National Aeronautics and Space Administration

Open-Source Science at NASA

Andrew Mitchell
WGISS-54
Tokyo, Japan
October 6, 2022

Executive Summary: Open Science Advancement

- NASA recently established a Chief Science Data Office (CDSO) which is actively implementing many of NASA's Open Source Science initiatives.
- **NASA Science Divisions are moving towards expansion of cloud activities**
- Multiple Divisions engaging in **new AI/ML activities**
- Divisions are facilitating **improved information and knowledge discovery**
- **Receive feedback** from the community on the implementation of the strategy for Open Science
- Identify tangible activities for **collaboration and development** between NASA science data repositories
- **Build networks** to support our Open Science vision

Images: Noun Project

Open Source Science Core Values

- As open as possible, as restricted as necessary, always secure
- Increase the accessibility, inclusion, and reproducibility of NASA’s Science Mission Directorate scientific activities
- When possible, minimize the burden

NASA Science Mission Directorate (SMD) Strategy for Data Management and Computing for Groundbreaking Science

SMD identifies strategic data management and science computing as a priority (February 2018)

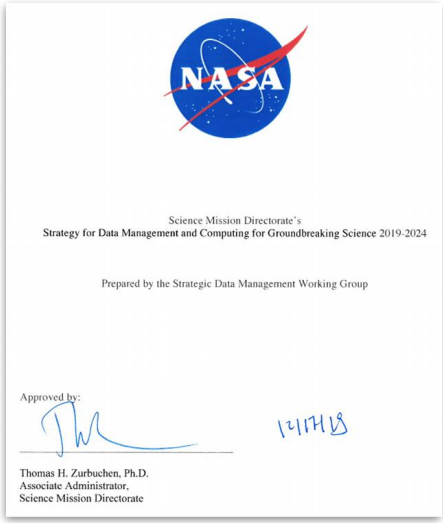
Archives Processing and Data Exploitation Meeting (GRC) (August 2018)

SMD's Strategy for Data Management and Computing (December 2019)

SMDWG KICKOFF | RFI ACTIVITY | WORKSHOP #1 | WORKSHOP #2 | STRATEGY DEVELOPMENT

67 RFI Responses with five common themes (July-September 2018)

Maximizing the Scientific Return of NASA Data (DC) (October 2018)

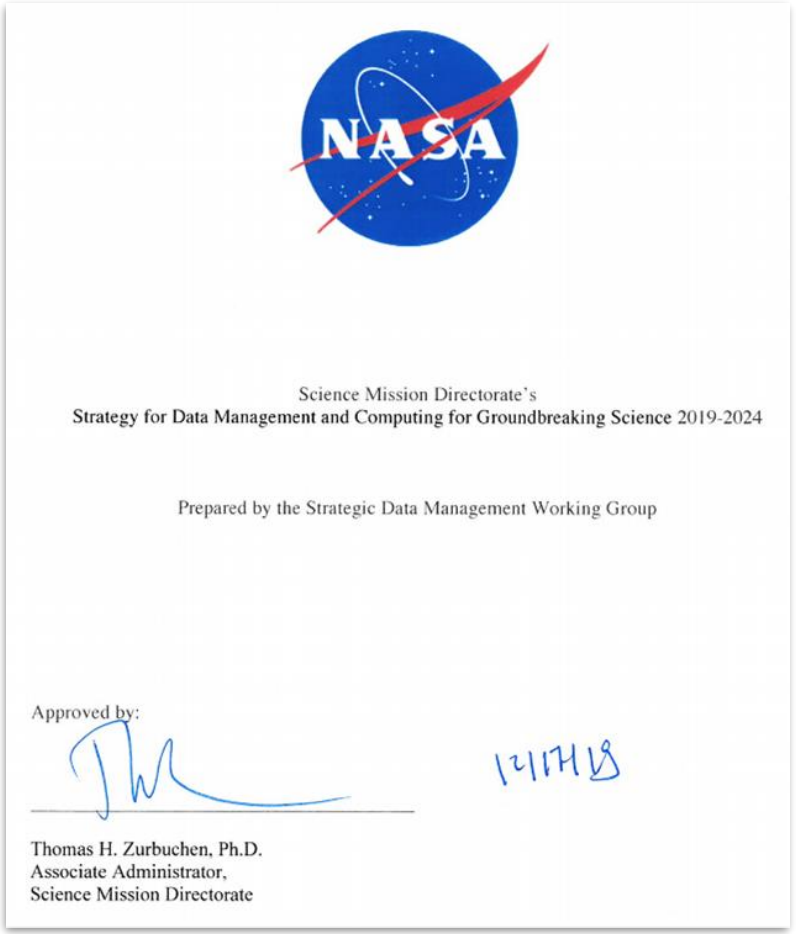

Science Mission Directorate's
Strategy for Data Management and Computing for Groundbreaking Science 2019-2024

Prepared by the Strategic Data Management Working Group

Approved by:

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

What is the SMD Strategy for Data and Computing?



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

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

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		

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.**

Overview of the implementation SPD-41a

Future implementation plans include:

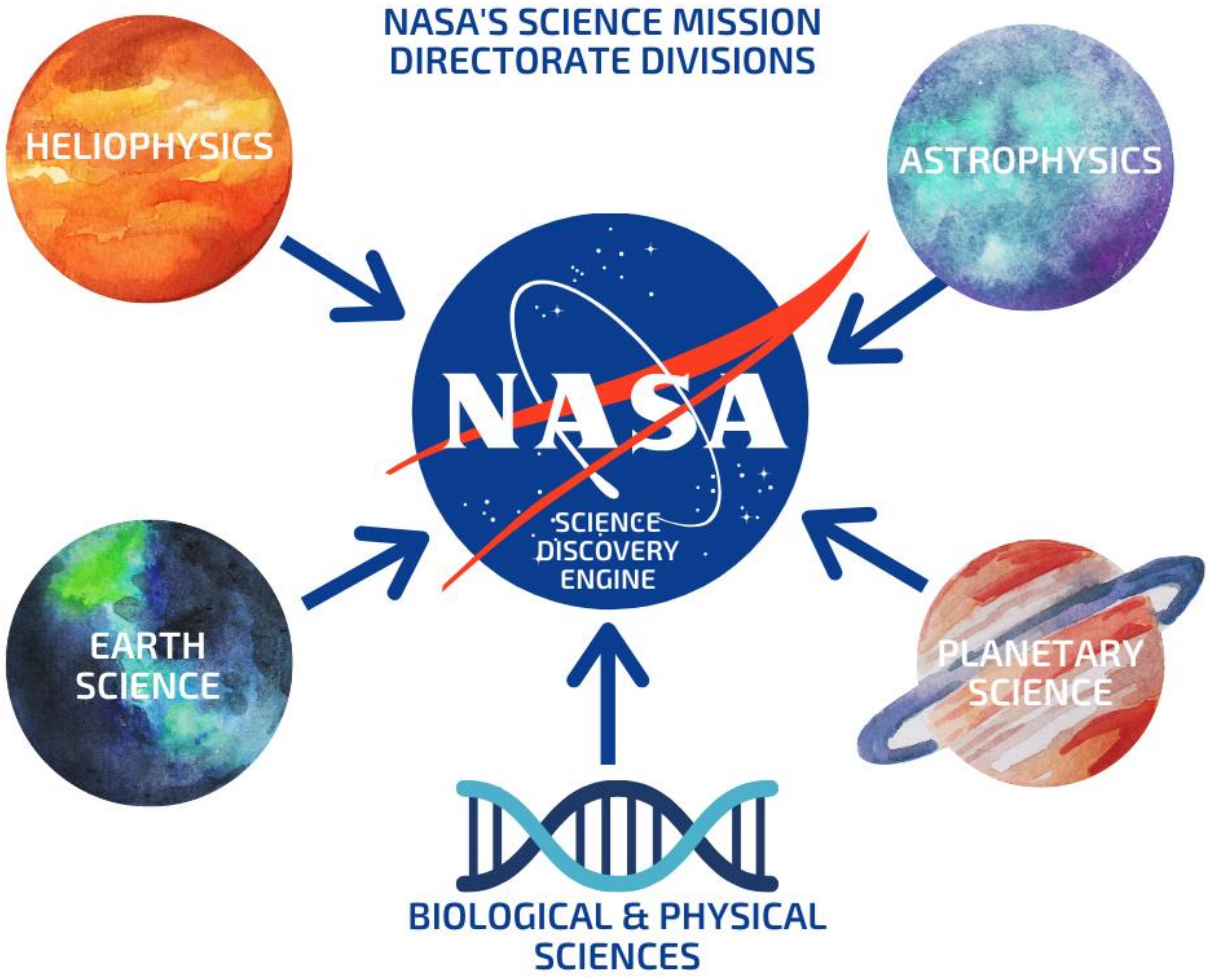
- Software release policy
- Guidance for awards, contracts, ROSES, and Announcement of Opportunities; PIs should include these costs in proposals
- Incorporated text into AOs
- Incentives for the community to make the transition - e.g ROSES22 F8. Supplement for Open Source Software

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.

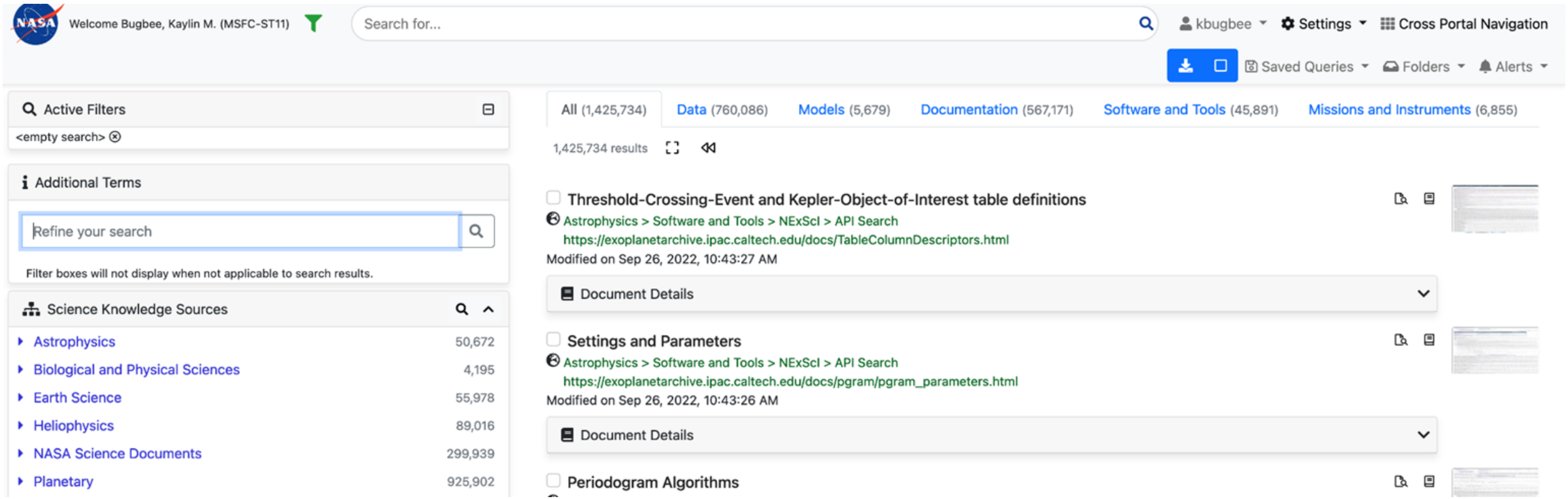
SMD Science Discovery Engine

Create an SMD discovery capability to enable open source science. Scope includes:

- Astrophysics: NAVO registry
- BPS: GeneLab, Life Sciences Data Archive
- Earth Science: Common Metadata Repository
- Heliophysics: SPASE registry, Events Knowledgebase
- Planetary Science: PDS API
- + Models, software, tools and other contextual information from all 5 divisions
- Over ***1 million documents & metadata*** included at this time.
- Incorporated 3 SMD relevant facets into the interface
    - Platforms
    - Instruments
    - Missions



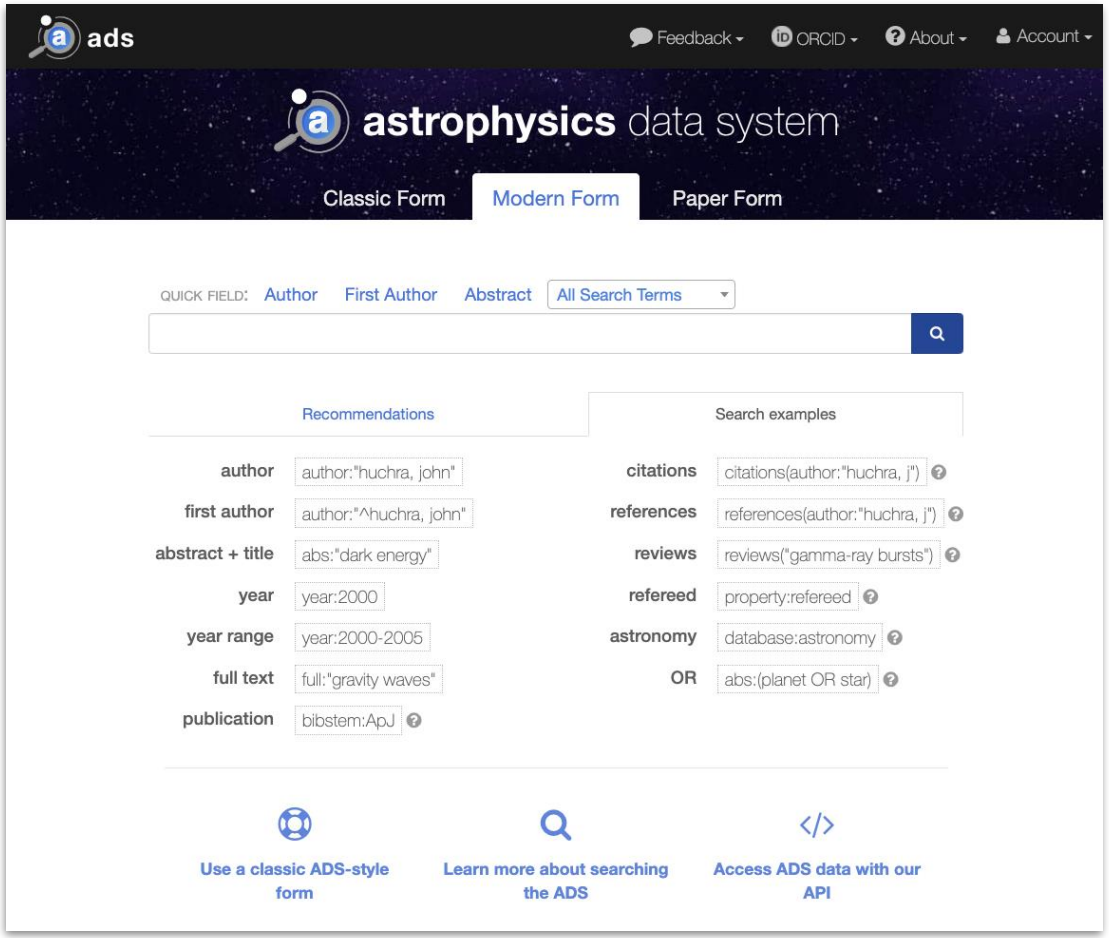
SMD Science Discovery Engine



The NASA Astrophysics Data System (ADS)

ADS is a NASA-funded project which provides discovery services for scholarly literature in Astronomy & Physics

- 15M metadata records, most of them traditional publications
- 6M full-text documents from all major publishers
- A citation graph with over 8M nodes and 142M edges
- (Anonymous) usage data for 50k regular users



https://ui.adsabs.harvard.edu


We’re on track to
accomplish much more.

What is Transform to Open Science (TOPS)?

TOPS is a 5-year NASA SMD initiative to foster adoption of Open Science practices across the scientific community.

Strategic Objectives:

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

What is TOPS doing?

- NASA has allocated $3 million/year to fund projects related to Open Science Training via the “TOPST” ROSES 22 element.
  - Develop ScienceCore
  - OpenCore summer schools
  - OpenCore virtual cohorts
- ***OpenCore*** is a community developed introduction to open science
- CSDO is participating in the Office of Science and Technology Policy (OSTP) Subworking group on the Year of Open Science
- Maintaining GitHub to share resources and ensure an open and transparent working environment

Context for Data and Computing Architecture Study

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.

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-		

For more information visit:
earthdata.nasa.gov