High-End Computing Program User Needs Assessment 2020**

CC BY

Priorities for FY22

Conduct studies

- Consolidation NASA Managed Cloud Environments
- Hybrid Cloud - HECC Architecture Study

Archive and common service prototypes in cloud environments

- Policy development and prototype development

TOPS

ROSES element for curriculum development and summer schools/cohorts/events for summer and fall of 2022

- Plan longer-term training (4-years)
- Cloud ready high value data sets
- Prizes and challenges

Open Science Activities

- AI/ML
- ADS
- ROSES for software
- SMD Catalog

CC BY

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
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

CC BY

TOPS KPIs for end of 2027

1	**Increase understanding and adoption of open science principles and techniques in our Mission and Research Communities** - 75% of mission and research principal investigators certified in open science principles - 20K scientists achieve open science certification
2	**Accelerate major scientific discoveries through supporting the adoption of open science** - One major scientific discovery using open science methods supported in each division (5 community moon shots) within 5 years
3	**Broaden participation by historically excluded communities** - Increase in participation by historically excluded communities in submitted proposals, applications from students, and participation in mission team to 20%

CC BY