NASA Support for Open Science Activities

June 27, 2023

Roopesh Ojha
Astrophysics Division
Science Mission Directorate
@NASAUniverse @NASAExoplanets

Open Science

Open science is the principle and practice of making research products and processes available to all, while respecting diverse cultures, maintaining security and privacy, and fostering collaborations, reproducibility and equity.

Open-Source Science is NASA's method to put Open Science into practice

- **Open** the entirety of the scientific process, ***from start to finish***
- **Broaden** community involvement in the scientific process
- **Increase** accessibility of data, software, & publications
- **Facilitate** inclusion, transparency, and reproducibility of science

CC BY

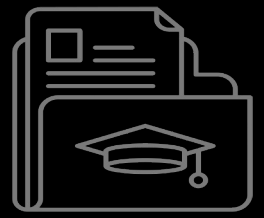
NASA Community Engagement

NASA's Transform to Open Science (TOPS) is a $40 million 5-year mission to accelerate adoption of open science

TOPS' Strategic Goals:

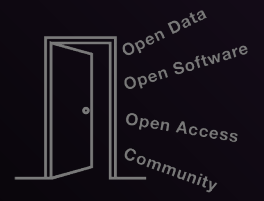
- Support 20K researchers to earn NASA's open science badge
- Double the participation of historically excluded groups across NASA science
- Enable five major scientific discoveries through open science principles

Engagement



Capacity Sharing

Incentives



Coordination

Join us as we embark on A Year of Open Science with NASA TOPS!

CC BY

NASA's 2023 Year of Open Science

TOPS Priorities:

…is the spark

CURRICULUM - Introduction to open science

- 1000 scientists earn NASA TOPS open science certificates

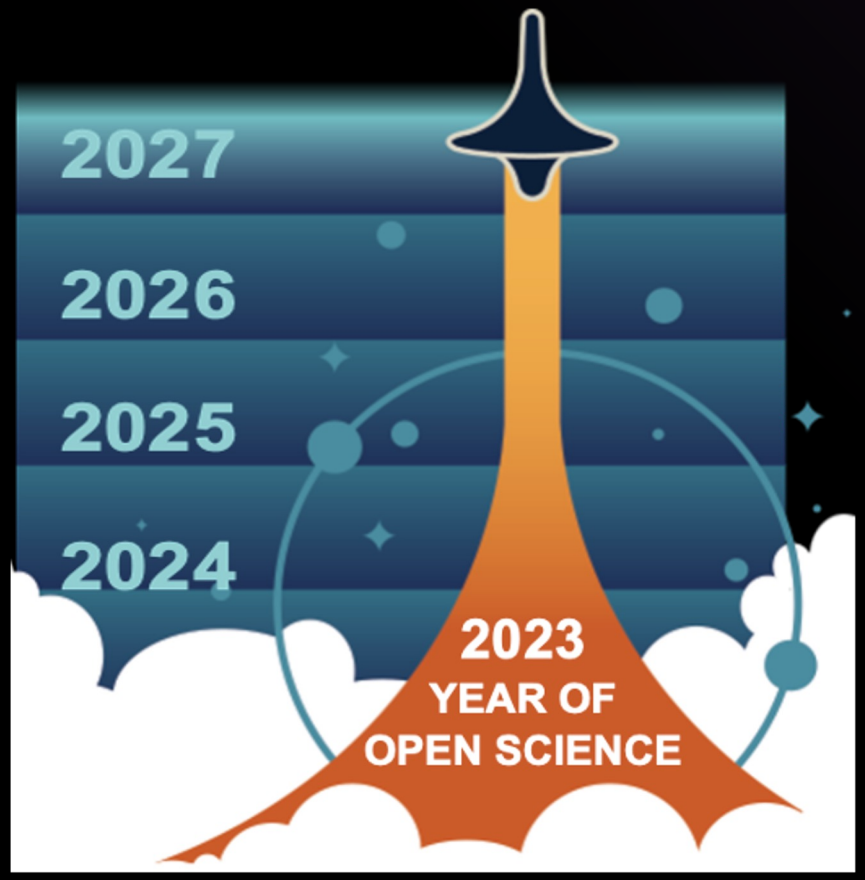
EVENTS - Engage historically underrepresented groups

- Partner w/ Minority Serving Institutions (MSIs) for marquee events, TOPS Annual Hackathon, Annual TOPS Internship program (3-4 people/year; All-in)
- Establishing Strategic Partnerships with external organizations focused on STEM engagement

OPPORTUNITIES - Develop incentives

- Pilot including open science activities in evaluations of 5+ NASA ROSES23 elements and at 5+ universities (Tenure and Promotion)
- Partner with societies on open science awards

TOPS is Increasing Understanding and Adoption of Open Science



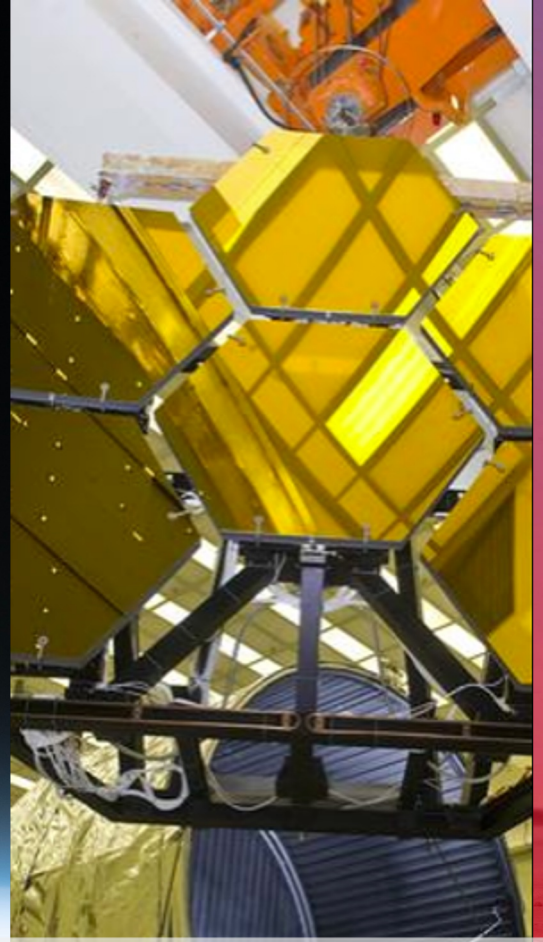
Developing the infrastructure to train 20,000 scientists and researchers as part of our five-year mission

- Introduce those beginning their open science journey to important definitions, tools, and resources
- Provide participants at all levels recommendations on best practices from subject matter experts

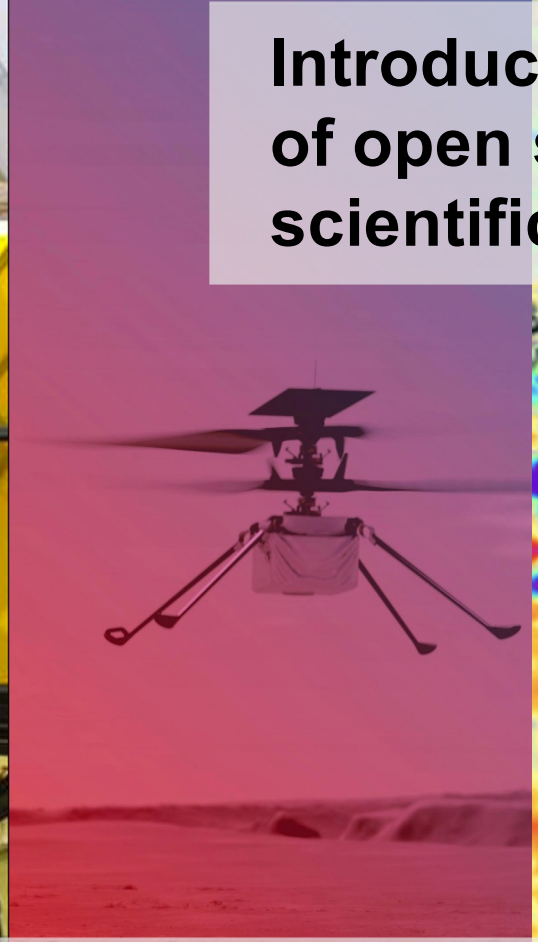
Earn NASA Open Science Certification

Introduction to Open Science: Benefits of open science presented as a scientific workflow

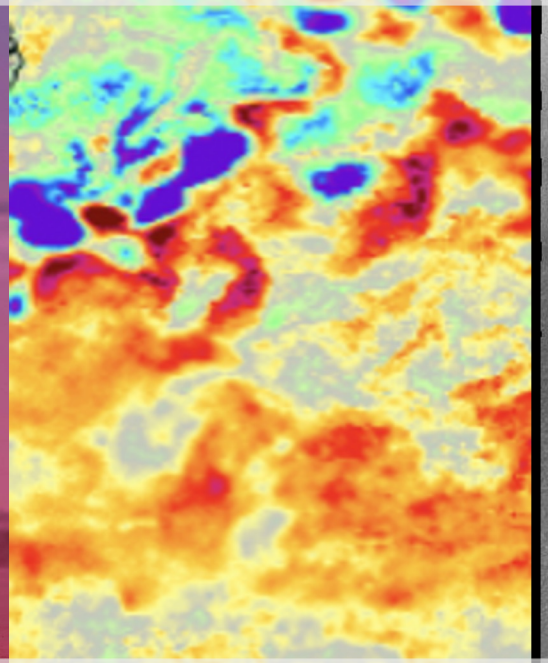
Ethos of Open Science



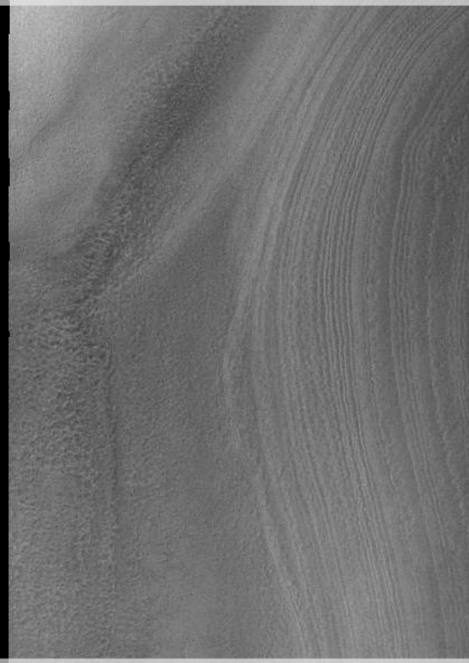
Open Science Tools



Open Source Software



Open Data



Open Results

SPD-41a is SMD's updated Scientific Information Policy.

- SPD-41a updates the previously released SPD-41, which consolidated existing Federal and NASA policy on sharing scientific information.
- SPD-41a is *forward looking* and will apply to all future SMD-funded scientific activities
- Policy updates were developed with community input and based on the OSTP Memo on Ensuring Free, Immediate, and Equitable Access to Federally Funded Research

Scientific Information Policy Website & FAQ

Major Policy Updates

- Peer-reviewed publications are made openly available with no embargo period.
- Research data and software are shared at the time of publication or the end of the funding award.
- Mission data are released as soon as possible and unrestricted mission software is developed openly.
- Science workshops and meetings are held openly to enable broad participation.

CC BY

Open-Source Science Guidance

- Narrative guidance on how to comply with SPD-41a; options and examples
- High-level, for relevance across SMD Divisions
- Adopts guidance published by other agencies
- Living document to be developed over time
- Contains:
  - Background and Motivation
  - Open Science and Data Management Plan
  - Sharing Publications
  - Data Management and Sharing
  - Software Management and Sharing
  - Sharing Materials for Science Events
  - Glossary of Open-Source Science Terms

CC BY

Astrophysics Scientific Information Management Policy

https://science.nasa.gov/astrophysics/astrophysics-data-centers

Clarifies details of the implementation of SPD-41a to the scientific information produced by Astrophysics Division funded activities

Provides guidance through the complete information lifecycle from measurements and information creation to persistent, accessible curation and sustainment

Approach

Closely coupled to SPD-41a. Generally avoids repeating material already covered there.

Minimalist approach. Avoid being too prescriptive

Living document. Will evolve with need

Building on the recommendations of the SMD's Strategy for Data Management and Computing for Groundbreaking Science 2019-2024, existing Government directives, NASA policy, community best practices, studies by the national academies, and community-led studies

Scientifically Useful Data

Full and open sharing of scientifically useful data produced by projects of all scales (including missions, sub-orbital experiments, and research and analysis investigations) with all users in an accessible manner as soon as such data become available.

Variations for:

- Initial calibration and validation (as short as practicable)
- Collaborations with international partners (rare occurrence with justification unique to the mission)
- Waiver process for GO/GI

Waiver Process

Mission PI/project scientist/observatory director may provide a process for waivers that increase (or decrease) the default data exclusive use period.

Waivers should be on a case-by-case basis, infrequent, and always with compelling justification preferably based on panel review evaluation, e.g., to protect the timely completion of a graduate student's thesis.

The HQ program scientist must be informed of all waiver decisions. Waivers that include a cumulative extension to the default exclusive access period greater than 6 months must be submitted to the HQ program scientist for approval.

Astrophysics Science Platform Project

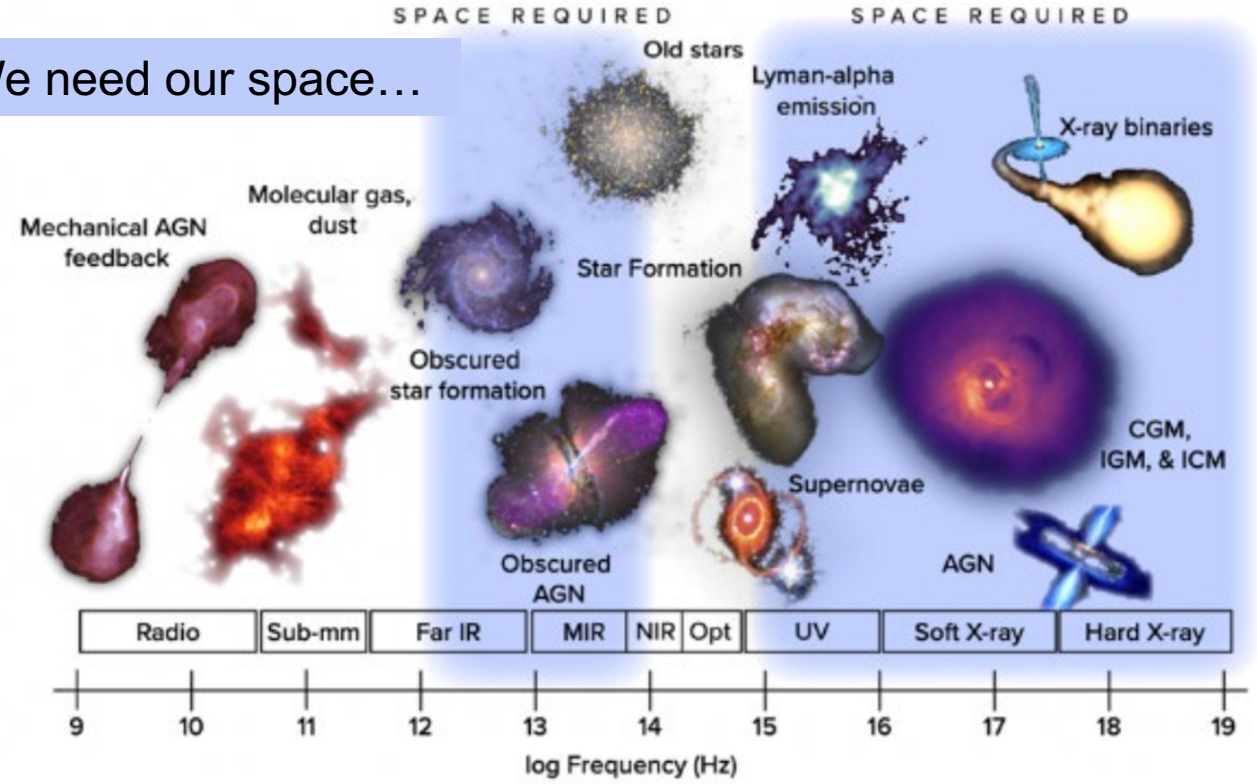
Astro2020 Decadal Survey Section 4.5.1 on Data Archiving, Curation, and Pipelines:

"The importance of joint analysis of observations from different facilities and wavelengths, and of sophisticated archiving with associated science platform tools, will grow dramatically over the next decade."

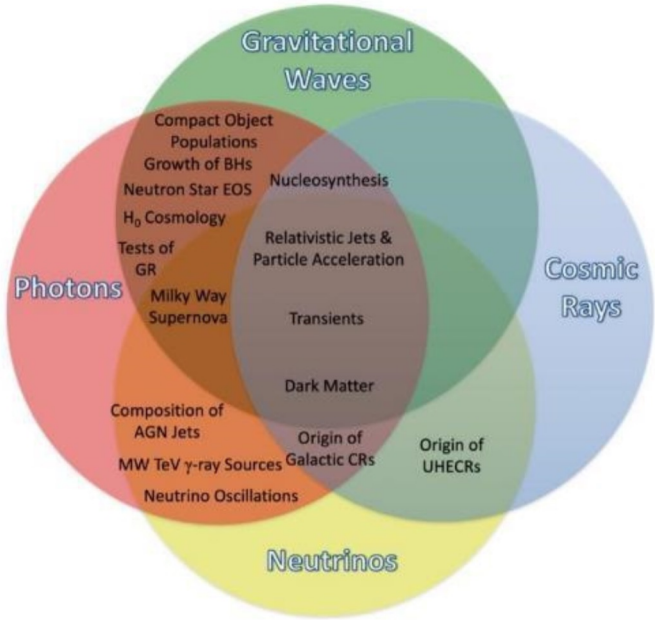
In October 2021, APAC advised the Astrophysics Division to "strategically identify appropriate cloud-based infrastructure options to facilitate analysis and theoretical modeling of the large data volumes from missions like Euclid by the wide community."

Astrophysics Data Landscape in 2030

We need our space…



"All priority science areas require multiwavelength observations." - Astro2020 Decadal Report



"Exploring the cosmos in the multi-messenger and time domains is a key scientific priority for the coming decade". – Astro2020 Decadal Report

Astro2020: multi-wavelength and multi-messenger astronomy

Data Analysis Landscape in 2030

More and better data will demand new Astrophysics Archive capabilities

- **Data volumes** in the Astrophysics Archives **double roughly every 2 years**
- **Large data sets are on the way:**
  - **Euclid** plans ~2PB in FY26 with ~6PB in FY28
  - **Roman** expects 0.75PB of level-2 science data (scan-by-scan) and level-3 data (co-added) after the first 6 months (~FY28), then 1.5PB per year
  - **Rubin Observatory (NSF-DoE)** will release 20TB/night from ~FY24, roughly 1.5PB per year, with a final data release of 15PB after 10 years of operation
- Multiwavelength astronomy requires **analyzing data sets jointly**; science often requires analyzing **data along with simulations**
- **Advanced data science tools** (e.g., AI/Machine Learning) can extract new knowledge, given adequate **computing resources close to the data**

Big Data Analytics and Machine Learning

Astrophysics Science Platform: 3 components

1. Cloud-based interface and supporting system rooted in open-source software, providing compute resources “proximate” to the data (low-latency access)
2. A rich collection of notebooks and pre-configured software containers for multi-wavelength and big-data science
3. Advanced data services that enable fast, location-agnostic services for data held on the platform, in NASA Astrophysics archives, and beyond.

Community users:
scientists needing server-side analysis for large amounts of data;
those needing computational facilities for big data analytics, Machine Learning, etc;
collaborators at different institutions wishing to share a computational environment;
scientists who cannot easily build the software they want to use; etc.

SMD RFI: Scientific Data & Computing Architecture to Support Open Science-I

- 75 responses (Govt. agencies, Universities, non-profits, companies, etc.)
- Addressed NASA's scientific data and computing architecture, cloud infrastructure, High-Performance Computing, open-source software and tools, training, and user support.
- Broadly, suggested should enable authorized users to easily collaborate, access and share information (data & software), and accommodate users with a range of expertise levels.
- Cloud Computing: - efficient cloud data services
  - the ability to quickly and easily share data and software
  - easy access to tools and libraries for authorized users

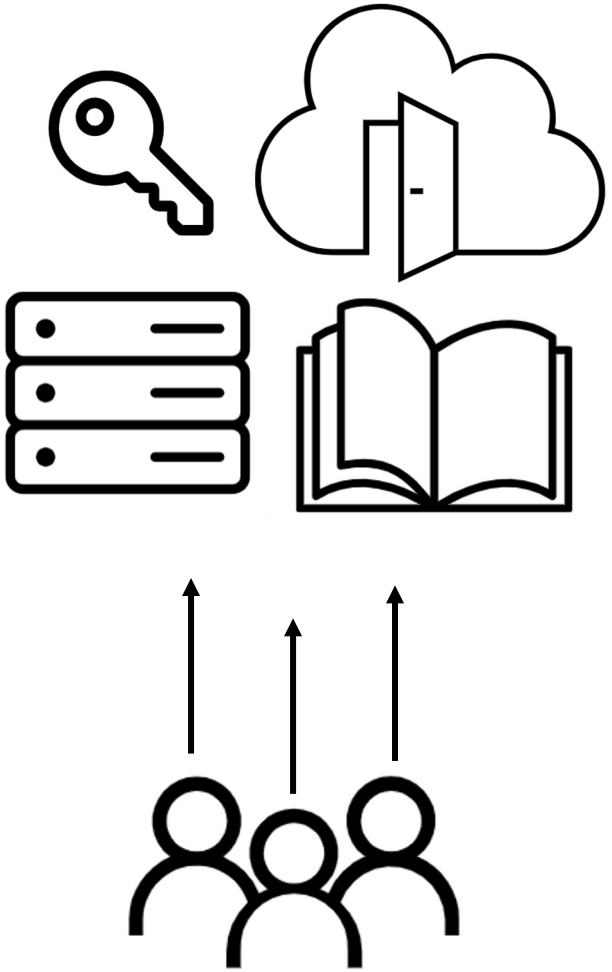
SMD RFI: Scientific Data & Computing Architecture to Support Open Science-II

- Cloud Onboarding and Training:
  - critical for NASA to offer resources to help users onboard to and utilize cloud computing and data services.
  - resources should be comprehensive and targeted towards a range of users (internal and external) and use cases.
- Astrophysics Community responses emphasized:
  - enabling data sharing
  - developing an open science data and computing infrastructure that will continue to support (with minimal disruption) existing activities of Astrophysics data archives that are critical to scientific collaboration and discovery
  - increasing accessibility of computing services

Thank you so much!

Backup Slides

What are SMD Core Services?



Existing and new data and computing capabilities that are reusable by all Divisions

Core services will include:

- ROSES SPD-41 Compliance
  - Research Data and Software Archive(s)
  - Independent FAIR assessments
- Hybrid Cloud Computing Environments
  - Cybersecurity/ATOs (high, moderate, low)
  - Baseline capabilities such as egress optimization, deep archival, user self registration and usage metrics, data transfer service
  - Account management, billing, CM and notifications
  - Discounts - egress, storage and processing
- Publically Available Collaboration Tools (e.g. GitHub, Jira, Slack)
- Scientific Information/Knowledge Management
- Open Source Science Training

Scientifically Useful

• Scientifically Useful: any material that you as a scientist would need in order to assess, interpret, evaluate, validate, or reproduce the results of a scientific project. Questions and clarifications will be provided by the responsible Program Officer/Program Scientist and the final authority rests with the Division Director. As the Astrophysics Division is gaining practical experience with the application of “scientifically useful” to the broad range of information circumstances within NASA astrophysics, community input will be sought and community standards will be considered.