



OPEN SOURCE SCIENCE DATA REPOSITORIES MEETING 2026

SEPTEMBER 1–3 | ARLINGTON, VIRGINIA

DAY 1

Tuesday, September 1, 2026

EDT

8:30 – 9:00 AM

Registration Opens

9:00 – 9:15 AM

Welcome Session

9:15 – 10:00 AM

Plenary: Foundation Models for Scientific Discovery: Implications for Data, AI, and Open Science

Speaker: *Dr. Vipin Kumar*, Regents Professor and William Norris Chair in Large Scale Computing, Department of Computer Science and Engineering Director, CSE Data Science Initiative

10:00 – 10:30 AM

Break

10:30 – 11:15 AM

Plenary: TBD

Speaker: *Dr. Vanessa McBride*, Science Director, Acting Head of the Centre for Science Futures, International Science Council

11:15 AM – 12:15 PM

From Vision to Practice: Division Perspectives on AI Adoption

Moderator: *Kaylin Bugbee*

This session will bring together leaders from across the SMD divisions to share their strategic vision, current efforts, and emerging priorities for integrating AI into scientific workflows and operations. The hybrid session will begin with brief division presentations, followed by a panel discussion exploring opportunities, challenges, and lessons learned as SMD moves from AI aspirations toward practical adoption.

12:15 – 1:30 PM

Lunch

1:30 – 3:00 PM

Breakout Sessions

Capital Classroom | Main Room:

Discovery and Metadata for AI

Moderator: *Valerie Dixon*

This discussion will look at ways AI and ML can improve search and discovery of metadata, followed by approaches for making your data more accessible and usable by AI systems.

- Global Change Master Directory (GCMD) AI/ML Keywords: *Valerie Dixon*
- Common Metadata Repository (CMR) NLP/EDSearch Demo: *Nick Doty*
- Geospatial “AI-Readiness” Key Areas: *Nick Doty*
- AI-Optimized Science Metadata Formats: *Bishwas Praveen, Bernard Benson*
- OpenAPI for LLMs: *Kirill Grigorev*
- Information Governance for AI/ML: *Scott Simmons*

National Landing Classroom | Breakout Room:

Toward a Shared Understanding of Analysis-Ready, Cloud-Optimized, and AI-Ready Data

Moderators: *Sam Gebre, Kaylin Bugbee*

Analysis-Ready Data, Cloud-Optimized Data, and AI-Ready Data are widely used concepts with overlapping goals but distinct emphases. This session will explore how these approaches relate to one another, their implications for repositories and scientific workflows, and how the community can build a shared understanding to guide future data practices across NASA science.

3:00 – 3:30 PM

Break

DAY 1

Tuesday, September 1, 2026

EDT
3:30 – 5:00 PM

Breakout Sessions

Capital Classroom | Main Room:
Moderator: Kaylin Bugbee

Stewarding Scientific Knowledge in the AI Era: New Questions for Data Governance, Provenance, and Trust

As AI becomes an increasingly important part of scientific workflows, how should we capture provenance, ensure transparency and reproducibility, and maintain trust in AI-enabled science? This session will explore emerging challenges related to documentation, provenance, preservation and the evolving role of data stewardship in supporting open and trustworthy science.

National Landing Classroom | Breakout Room:

AI Tools for Data Stewardship

Moderators: Nidhi Jha, Sam Gebre, Danielle Lopez

Through short talks and open discussion, this session examines the current landscape of AI tools for data stewardship across SMD's data repositories — what's available today, what's most widely used, and where the gaps are in metadata, standards, and interoperability across archives.

DAY 2

Wednesday, September 2, 2026

EDT
8:00 – 8:30 AM

Registration Opens

8:30 – 9:30 AM

Agency Chief Data and AI Officer (CDAIO) Talk
Speaker: Sara Cermenaro

9:30 – 10:30 AM

OCSDO Programmatic Updates
Moderator: DeVon Griffin

This session will provide brief updates from key OCSDO-supported initiatives, including the Data and Analysis Services Project (DASP), Science Cloud Infrastructure Project (SCIP), High-End Computing (HEC), and Science Explorer (SciX). Speakers will highlight recent progress, upcoming priorities, and opportunities for community engagement across these programs.

10:30 – 11:00 AM

Break

11:00 – 12:00 PM

Responsible AI Tool Development, Use & Trust
Moderator: Shubha Ranjan

This session will feature an open discussion on the various LLMs used across NASA (Copilot and Claude) along with several novel use cases. We will also briefly cover AI security best practices and demonstrate a use case for connecting MCP agents with JupyterHub.

- AI Governance: Dhanur Sharma
- AI Security & Best Practices: Ramon Ramirez-Linan

12:00 – 1:30 PM

Lunch

1:30 – 1:40 PM

Building a Community for AI Agents for Science
Speaker: Rahul Ramachandran

As different groups are exploring, prototyping and building agents, there is a need to coalesce as a community to ensure that the practices, tools, and design artifacts that enable agent quality can be shared, peer-reviewed, and improved across the broader data-repository community. This session brings scientists, data stewards, and developers into one room to start addressing this question.

EDT

1:40 – 2:40 PM

Division Presentations on AI Agents for Science

Moderator: *Rahul Ramachandran, Sam Gebre*

This session will highlight a range of AI agents being developed and used to advance scientific research across NASA's SMD divisions.

- Multisource Integrated Observatory (MIO) Worldview Agents: *Cécile Rousseaux*
- Accelerated Knowledge Discovery (AKD) Demo: *Nidhi Jha*
- Knowledge Graph, MCP: *Amanda Saravia-Butler*
- MCP and MCP Registry: *Ramon Ramirez-linan*

2:40 – 3:00 PM

Break

Lessons Learned from Building AI Agents for Science

Moderator: *Rahul Ramachandran*

Participants will collaboratively review and refine a draft set of community norms and best practices for AI agents in science, covering design, development, evaluation, trust, and responsible deployment. Using a shared GitHub repository, attendees can suggest additions, modifications, and improvements, helping build a community-driven, consensus-based set of best practices that evolves with real scientific needs and experiences.

Let's Build an Agent Together & Close Out

Moderators: *Rahul Ramachandran, Nidhi Jha*

In this interactive hands-on session, participants will collaboratively design and build a Session Report Agent using the AKD Co-Design studio. Rather than simply watching a demo, attendees will actively contribute throughout the agent design process by helping define the agent's purpose, providing context, suggesting prompts, identifying required tools, and refining the expected outputs.

EDT

8:00 – 8:30 AM

Registration Opens

NASA Science Mission Directorate's Strategy for Data Management and Computing to Accelerate Scientific Discovery, 2025–2030

Speaker: *Chelle Gentemann*

This session will explore NASA Science Mission Directorate's vision for modernizing scientific data and computing, including AI-ready infrastructure, integrated data and compute environments, and community-driven capabilities that accelerate scientific discovery. Discussion will focus on implementation priorities, emerging challenges, and opportunities for collaboration across the science community.

Commercial Data and the Future of NASA Science Data Management

Speakers: *Dana Ostrenga, Robin Ferguson*

Commercial data will become an increasingly important part of NASA's scientific data ecosystem, augmenting traditional mission data with new observations, capabilities, and pathways for discovery. The rise of commercial data challenges traditional assumptions about open access and long-term preservation, driving the need for new data management frameworks that support both scientific reuse and commercial partnerships.

10:30 – 11:00 AM

Break

EDT

11:00 – 12:00 PM

Envisioning Scientific Data Stewardship in the AI Era

Moderators: Kaylin Bugbee, Kerstin Lehnert

This interactive session will invite participants to explore plausible futures for what scientific data stewardship should look like five, ten, or twenty years from now. Together, participants will identify emerging signals of change, discuss the trends driving those signals, and develop future scenarios for scientific data stewardship in an AI-enabled research ecosystem.

12:00 – 1:00 PM

Lunch

1:00 – 2:30 PM

Breakout Sessions

Capital Classroom | Main Room:

Research Data Repository Update and Feedback Session

Moderator: Siddharth Chaudhary

In this session, we will provide a live demonstration of the NASA Research Data Repository capabilities, followed by an interactive discussion on metadata curation, repository features, discovery, metrics, and future priorities. Participants will have the opportunity to share their experiences, discuss challenges, and provide feedback to help shape the repository's continued development.

National Landing Classroom | Breakout Room:

Stewarding Data Through the Data Management Lifecycle

Moderators: Nidhi Wahi, Danielle Lopez

Through cross-division discussion, this session will introduce the NASA SMD Data Management Lifecycle and explore where repository-specific approaches are appropriate and where increased alignment could strengthen stewardship, improve interoperability, and enhance support for future scientific discovery.

- Earth Science Data and Information System (ESDIS): Beth Huffer

2:30 – 3:00 PM

Break

3:00 – 4:30 PM

Breakout Sessions

Capital Classroom | Main Room:

Governance Initiatives & GUAC Update and Feedback Session

Moderators: Nidhi Wahi, Dhanur Sharma

In this session, we will cover the current state of several ongoing Governance initiatives and provide a live demonstration of the Governance for Understanding, Accessibility, & Collaboration (GUAC) Knowledge Hub, a unified, role-based governance application striving towards collaboration across the SMD disciplines and the support of increasingly complex governance processes and decision-making.

National Landing Classroom | Breakout Room:

What Makes Scientific Data 'Authoritative?'

Moderators: Kaylin Bugbee, Valerie Dixon

This session will examine what authoritative means in the context of scientific data and whether it implies different stewardship responsibilities for data providers, repositories, and the broader scientific community.

4:30 – 5:00 PM

Workshop Closeout