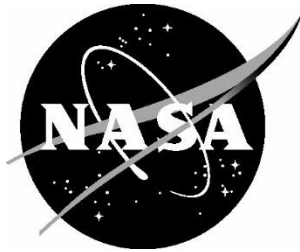


NASA/ SP–20260002277



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

Nicola J. Fox
National Aeronautics and Space Administration

Kevin Murphy
National Aeronautics and Space Administration

April 2026

NASA STI Program Report Series

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NTRS Registered and its public interface, the NASA Technical Reports Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- **TECHNICAL PUBLICATION.** Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.
- **TECHNICAL MEMORANDUM.** Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.
- **CONTRACTOR REPORT.** Scientific and technical findings by NASA-sponsored contractors and grantees.

- **CONFERENCE PUBLICATION.** Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.
- **SPECIAL PUBLICATION.** Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.
- **TECHNICAL TRANSLATION.** English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at <http://www.sti.nasa.gov>

NASA/ SP–20260002277



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

Nicola J. Fox
National Aeronautics and Space Administration

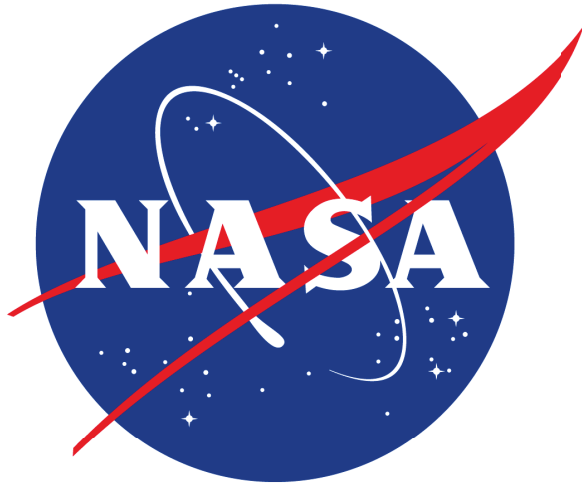
Kevin Murphy
National Aeronautics and Space Administration

National Aeronautics and
Space Administration

April 2026

National Aeronautics and
Space Administration

April 2026



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

Prepared by the Office of the Chief Science Data Officer

Approved by:

Nicola Fox Digitally signed by Nicola Fox
Date: 2026.04.23 07:30:29
-04'00'

Dr. Nicola J. Fox
Associate Administrator (AA)
Science Mission Directorate

Kevin Murphy Digitally signed by Kevin Murphy
Date: 2026.03.24 13:08:27
-05'00'

Kevin J. Murphy
Chief Science Data Officer
Science Mission Directorate

Suggested Citation: NASA (2026) NASA Science Mission Directorate’s Strategy for Data Management and Computing to Accelerate Scientific Discovery 2025–2030, <https://doi.org/10.64631/KUER1963>.

Artificial Intelligence (AI) Usage Disclosure: This document was created with assistance from ChatGSFC to combine, clarify, and summarize information. The content has been reviewed and edited by the author team. More information on the extent and nature of AI usage is included in the Acknowledgements section of the document.

TABLE OF CONTENTS

TABLE OF CONTENTS.....	3
INTRODUCTION	4
BACKGROUND	4
OVERVIEW	5
APPROACH	6
VISION, MISSION, AND GOALS	8
VISION.....	8
MISSION	8
GOALS	8
STRATEGIES TO ACHIEVE THESE GOALS.....	8
FINDINGS AND RECOMMENDATIONS.....	12
ACKNOWLEDGEMENTS.....	15
REFERENCES	16
APPENDIX A: STATUS OF 2019–2024 STRATEGY RECOMMENDATIONS.....	20
Area 1: Open Data/Open Software	20
Area 2: High-End Computing (HEC) Program	21
Area 3: Archives Modernization.....	21
Area 4: Advanced Capabilities	24
Area 5: Management.....	25
APPENDIX B: LANDSCAPE ANALYSIS.....	26
OCSDO Organizational Structure and Programs	26
Current Challenges	26
Computational Resource Limitations	27
Data Standards and Interoperability	28
Software Recognition and Support.....	29
Workforce Skills Gap	29
APPENDIX C: DETAILED FINDINGS AND RECOMMENDATIONS	31
Area 1: Infrastructure and Management	31
Area 2: Standards and Accessibility	33
Area 3: Computation and Analysis Environments	35
Area 4: Artificial Intelligence and Quantum Information Processing.....	36
Area 5: Documentation and Training	38
Area 6: Governance and Strategy	39
APPENDIX D: ALIGNMENT WITH OCSDO AND NASA	41
APPENDIX E: DEFINITIONS AND ACRONYMS	45
Definitions	45
Acronyms.....	45

INTRODUCTION

NASA's Science Mission Directorate (SMD) seeks to “discover the secrets of the universe, to search for life elsewhere in the Universe, and to protect and improve life on Earth and in space” ([NASA, 2025](#)). To achieve these ambitious exploration goals, SMD manages its scientific data infrastructure and provides high-end computing (HEC) capabilities that support NASA's entire mission portfolio. These capabilities inform critical decision-making across multiple areas of national importance: precision agriculture support that safeguards our food supplies, space weather predictions for protecting critical infrastructure, and advanced simulations for rocket development. From Mars mission simulations to deep space data processing, this infrastructure advances our understanding of Earth, our solar system, and the universe beyond while pushing the boundaries of human exploration. SMD continuously prioritizes the development and evolution of these systems so that they continue to enable breakthrough discoveries and deliver direct benefits to the American people.

BACKGROUND

In 2019, NASA's Science Mission Directorate's (SMD) Strategy for Data Management and Computing for Groundbreaking Science 2019–2024 ([SDMC24](#)) laid the groundwork for open science and set out a strategy to evolve data and computing resources to meet researchers' increasingly complex needs and adapt to rapidly changing technology. While many [SDMC24](#) recommendations have been successfully implemented ([Appendix A](#)), both scientific practices and enabling technologies have continued to evolve rapidly, requiring ongoing adaptation of our systems to sustain and accelerate scientific progress. The Strategy for Data Management and Computing to Accelerate Scientific Discovery 2025–2030 (SDMC30) provides a roadmap for NASA's scientific data and computing evolution, focusing on computational efficiency and productivity-enhancing tools that will accelerate scientific discovery.

SMD operates NASA's HEC, a critical resource for NASA's Aeronautics, Exploration, and Science Mission Directorates. However, unmet needs in HEC capabilities are limiting simulations that support exploration and next-generation aircraft development. SMD also manages over 150 Petabytes (PB) of observational data representing decades of exploration and discovery. These data are needed to prepare for exploration of the Moon, Mars, and beyond (e.g., Lunar Reconnaissance Orbiter, Mars Reconnaissance Orbiter, Solar Dynamics Observatory). By 2030, the projected total scientific data volume, including exciting data from new worlds and extreme environments, is expected to be 530 PB. Together, managing both this vast data enterprise and essential computing infrastructure requires a coordinated, strategic approach that encompasses the entire scientific research lifecycle: data acquisition and management, advanced analytical capabilities, and long-term preservation.

SMD views computing and data environments as deeply interconnected components essential for modern scientific workflows, recognizing that effective discovery and exploration increasingly depend on their seamless integration. To address this critical relationship, SMD has designated the Office of the Chief Science Data Officer (OCSDO) to coordinate with divisional science data officers while also managing SMD's HEC capabilities. This approach ensures efficiency,

interoperability, and strategic alignment across both domains, reflecting the reality that modern analytical methods require coordinated evolution of data systems and computational resources. The integrated management structure has enabled significant progress in implementing recommendations from the 2019-2024 strategy, including increasing alignment with open science principles, modernizing data systems, developing cross-divisional search capabilities, conducting assessments of HEC and cloud resources, and developing Artificial Intelligence (AI) capabilities.

By advancing data management, software development, and computational infrastructure, this strategy aims to transform SMD's data and computing enterprise into a modern, efficient ecosystem that accelerates both scientific discovery and exploration capabilities, enabling the next generation of missions to Earth, the Moon, Mars, and beyond while maximizing return on investment for the American public.

OVERVIEW

Science and technology have evolved dramatically since [SDMC24](#) was released. Since 2019, exponential growth in data volumes, the emergence of AI systems as scientific collaborators, increased demands for transparent, reproducible research, and growing computational needs have transformed the discovery process. Federal agencies are mandated to modernize their data infrastructure to enhance accessibility, transparency, and reproducibility^{1,2}, support American AI leadership through optimized infrastructure and data capabilities³, provide a framework for responsible AI implementation⁴, and establish guidelines for strategic AI procurement⁵. These converging factors necessitate a strategic evolution of NASA's data and computing infrastructure to advance science, harness emerging technologies, and advance the agency's research priorities.

To ensure our systems operate efficiently, evolve with technological advancements, and remain aligned with NASA's objectives, SMD conducts comprehensive internal assessments that rely public community workshop reports, decadal surveys, and Office of Inspector General (OIG) findings, every five years. For this strategy, OCSDO evaluated each division's data and computing resources (detailed in the NASA Science Mission Directorate's FY2024 Science Data Repository Metrics Report ([2026](#))). This assessment identified five critical challenges that this strategy addresses:

¹ U.S. Congress. (2019). [Foundations for Evidence-Based Policymaking Act of 2018](#).

² Executive Order No. 14303. (2025, May 28). "[Restoring Gold Standard Science](#)". Federal Register, 90 FR 22601.

³ Executive Order No. 14320, "[Promoting the Export of the American AI Technology Stack](#)," Federal Register, Vol. 90, No. 142 (July 28, 2025), 35393.

⁴ Executive Office Of The President Office of Management and Budget (2025). Memorandum M-25-21, "[Accelerating the Use of AI in Federal Agencies](#)," April 3, 2025.

⁵ Executive Office Of The President Office of Management and Budget (2025). Memorandum M-25-22, "[Driving Efficient Acquisition of Artificial Intelligence in Government](#)," April 3, 2025.

1. **Infrastructure Fragmentation:** Duplicative systems limit efficiency and interoperability while increasing costs and security vulnerabilities.
2. **Computational Resource Limitations:** Scientific demands increasingly exceed available HEC and researchers struggle to access cloud computing.
3. **Data Standards and Interoperability:** Inconsistent implementation of Findable, Accessible, Interoperable, and Reusable (FAIR) principles and metadata standards hinders data discovery, citation, and reuse across disciplines.
4. **Software Recognition and Support:** Insufficient recognition of software as a critical scientific asset, as well as an increasingly significant infrastructure constituent, leads to inadequate support, preservation, maintenance, and reuse.
5. **Workforce Skills Gap:** Growing need for specialized expertise in data science, AI, and other complex methods outpaces NASA's ability to recruit and retain qualified personnel, hire contractors, or establish Space Act Agreements.

By advancing data management, software development, and computational infrastructure, this strategy provides a comprehensive roadmap for NASA's scientific data and computing evolution. The strategy is organized around three strategic goals with corresponding strategies and recommendations ([Table 1](#)), implemented through six focused areas of action ([Table 2](#)) that directly address the identified challenges. This organizational structure ensures both strategic vision and practical implementation, balancing immediate needs with long-term scientific objectives. The following section details the collaborative approach used to develop this strategy, ensuring it reflects the diverse requirements of NASA's scientific community while maximizing return on investment for the American public.

APPROACH

In October 2024, the OCSDO established a 2025-2030 Strategy Development Team ([Table 3](#)) that included a steering team from OCSDO and a tiger team with representatives from each science division. Development was centered on three core principles:

1. **Data-driven:** We conducted comprehensive analysis of existing reports, decadal surveys, Office of Inspector General (OIG) findings, and community workshops to identify persistent challenges and promising solutions. This approach ensures recommendations are grounded in demonstrated needs rather than technological novelty.
2. **Cross-divisional collaboration:** By establishing a Strategy Development Team with representation from each science division, we ensured divisional perspectives were incorporated throughout strategy development. This collaborative structure helped distinguish between truly unique divisional needs and common capabilities that can be standardized for increased efficiency and interoperability.
3. **Scientific Excellence:** We evaluated all recommendations against principles of Gold Standard Science (reproducibility, transparency) and interdisciplinary collaboration to

ensure NASA's data and computing infrastructure supports the highest standards of scientific integrity.

This 2025–2030 strategy is informed by multiple sources across all five science divisions. Each division has published recent decadal studies or midterm assessments: Astrophysics ([2023](#)), Earth Science ([2024](#)), Heliophysics ([2020](#)), Planetary Science ([2023](#)), and Biological and Physical Sciences ([2023](#)). Additionally, National Academies of Sciences, Engineering and Mathematics (NASEM) reports, NASA Office of Inspector General (OIG) reports, Executive Orders, OMB memorandums, and community workshops have contributed valuable insights (see [References](#)).

The Strategy Development Team completed a systematic review of existing reports, decadal surveys, Office of Inspector General (OIG) findings, community workshops, division-specific data holdings, computing resources, and users ([Appendix B](#)), looked across to understand the current landscape, identify existing challenges and emerging opportunities ([Appendix C](#)), and develop recommendations. This strategy will guide SMD's data and computing evolution over the next five years, positioning NASA at the forefront of scientific discovery in support of NASA's mission, while maintaining the highest standards of scientific integrity.

VISION, MISSION, AND GOALS

The Strategy Development Team developed a vision, mission, goals, and strategies to accelerate scientific discovery which guided the development of findings and recommendations.

VISION

Accelerating science through world-class data, computing, and AI capabilities that empower NASA scientists and partners to move from data to discovery faster.

MISSION

To provide seamlessly integrated data, computing systems, and advanced analytical tools that accelerate discovery, enable exploration, and maximize the return on America's investment in Earth and space sciences.

GOALS

This strategy defines three complementary goals that transform SMD's approach to scientific data and computing over the next five years:

Goal 1: Build an open, interoperable ecosystem for scientific discovery and economic opportunity

Goal 2: Provide efficient computing infrastructure to meet NASA's mission needs

Goal 3: Accelerate scientific productivity through AI, quantum computing, and workforce development

Together, these goals will create a more efficient research environment that maximizes NASA's scientific impact and makes scientific assets more accessible and actionable for both researchers and industry partners.

STRATEGIES TO ACHIEVE THESE GOALS

To accomplish these goals, SMD will employ several key strategies: embracing standards and interoperable systems, developing AI-ready workflows, and establishing transparent governance frameworks. These strategies will enhance SMD's capacity to pursue high-risk, impactful research while ensuring efficient resource allocation and broad accessibility of NASA's scientific assets.



GOAL 1: BUILD AN OPEN, INTEROPERABLE ECOSYSTEM FOR SCIENTIFIC DISCOVERY AND ECONOMIC OPPORTUNITY

Create productive systems, efficient tools, and practices that maximize collaboration, minimize administrative burdens, and enable humans and machines to find, access, and accurately use NASA data.

Strategy 1.1: Provide SMD divisions with standardized approaches for data management and software preservation that prioritize collaboration, discoverability, interoperability, reusability, and commercial potential.

Strategy 1.2: Establish consistent metadata standards, documentation requirements, and levels of service framework to ensure appropriate resource allocation based on scientific and economic value.

Strategy 1.3: Create standardized interface designs, single sign-on capabilities, and a clear authoritative pathway for datasets to simplify discovery and use of data for both humans and machines.

Strategy 1.4: Establish transparent governance frameworks for data, software, and artificial intelligence lifecycles that include prioritization criteria based on scientific and economic value to ensure accountability, sustainability, and proper stewardship of scientific assets.

NASA's scientific mission is fundamentally tied to its ability to collect, manage, and share scientific data with researchers, commercial entities, and the public. Building on the agency's commitment to a "one-data culture free of siloes" and Findable, Accessible, Interoperable, Reusable, Understandability, Security, and Trustworthiness (FAIRUST) principles (NASA, 2021), this goal focuses on establishing an integrated scientific data and software ecosystem that is in compliance with the Open, Public, Electronic, and Necessary (OPEN) Government Data Act,⁶ the Evidence Act,⁷ and the Executive Order on Restoring Gold Standard Science.⁸

Despite improvements over the past five years, researchers still face friction when working with NASA resources, reducing scientific output and limiting potential applications. The challenges are complex and deeply embedded into the existing data ecosystem: fragmented interfaces built on legacy on-premises mission processing, inconsistent metadata (reflecting its labor-intensive nature), duplicative organizational structures, and varied access protocols. These are not simple technical problems; they require coordinated transformation across the complex organizational landscape spanning multiple science divisions, data repositories, and NASA-funded missions with varying data delivery timelines and requirements. This goal addresses these issues with standardization of interfaces, metadata, infrastructure, and transparent governance frameworks that prioritize high-value assets and will enhance scientific productivity while creating commercial opportunities.

⁶ U.S. Congress. (2019). [Open, Public, Electronic, and Necessary Government Data Act](#) (OPEN Government Data Act).

⁷ U.S. Congress. (2019). [Foundations for Evidence-Based Policymaking Act of 2018](#).

⁸ Executive Order No. 14303. (2025, May 28). "[Restoring Gold Standard Science](#)". Federal Register, 90 FR 22601.



GOAL 2: PROVIDE EFFICIENT COMPUTING INFRASTRUCTURE TO MEET NASA'S MISSION NEEDS

Maintain existing capabilities while evolving computational infrastructure and advanced analysis platforms to deliver cost-effective solutions that increase productivity and accelerate scientific discovery. This dual focus on efficiency addresses both scientific productivity and cost-effectiveness.

Strategy 2.1: Develop clear decision frameworks to guide researchers in selecting appropriate computing resources, such as HEC, cloud, or hybrid, based on cost-efficiency and reproducibility standards.

Strategy 2.2: Create integrated science platforms that collocate data and computing resources to facilitate complex workflows, with priority given to high-value datasets, and enable collaborative research across scientific and industry boundaries.

Strategy 2.3: Enable seamless integration across computing platforms and environments to prevent technological lock-in, ensure long-term sustainability, and maximize economic impact.

Access to scientific computing is essential to achieve NASA's mission and requires strategic evolution to meet rapidly expanding demands. This capability provides the foundational infrastructure necessary for maintaining U.S. leadership in AI technologies and space exploration, continuing NASA's investment in AI-optimized technology, and enabling breakthroughs across scientific disciplines. A strategic approach to computing resource evolution aligns with federal mandates for modernized infrastructure,⁹ data-driven decision making,¹⁰ and scientific reproducibility.¹¹

Despite advances in open-source tools and cloud technologies over the past five years, computational demands continue to grow in both scale and complexity, requiring access to specialized resources for emerging scientific methods like AI and advanced modeling. Researchers face multiple barriers to productive science: limited access to computational resources co-located with data (forcing inefficient transfers), fragmented computing environments, and insufficient guidance for selecting optimal resources. This goal addresses these challenges through strategies that streamline scientific workflows, enhance research productivity, and maintain America's leadership in scientific computing and AI.

⁹ Executive Order No. 14320, "[Promoting the Export of the American AI Technology Stack](#)," Federal Register, Vol. 90, No. 142 (July 28, 2025), 35393.

¹⁰ U.S. Congress. (2019). [Foundations for Evidence-Based Policymaking Act of 2018](#).

¹¹ Executive Order No. 14303. (2025, May 28). "[Restoring Gold Standard Science](#)". Federal Register, 90 FR 22601.



GOAL 3: ACCELERATE SCIENTIFIC PRODUCTIVITY THROUGH AI, QUANTUM COMPUTING, AND WORKFORCE DEVELOPMENT

Design systems, processes, and training to enable advanced analytical methods, ensuring that data are machine-readable, well-documented, and structured to support AI applications while maintaining scientific integrity, explainability, and reproducibility.

Strategy 3.1: Support developing open-source models and tools specifically designed for NASA's scientific data, co-created with domain experts to ensure scientific validity, technological leadership, and economic potential.

Strategy 3.2: Establish targeted training programs and educational resources to build analytical capabilities and expertise within the scientific community and broader workforce.

Strategy 3.3: Foster communities of practice across disciplines and sectors to share knowledge, identify emerging needs, and accelerate adoption of new technologies that advance both science and economic opportunity.

Strategy 3.4: Ensure scientific reproducibility and transparency in AI-driven and Quantum-driven research and other complex methods through explainable models, verifiable workflows, and comprehensive documentation.

NASA's pursuit of "high-risk, long-lead science and technology challenges" represents its greatest contribution to the nation (NASA at a Crossroads, [2024](#)). This goal supports requirements for enhancing data-driven decision making through advanced analytical techniques,¹² the mandate for machine-readable data formats,¹³ principles of scientific transparency and reproducibility in computational methods,¹⁴ mandates to extend American leadership in AI,¹⁵ responsible innovation that maintains American technological leadership,¹⁶ and accelerating federal AI use and acquisition.¹⁷

The computational landscape is evolving at unprecedented speed. Scientists are developing new AI and machine learning approaches specifically and investigating quantum algorithms for previously intractable problems. Scientists trained in traditional research methods must now navigate a complex ecosystem of models, tools, and methodologies without clear guidance on best practices. This goal addresses these challenges by supporting open-source scientific AI development, establishing training programs, fostering cross-disciplinary communities of practice, and ensuring AI applications maintain rigorous scientific standards.

¹² U.S. Congress. (2019). [Foundations for Evidence-Based Policymaking Act of 2018](#).

¹³ U.S. Congress. (2019). [Open, Public, Electronic, and Necessary Government Data Act](#) (OPEN Government Data Act).

¹⁴ Executive Order No. 14303. (2025, May 28). "[Restoring Gold Standard Science](#)". Federal Register, 90 FR 22601

¹⁵ Executive Order No. 14320, "[Promoting the Export of the American AI Technology Stack](#)," Federal Register, Vol. 90, No. 142 (July 28, 2025), 35393.

¹⁶ Executive Office Of The President Office of Management and Budget (2025). Memorandum M-25-21, "[Accelerating the Use of AI in Federal Agencies](#)," April 3, 2025.

¹⁷ Executive Office Of The President Office of Management and Budget (2025). Memorandum M-25-22, "[Driving Efficient Acquisition of Artificial Intelligence in Government](#)," April 3, 2025.

FINDINGS AND RECOMMENDATIONS

The future of exploration and scientific discovery at NASA depends on transforming how SMD's rapidly expanding data resources are managed, processed, and analyzed for insights. Through extensive analysis and consultation with stakeholders across all science divisions, this strategy identifies opportunities to evolve SMD's data and computing ecosystem into an efficient infrastructure that accelerates discovery and maximizes return on investment.

There are both significant strengths in SMD's current approaches and key areas requiring strategic evolution. The detailed findings and recommendations in [Appendix C](#) address six critical domains that will transform SMD's capabilities:

1. **Infrastructure and Management** - Establishing common frameworks, standardized interfaces, and authoritative data sources to reduce fragmentation and increase efficiency.
2. **Standards and Accessibility** - Implementing consistent metadata practices and optimized data formats to enhance discoverability and usability by humans and machines.
3. **Computation and Analysis Environments** - Integrating data and computing resources with analytics to accelerate scientific workflows.
4. **Artificial Intelligence and Advanced Information Processing** - Developing open-source foundation models with scientific validation protocols while establishing frameworks for explainability and reproducibility.
5. **Documentation and Training** - Building standardized educational materials and comprehensive training programs to empower researchers with necessary skills.
6. **Governance and Strategy** - Implementing metrics frameworks, supporting scientific software, and establishing ongoing community engagement processes.

These recommendations build upon the three strategic goals outlined earlier: building an open, interoperable ecosystem; optimizing computing infrastructure; and accelerating breakthroughs through AI, quantum computing, and workforce development. By implementing these recommendations, SMD will create a more efficient research environment that reduces technological barriers, enhances scientific productivity, and delivers greater economic value to the American people. Ultimately, these advances in data and computing capabilities will accelerate the scientific discoveries and technological innovations essential for NASA's exploration of Earth, the Moon, Mars, and beyond.

Table 1. SDMC30's vision, mission, goals, strategies, and recommendations. High-level goals cascade into specific strategies and recommendations, creating a framework to evolve SMD's data and computing ecosystem.

<p>Vision: Accelerating science through world-class data, computing, and AI capabilities that empower NASA scientists and partners to move from data to discovery faster.</p> <p>Mission: To provide seamlessly integrated data, computing systems, and advanced analytical tools that accelerate discovery, enable exploration, and maximize the return on America's investment in Earth and space sciences.</p>		
<p>Goal 1: Build an Open, Interoperable Ecosystem for Scientific Discovery and Economic Opportunity</p>	<p>Goal 2: Provide Efficient Computing Infrastructure to Meet NASA's Mission Needs</p>	<p>Goal 3: Accelerate scientific productivity through AI, quantum computing, and workforce development</p>
<p>Strategy 1.1: Provide SMD divisions with standardized approaches for data management and software preservation that prioritize collaboration, discoverability, interoperability, reusability, and commercial potential. Recommendations: 1a, 1e, 6b, 6c.</p>	<p>Strategy 2.1: Develop clear decision frameworks to guide researchers in selecting appropriate computing resources, such as HEC, cloud, or hybrid, based on cost-efficiency, and reproducibility standards. Recommendations: 3b, 3c.</p>	<p>Strategy 3.1: Support developing open-source models and tools specifically designed for NASA's scientific data, co-created with domain experts to ensure scientific validity, technological leadership, and economic potential. Recommendations: 4a, 4e.</p>
<p>Strategy 1.2: Establish consistent metadata standards, documentation requirements, and levels of service framework to ensure appropriate resource allocation based on scientific and economic value. Recommendations: 2a, 2d, 2e.</p>	<p>Strategy 2.2: Create integrated science platforms that collocate data and computing resources to facilitate complex workflows, with priority given to high-value datasets, and enable collaborative research across scientific and industry boundaries. Recommendations: 2b, 2c, 3d.</p>	<p>Strategy 3.2: Establish targeted training programs and educational resources to build analytical capabilities and expertise within the scientific community and broader workforce. Recommendations: 4e, 5a, 5b.</p>
<p>Strategy 1.3: Create standardized interface designs, single sign-on capabilities, and a clear authoritative pathway for datasets to simplify discovery and use of data for both humans and machines. Recommendations: 1b, 1c, 1f, 3a.</p>	<p>Strategy 2.3: Enable seamless integration across computational layers to prevent technological lock-in, ensure long-term sustainability, and maximize economic impact. Recommendations: 3a, 3d, 4d.</p>	<p>Strategy 3.3: Foster communities of practice across disciplines and sectors to share knowledge, identify emerging needs, and accelerate adoption of new technologies that advance both science and economic opportunity. Recommendations: 4e, 6e.</p>
<p>Strategy 1.4: Establish transparent governance frameworks for data, software, and artificial intelligence lifecycles that include prioritization criteria based on scientific and economic value to ensure accountability, sustainability, and proper stewardship of scientific assets. Recommendations: 1d, 6a, 6d, 6e.</p>		<p>Strategy 3.4: Ensure scientific reproducibility and transparency in AI-driven and Quantum-driven research and other complex methods through explainable models, verifiable workflows, and comprehensive documentation. Recommendations: 4b, 4c, 4e.</p>

Table 2. SDMC30’s vision, mission, areas, and recommendations. The strategic recommendations are organized into six implementation areas that provide clear action pathways for transforming NASA’s data ecosystem. Together, these areas address infrastructure modernization, standardization, computing optimization, AI advancement, community enablement, and governance—all essential components for realizing the strategy’s vision of seamlessly integrated, AI-ready data ecosystems that accelerate scientific discovery.

<p>Vision: Accelerating science through world-class data, computing, and AI capabilities that empower NASA scientists and partners to move from data to discovery faster.</p> <p>Mission: To provide seamlessly integrated data, computing systems, and advanced analytical tools that accelerate discovery, enable exploration, and maximize the return on America’s investment in Earth and space sciences.</p>					
Area 1: Infrastructure and Management	Area 2: Standards and Accessibility	Area 3: Computation and Analysis Environments	Area 4: Artificial Intelligence and Quantum Information Processing	Area 5: Documentation and Training	Area 6: Governance and Strategy
<p>Recommendations:</p> <p>1a: Common software frameworks</p> <p>1b: Standardized interface designs</p> <p>1c: Trusted data sources</p> <p>1d: Model and simulation output preservation</p> <p>1e: Software preservation catalog</p> <p>1f: Single sign-on service</p>	<p>Recommendations:</p> <p>2a: Levels of Service framework</p> <p>2b: Data migration prioritization</p> <p>2c: Optimized Data Access and Formats</p> <p>2d: Minimum metadata standards</p> <p>2e: Metadata tools</p>	<p>Recommendations:</p> <p>3a: Seamless collaboration</p> <p>3b: Compute capabilities process</p> <p>3c: HEC/cloud resource optimization</p> <p>3d: Data and analytics platforms</p>	<p>Recommendations:</p> <p>4a: Co-develop AI models</p> <p>4b: Explainable AI models</p> <p>4c: AI workflow best practices</p> <p>4d: Vendor-agnostic AI infrastructure</p> <p>4e: Quantum Computing</p>	<p>Recommendations:</p> <p>5a: Standardized educational materials</p> <p>5b: Strategic training program</p>	<p>Recommendations:</p> <p>6a: Comprehensive metrics framework</p> <p>6b: Supporting scientific software</p> <p>6c: Modern development tools</p> <p>6d: Streamlined software release process</p> <p>6e: Mid-cycle workshops</p>

ACKNOWLEDGEMENTS

This document was created with assistance from ChatGSFC, NASA Goddard Space Flight Center's AI assistant platform. ChatGSFC was used to combine, clarify, and summarize information. The content has been reviewed and edited by the team (Table 3).

Table 3. Strategy Development Team

Name	Division	Role/Affiliation
Chelle Gentemann	OCSDO	Chair - Steering Team
Kevin Murphy	OCSDO	Steering Team
Andrew Mitchell	OCSDO	Steering Team
J.L. Galache	OCSDO, former	Steering Team
Holly Norton	OCSDO, former	Steering Team
Emily Kosmaczewski	OCSDO, former	Steering Team
Bill Miller	NSF	Observer
Alessandra Aloisi	Astrophysics Division	Team
Sanaz Vahidinia	Astrophysics Division	Team
Amanda Saravia-Butler	Biological and Physical Sciences Division	Team
Shawn Reagan	Biological and Physical Sciences Division	Team
Jim O'Sullivan	Earth Science Division	Team
Patrick Quinn	Earth Science Division	Team
Jared Bell	Heliophysics Division	Team
Rebecca Ringuette	Heliophysics Division	Team
Rebekah Dawson-Rigas	Planetary Science Division	Team
Robin Fergason	Planetary Science Division	Team
Amanda Adams	OCSDO	Communications Lead
Michael Allen	OCSDO	Report Formatting

REFERENCES

- Congress (2019) "H.R.4174 - 115th Congress (2017-2018): Foundations for Evidence-Based Policymaking Act of 2018." Congress.gov, Library of Congress, 14 January 2019, <https://www.congress.gov/bill/115th-congress/house-bill/4174>.
- Congress (2018) "H.R.5515 - 115th Congress (2017-2018): John S. McCain National Defense Authorization Act for Fiscal Year 2019". Pub. L. No. 115-232, 132 Stat. 1636. <https://www.congress.gov/115/bills/hr5515/BILLS-115hr5515enr.pdf>.
- Congress (2017) "H.R.1770 - 115th Congress (2017-2018): Open, Public, Electronic, and Necessary Government Data Act or the OPEN Government Data Act, Congress.gov, Library of Congress, 29 March 2017, <https://www.congress.gov/bill/115th-congress/house-bill/1770>.
- Executive Office Of The President Office of Management and Budget (2025). Memorandum M-25-21, "Accelerating the Use of AI in Federal Agencies," April 3, 2025, <https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf>.
- Executive Office Of The President Office of Management and Budget (2025). Memorandum M-25-22, "Driving Efficient Acquisition of Artificial Intelligence in Government," April 3, 2025, <https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-22-Driving-Efficient-Acquisition-of-Artificial-Intelligence-in-Government.pdf>.
- Executive Order No. 14179. (2025, Jan 31) "Removing Barriers to American Leadership in Artificial Intelligence", Federal Register, 90 FR 8741. <https://www.federalregister.gov/documents/2025/01/31/2025-02172/removing-barriers-to-american-leadership-in-artificial-intelligence>.
- Executive Order No. 14303. (2025, May 28). "Restoring Gold Standard Science". Federal Register, 90 FR 22601. <https://www.federalregister.gov/documents/2025/05/29/2025-09802/restoring-gold-standard-science>.
- Executive Order No. 14320. (2025, July 23). "Promoting The Export Of The American AI Technology Stack". Federal Register, 90 FR 35393. <https://www.federalregister.gov/documents/2025/07/28/2025-14218/promoting-the-export-of-the-american-ai-technology-stack>.
- NASA (2026) NASA Science Mission Directorate's FY2024 Science Data Repository Metrics Report, <https://doi.org/10.64631/BJDW5471>.
- NASA (2025) NASA Science Mission Directorate 2024 NASA open-source Science Data Repositories Workshop Report, <https://assets.science.nasa.gov/content/dam/science/cds/about-us/ocsdo/reports/2024%20NASA%20Open%20Source%20Science%20Data%20Repositories%20Workshop%20Report.pdf>.
- NASA (2025) 2025-2026 NASA Science Plan: A Vision for Scientific Excellence. <https://science.nasa.gov/about-us/science-strategy/>.

- NASA (2024) NASA Office of Inspector General (OIG) Report IG-24-009, “NASA OIG Audit of NASA’s High-End Computing Capabilities”, March 14, 2024, <https://oigforms.nasa.gov/docs/IG-24-009.pdf>.
- NASA (2024) NASA Office of Inspector General (OIG) Report IG-24-019, “Evaluation of NASAs Information Security Program for Fiscal Year 2024”, September 12, 2024, <https://oig.nasa.gov/wp-content/uploads/2024/09/ig-24-019.pdf>.
- NASA (2024) NASA Science Mission Directorate Artificial Intelligence Workshop Report, <https://ntrs.nasa.gov/citations/20240015462>.
- NASA (2024) NASA Science Mission Directorate Data and Computing Architecture Study, <https://science.nasa.gov/wp-content/uploads/2024/08/data-and-computing-architecture-study-final-report-aug-2024.pdf>.
- NASA (2023) NASA Office of Inspector General (OIG) Report IG-23-006, “NASA Federal Information Security Modernization Act of 2014 Evaluation Report for Fiscal Year 2022”, <https://oigforms.nasa.gov/docs/IG-23-006.pdf>.
- NASA (2023) NASA Office of Inspector General (OIG) Report IG-23-012, “NASA's Management of Its Artificial Intelligence Capabilities”, <https://oig.nasa.gov/wp-content/uploads/2023/12/ig-23-012.pdf>.
- NASA (2023) NASA Office of Inspector General (OIG) Report IG-23-017, “NASA Federal Information Security Modernization Act of 2014 Evaluation Report for Fiscal Year 2023”, <https://oig.nasa.gov/wp-content/uploads/2024/02/ig-23-017.pdf>.
- NASA (2023) NASA Procedural Requirements (NPR) 2210.1E, June 14, 2023, https://nodis3.gsfc.nasa.gov/displayDir.cfm?Internal_ID=N_PR_2210_001E_&page_name=main.
- NASA (2023) NASA Science Mission Directorate Data Repositories Workshop Informal Report, <https://zenodo.org/records/11206267>.
- NASA (2021) NASA Data Strategy. Retrieved February 13, 2025, from https://www.nasa.gov/wp-content/uploads/2023/02/nasa_data_strategy.pdf.
- NASA (2020) NASA Office of Inspector General (OIG) Report IG-20-011, “NASA’S Management of Distributed Active Archive Centers”, <https://oig.nasa.gov/wp-content/uploads/2024/02/IG-20-011.pdf>.
- NASA (2019) Science Mission Directorate's Strategy for Data Management and Computing for Groundbreaking Science 2019–2024, https://science.nasa.gov/wp-content/uploads/2023/05/SDMWGStrategy_Final-1.pdf.
- NASA (2017) Ad-Hoc Task Force on Big Data of the NASA Advisory Council Science Committee Meeting Minutes June 28-30, 2016 NASA Goddard Space Flight Center. https://science.nasa.gov/wp-content/uploads/2023/05/BDTFJune2016_minutes_finalTAGGED.pdf.

- National Academies of Sciences, Engineering, and Medicine (2020) Progress Toward Implementation of the 2013 Decadal Survey for Solar and Space Physics: A Midterm Assessment. Washington, DC: The National Academies Press. <https://doi.org/10.17226/25668>.
- National Academies of Sciences, Engineering, and Medicine (2021) Planning the future Space Weather Operations and Research Infrastructure: Proceedings of a Workshop. Washington, DC: The National Academies Press. <https://doi.org/10.17226/26128>.
- National Academies of Sciences, Engineering, and Medicine (2022) Automated Research Workflows for Accelerated Discovery: Closing the Knowledge Discovery Loop. Washington, DC: The National Academies Press. <https://doi.org/10.17226/26532>.
- National Academies of Sciences, Engineering, and Medicine (2023) Origins, Worlds, and Life: A Decadal Strategy for Planetary Science and Astrobiology 2023-2032. Washington, DC: The National Academies Press. <https://doi.org/10.17226/26522>.
- National Academies of Sciences, Engineering, and Medicine (2023) Pathways to Discovery in Astronomy and Astrophysics for the 2020s. Washington, DC: The National Academies Press. <https://doi.org/10.17226/26141>.
- National Academies of Sciences, Engineering, and Medicine (2023) Thriving in Space: Ensuring the Future of Biological and Physical Sciences Research: A Decadal Survey for 2023-2032. Washington, DC: The National Academies Press. <https://doi.org/10.17226/26750>.
- National Academies of Sciences, Engineering, and Medicine (2024) Accelerating and Deepening Approaches to FAIR Data Sharing: Proceedings of a Workshop, in Brief. Washington, DC: The National Academies Press. <https://doi.org/10.17226/27786>.
- National Academies of Sciences, Engineering, and Medicine (2024) AI for Scientific Discovery: Proceedings of a Workshop. Washington, DC: The National Academies Press. <https://doi.org/10.17226/27457>.
- National Academies of Sciences, Engineering, and Medicine (2024) NASA at a Crossroads: Maintaining Workforce, Infrastructure, and Technology Preeminence in the Coming Decades. Washington, DC: The National Academies Press. <https://doi.org/10.17226/27519>.
- National Academies of Sciences, Engineering, and Medicine (2024) Promoting Equitable and Inclusive Implementation of Open Scholarship Policies: Proceedings of a Workshop, in Brief. Washington, DC: The National Academies Press. <https://doi.org/10.17226/27801>.
- National Academies of Sciences, Engineering, and Medicine (2024) Thriving on Our Changing Planet: A Midterm Assessment of Progress Toward Implementation of the Decadal Survey. Washington, DC: The National Academies Press. <https://doi.org/10.17226/27743>.
- National Academy of Sciences (2024) Toward a New Era of Data Sharing: Summary of the US-UK Scientific Forum on Researcher Access to Data. Washington, DC: The National Academies Press. <https://doi.org/10.17226/27520>.

National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). U.S. Department of Commerce.
<https://doi.org/10.6028/NIST.AI.100-1>

Office of Management and Budget. (2025). *Accelerating federal use of AI through innovation, governance, and public trust* (Memorandum M-25-21). <https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf>

Office of Management and Budget. (2025). *Driving efficient acquisition of artificial intelligence in government* (Memorandum M-25-22). <https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-22-Driving-Efficient-Acquisition-of-Artificial-Intelligence-in-Government.pdf>

Planetary Data Ecosystem Independent Review Board (2021) Final Report.
<https://science.nasa.gov/planetary-science/data/pde-irb/>.

White House Office of Science and Technology Policy, “Memorandum for the Heads of Executive Departments and Agencies, Agency Guidance for Implementing Gold Standard Science in the Conduct & Management of Scientific Activities”, June 23rd 2025,
<https://www.whitehouse.gov/wp-content/uploads/2025/03/OSTP-Guidance-for-GSS-June-2025.pdf>

APPENDIX A: STATUS OF 2019–2024 STRATEGY RECOMMENDATIONS

Status current as of June 2, 2025

Area 1: Open Data/Open Software

Recommendation 1a: NASA Science should develop a standard open data policy for all new missions, and current operating missions also should be encouraged to follow the policy. For new missions, the data requirements should be documented in the standard AO and SALMON templates, and compliance should be evaluated as part of PLRAs and appropriate lifecycle reviews. NPR 7120.5, and other NASA guiding documents, should be updated to include data systems in the compliance matrix.

Status 1a: **Completed** NASA's Science Mission Directorate Policy Document-41a ([SPD-41a](#)) became effective on December 22, 2022. This applies to all new research proposed from ROSES-2023 onwards and all existing missions that are at pre-Key Decision Point-B as of March 2023. Work to incorporate SPD-41a open data policies into NPR 7120.5 is ongoing.

Recommendation 1b: Standard open data requirements should be included in ROSES for all new solicitations and subsequent award requirements. These requirements should reflect the most current guidance and consider the recommendation from the NASEM report.

Status 1b: **Completed** NASA's [SPD-41a: Scientific Information Policy for the Science Mission Directorate](#) became effective on December 22, 2022. This policy reflects the most current OSTP and NASEM report recommendations.

Recommendation 1c: Missions should take a lifecycle approach in planning how their data will be managed for long-term curation after KDP-F. The Earth Science Division's data management schedule provides a reference for how this can be done and should be used as a template as each division develops their own approach as is available at: <https://earthdata.nasa.gov/collaborate/new-missions>.

Status 1c: **Completed** NASA's [SPD-41a: Scientific Information Policy for the Science Mission Directorate](#) became effective on December 22, 2022. This applies to all existing missions that are at pre-Key Decision Point-B as of March 2023 and includes planning for long-term curation.

Recommendation 2a: SMD should adopt a requirement for all new software development to be open-source, except in instances of ITAR, EAR, national security, PII, or other similar restrictions. This is consistent with OMB-M-13-13. Any waiver to not having publicly accessible software will need to be approved by the Deputy Associate Administrator for Research.

Status 2a: **Completed** NASA's [SPD-41a: Scientific Information Policy for the Science Mission Directorate](#) became effective on December 22, 2022, requiring all new software to be open-source except in instances of ITAR, EAR, national security, PII, or other similar restrictions. This applies to all new software proposed from ROSES-2023 onwards.

Recommendation 2b: Any software developed under the ROSES NRA must be released as open-source Software (OSS); opt-outs and alternatives must be justified and approved by the Deputy Associate Administrator for Research. The Earth Science Division currently has an OSS policy that could be adapted for a Directorate-wide policy. This software policy must be added to the standard ROSES language. NPR 7120.8, and other NASA guiding documents, should be updated to include software in the compliance matrix.

Status 2b: **Completed** NASA's [SPD-41a: Scientific Information Policy for the Science Mission Directorate](#) became effective on December 22, 2022, requiring that all new software proposed under the ROSES be open-source from ROSES-2023 onwards.

Recommendation 2c: For new missions, an open-source requirement must be included in the standard AO and SALMON language, and compliance should be evaluated as part of PLRAs and appropriate lifecycle reviews. For missions already in operation, software accessibility should be part of the senior review process. NPR 7120.5, as well as the PE Handbook and relevant mission documents, should be updated to include software in the compliance matrix.

Status 2c: **Completed** NASA's [SPD-41a: Scientific Information Policy for the Science Mission Directorate](#) became effective on December 22, 2022. This applies to all new software for existing missions that are at pre-Key Decision Point-B as of March 2023. Work to incorporate SPD-41a open software policies into NPR 7120.5 is ongoing. More mission specific guidance is online at <https://github.com/nasa/smd-open-science-guidelines/>.

Area 2: High-End Computing (HEC) Program

Recommendation 3: SMD should conduct assessments no less than every five years for high end computational resources. These assessments should evaluate HEC capacity, computing needs, and allocations across science divisions. This information should be used to develop a strategy for tracking and allocating cycle time for NASA resources and exploring whether there is the demand for additional or new computational resources to support SMD's objectives. The output of this assessment will also be used as part of Agency-level planning through the Office of Strategic Infrastructure's Shared Capability Assets Program to determine if additional resources are required.

Status 3: **Completed** Recommendation 3 was addressed. SMD conducted a HEC Tiger Team Study under the auspices of the Office of the Chief Science Data Officer. The High-End Computing (HEC) Capability Strategic Working Group 2 Recommendations and Findings in response to the 2024 Inspector General's report ([IG-24-009](#)) has been completed and its recommendations approved by the NASA AA in January 2025. It is not public.

Area 3: Archives Modernization

Recommendation 4a: Digital data derived from NASA-funded research is required to go to a NASA archive for long-term curation and public availability. There are shared responsibilities between the principal investigators (PIs) and the archives to ensure transition of data. Archives must provide clear guidance on requirements to ingest data and investigators must adhere to the requirements. The Earth Science Division provides one example of such guidance at <https://earthdata.nasa.gov/collaborate/new-missions>. Additionally, the current archives must

prepare to receive and validate data from heterogeneous sources, which may involve more “hands-on” work with PIs submitting data, metadata, and documentation. Such additional work will require additional funding, which should be addressed in the PPBE process.

Status 4a: **Completed**

In addition to ESD, the remaining SMD Divisions also maintain pages with information for data providers on how to process and deliver data to their respective repositories:

- APD: <https://science.nasa.gov/astrophysics/data/>
- BPS: <https://www.nasa.gov/osdr-submission-portal/>
- HPD: https://spdf.gsfc.nasa.gov/archive_newdata_req.html
- PDS: <https://pds.nasa.gov/home/providers/>
- ESD: <https://www.earthdata.nasa.gov/engage/data-management-guidance>

Additionally, SMD provides public guidance: <https://github.com/nasa/smd-open-science-guidelines/>

Recommendation 4b: SMD should include an evaluation criterion for future ROSES solicitations to assess the adequacy of data management plans associated with individual proposals to support delivery of data to a final archive. SMD should evaluate whether an augmentation of funding may be required, either at the award or program element level, to enable this. As this becomes the norm, SMD should reassess the level of funding required.

Status 4b: **Completed** As of 2023, most ROSES elements require an OSDMP (Open Science Data Management Plan). Guidance for evaluation of OSDMPs and a list of preferred archives [was provided for ROSES 2023](#). In FY22-24, the F.8 Supplements for Open-Source Science (SOSS) ROSES element provided additional funding to existing projects that may need resources to comply with SPD-41a; all new projects are expected to include these costs in their budget. The mission OSDMP guidance was released 3/13/2025 and is public at https://assets.science.nasa.gov/content/dam/science/cds/open-science/2025-03-13_SMD-Mission-OSDMP-Template-v1.0.docx.

Recommendation 5: SMD should strongly encourage collaboration and cooperation of data professionals in academia, industry, and elsewhere to enable cross-cutting scientific discovery. There are currently small-scale efforts in the scientific communities to provide data stewardship training, including cleaning and curation, but these will need to be expanded to support anticipated future demand. This activity should be sustained to ensure that new techniques (science or data), methods and algorithms for analysis, and management of data can be incorporated as they are developed.

Status 5: **Completed** NASA has signed Space Act Agreements with AWS, Microsoft, and IBM, leaning on industry to help meet our milestones and goals. Through these agreements NASA has released or is developing several AI foundation models with Microsoft and IBM. To encourage collaboration and cooperation NASA has supported several [Openscape cohorts](#) in Earth Science and Hackathons ([OceanHackWeek](#), [Phase Change Phenomena in a Decarbonizing World](#)). Basic

data stewardship is also covered by NASA's [Open Science 101](#), released in December 2023, that over 4,000 scientists have completed. NASA has organized [several data repository workshops](#) to encourage collaboration between SMD repositories on common tools and lessons learned. Within the newly established OCSDO there are two core activities that support incorporation of new data science techniques (DASP) and development of common tools (Core Data and Computing Services).

Recommendation 6: SMD should communicate the value of this capability and work with the NASA Centers to increase their recruitment of data science professionals. While NASA is at a competitive disadvantage for these types of positions, there are unique opportunities provided to data science professionals that only exist in the NASA environment.

Status 6: In Progress NASA requested a report from NASEM: [NASA at a Crossroads: Maintaining Workforce, Infrastructure, and Technology Preeminence in the Coming Decades](#) (2024). NASA's [Data and Computing Architecture Study](#) (2024) highlights the need for increased data science capabilities within NASA. OCSDO has organized several community workshops for software developers and data scientists.

Recommendation 7: A process should be developed for data and software usability, discoverability, and accessibility to be evaluated as part of future senior reviews. This should also include utilization of advanced computing capabilities (e.g. high-end computing and commercial cloud computing). Information gathered through this process should be used to inform decision-making by the relevant archives.

Status 7: In Progress SPD-41a gives specific guidance on when and how data and software usability, discoverability, and accessibility will be evaluated. Data accessibility and usability is already part of mission senior review for the Astrophysics and Heliophysics divisions, likely other divisions too. HEC and cloud computing use by the missions (vs scientists analyzing the data) are part of mission operations and are reviewed as such. Archives work closely with the missions to scope e.g. capacity to serve anticipated data. A process has not been developed for software at this time.

Recommendation 8: In order to enable cross-disciplinary science, SMD must facilitate greater discoverability of like data holdings in various archives. SMD should engage with the science community to develop a metadata description that would cross science boundaries in order to facilitate these new areas of study.

Status 8: Completed NASA has developed and released the Science Discovery Engine ([SDE](#)) and Science Explorer ([SciX](#)) to provide a single search capability for NASA funded research, including publications, data, and software to enable cross-disciplinary science and discovery. With the SDE users can perform free-text searches to quickly find what they need within this vast collection of science data, documentation, and code. The SciX portal links users to scientific literature, data, and software in the fields of astronomy, Earth science, heliophysics, the biological and physical sciences, and planetary science. A cross-Divisional metadata standard has been a topic of discussion at the NASA Data Repository Workshops.

Recommendation 9: SMD should create a free and open, unified journal server along the lines of PubSpace, ADS or ERS to make science papers more accessible to the public. NASA Science

should also consider adopting the National Science Foundation’s requirements to reinstate the need for grant recipients to provide copies of all published research as part of their annual reports.

Status 9: Completed NASA funds SciX (the evolution of the ADS) that will search for publications in all NASA science domains (but doesn’t store publications themselves). SPD-41a requires all publications from NASA-funded science to be freely available, e.g., as preprints. NASA is not requiring grant recipients to provide copies of all published research as part of their annual reports, but does encourage all funded researchers to identify the correct funding sources at the time of publication and report publications in all grants required reporting.

Area 4: Advanced Capabilities

Recommendation 10: SMD should encourage the science divisions to explore novel computational techniques, including cloud computing and AI/ML, through various avenues, including ROSES NASA Research Announcements, technology calls in cooperation with the Office of the Chief Information Officer and Space Technology Mission Directorate, and mission Announcements of Opportunity.

Status 10: Completed SMD Divisions are exploring these opportunities with IMPACT (Interagency Implementation and Advanced Concepts Team) and releasing Division subject matter NASA Foundation Models. The OCSDO’s Data Analysis Services Program (DASP) will advance NASA’s strategic goal of developing and leading innovative and sustainable programs to enable open science through continuous evolution of science data and computing systems for SMD. Many ROSES calls already encourage the use of AI/ML in proposed projects. We are also making significant progress in hosting data and computation on the cloud through examples such as VEDA and NASA SMD data repositories. Community platforms on the cloud are also in initial stages of adoption as demonstrated by the landmark success of CryoCloud.

Recommendation 11: SMD should make investments to incentivize and educate the community on how to use AI/ML to approach science in new ways. Hands-on training can be achieved through expansion of hackathons, competitions, and grant programs. Science results and lessons learned about the use of AI/ML will be shared at community meetings to increase awareness of the potential of these techniques.

Status 11: Completed NASA has funded several TWSC workshops ([Workshop on Impacts of Unusual Weather Events and Climate Anomalies on a Tropical Rainforest](#)), TOPS ScienceCore, and F.14 High Priority open-source Science [projects](#), that focused on use of innovative data science techniques. The OCSDO organized the [2023 NASA open-source Science Data Repositories Workshop](#), [2024 NASA SMD AI Workshop](#), [2025 SPARK AI campaign](#), [2024 Software for NASA SMD Workshop](#), and the CryoCloud Workshop; it is also investing in additional hands-on training opportunities. The Center for HelioAnalytics provides AI/ML tutorials on the cloud through community workshops such as the biannual [PyHC Summer School](#), also funded with NASA resources.

Area 5: Management

Recommendation 12: SMD should appoint a directorate-level Scientific Data and Information Officer to oversee this effort and serve as an interface between the Office of the Chief Information Officer, Office of the Chief Scientist, international partners, commercial providers, and others. The Scientific Data and Information Officer’s responsibilities should include: 1. Setting policy for SMD, verifying compliance, and maintaining awareness of external policy drivers; 2. Maintaining a cross-divisional innovation forum to identify, select, and fund new opportunities; 3. Investing in targeted capabilities and tracking their progress; 4. Cultivating SMD’s data and computing community through workshops, studies, training, etc., and; 5. Conducting periodic independent evaluation of the structure and content of the SMD data and computing portfolio.

Status 12: Completed Kevin Murphy was appointed as NASA’s Chief Science Data Officer in March 2021 and stood up the Office of the Chief Science Data Office (OCSDO). The Divisions have each named a Data Officer to manage the science data needs of each Division and liaise with the repositories and the OCSDO.

APPENDIX B: LANDSCAPE ANALYSIS

To develop SDMC30, the strategy team conducted a comprehensive assessment of NASA's evolving scientific data and computing landscape since the publication of [SDMC24](#). This analysis examined organizational changes, technological developments, capability gaps, and emergent opportunities across the entire Science Mission Directorate (SMD) ecosystem. This assessment draws upon multiple authoritative sources including National Academy of Science, Engineering, and Medicine (NASEM) decadal surveys and reports, Office of the Inspector General (OIG) findings, community workshops, and detailed capability inventories (see [References](#)). By synthesizing insights from across SMD's five scientific disciplines, this section provides a clear picture of current capabilities, identifies critical gaps, and establishes the foundation for the strategy's forward-looking recommendations.

OCSDO Organizational Structure and Programs

As recommended by the [SDMC24](#), in 2021 NASA SMD established the Office of the Chief Science Data Officer (OCSDO) to advance NASA's unique science missions and enable groundbreaking discoveries through three core functions:

1. Increase efficiency and usability of scientific and engineering data and computing infrastructure through maintenance, operations, and continuous evolution.
2. Advance discovery through innovative applications and partnerships in data science, advanced analytics, quantum information processing, and artificial intelligence.
3. Enable transparent, trustworthy, and reproducible open science through guidance, training, and open-source tools.

Recognizing the interdependence of computation, data science, and scientific data in supporting NASA's scientific and engineering activities, OCSDO implements these functions through three programs and manages NASA's High-End Computing (HEC) resources:

- **Core Data and Computing Services (CDCS):** Provides a foundation for layered architecture enabling divisions to integrate discipline-specific services, including cloud infrastructure and high-end computing
- **Data Science and Innovation (DSI):** Develops governance frameworks, evaluates compliance, maintains data science capabilities, and provides research-to-operations frameworks for data science projects
- **Open Science Implementation (OSI):** Manages grant portfolio, coordinates science data policies, and provides guidance for alignment with open information policies

Current Challenges

NASA SMD currently manages over 150 Petabytes (PB) of observational data and simulation results across 10 repositories. Each of SMD's five divisions maintains its own data systems with varying budgets, repositories, data providers, user interfaces, and capabilities, a legacy of mission-specific needs that has created siloed infrastructures with overlapping functionalities

(see details in NASA Science Mission Directorate’s FY2024 Science Data Repository Metrics Report (2026)). As highlighted in the Data and Computing Architecture Study (2024), this volume is expected to increase exponentially, driven by upcoming missions and the expansion of scientific simulation capabilities.

While the data system fragmentation has greatly diminished with the implementation of SDMC24 (see [Appendix A](#)), it still complicates cross-disciplinary research, increases overall operational costs, creates unnecessary barriers to discovery, and use of NASA's data and software assets. While some divisions have created a single point of entry, others have a multitude of entry points. Useful software tools are often not clearly linked, and tutorials are developed inconsistently.

Multiple NASEM decadal surveys (see [References](#)) have consistently emphasized the need for improved data sharing, interoperability, standardization, and access. These surveys identify persistent barriers to cross-disciplinary research and recommend NASA develop best practices for data storage, common data formats, and tools to facilitate analysis across multiple archives. Similarly, the *Toward a New Era of Data Sharing* NASEM report (2024) highlighted that effective data sharing requires infrastructure, frameworks, standards, trained personnel, and sustainable funding streams.

The Data and Computing Architecture Study (2024) recommended developing a core architecture that organizes general Core Services across SMD Divisions while preserving data-management autonomy. It noted that a more coordinated approach could reduce redundancies in repository functions.

Similarly, the *Planning the Future Space Weather Operations and Research Infrastructure* (2021) workshop proceedings highlighted the need for centralized access points to platforms and associated tools, which would help reduce the current fragmentation across systems.

Computational Resource Limitations

Current computational resources are increasingly inadequate for cutting-edge scientific workloads, as documented in the Data and Computing Architecture Study (2024) and OIG findings:

- Modern Artificial Intelligence / Machine Learning (AI/ML) applications require approximately 5,000-10,000 times more computing power than NASA can currently deliver as an agency
- High-value scientific data analysis from billion-dollar missions exceeds available computational resources
- Scientific domains frequently must limit model resolution and complexity due to computational and data storage constraints

The 2024 NASA SMD AI Workshop Report (2024) emphasized foundation models' potential to transform scientific discovery but noted substantial computational requirements that exceed current capabilities. The workshop identified critical needs for infrastructure investments, specially tuned scientific foundation models, and AI-ready data architecture.

The Data and Computing Architecture Study (2024) found that "efficient data and computing capabilities that allow widespread collaboration will be critical to NASA's future groundbreaking discoveries." As NASA's data volumes continue to grow exponentially and analytical techniques become more computationally intensive, optimizing and collocating computing infrastructure with data resources becomes essential for enabling scientific discovery.

The 2024 Software for NASA SMD Workshop Report (2024) identified a "need for better guidance and exposure to cloud resources and High-Performance Computing (HPC)." NASA's computing ecosystem now spans traditional HEC facilities like the NASA Advanced Supercomputing (NAS) facility's HEC and cloud-based platforms through the Computational and Information Science and Technology Office's (CISTO) Science Managed Cloud Environment (SMCE), creating a complex decision landscape for researchers.

Multiple NASEM decadal surveys and reports have emphasized the importance of science platforms that allow researchers to analyze data without downloading it locally. As detailed in the Planetary Data Ecosystem Independent Review Board (IRB) Final Report (2021), these capabilities require specialized infrastructure that brings computing to the data rather than requiring data download—a capability currently limited by computational resource constraints.

Data Standards and Interoperability

The Toward a New Era of Data Sharing NASEM report (2024) highlighted that effective data sharing requires infrastructure, frameworks, standards, trained personnel, and sustainable funding streams. These elements are critical for ensuring interoperability across NASA's data systems.

The NASA SMD Data Repositories Workshop Informal Report (2024) specifically called for repositories to offer different Levels of Service with clear procedures for applying them to data. The workshop also identified the need for better coordination on metadata both to enable data discovery, interdisciplinary research, accurate use, and to adhere to federal requirements.

The NASA SMD Data Repositories Workshop Informal Report (2024) report identified critical needs for standardized service levels, clear procedures for accepting researcher-submitted data, persistent identifiers (PIDs) for simulation outputs, better coordination on metadata, and automated assessment of Findable, Accessible, Interoperable, and Reusable (FAIR) compliance.

There exists a need for clearer definition of metrics and capabilities to automatically collect user metrics across all the repositories and divisions. The NASA SMD Data Repositories Workshop Informal Report (2024) also identified the need for "automated assessment of how data complies with FAIR and tools to simplify the creation of FAIR data." By implementing standardized metrics collection, NASA can make data-driven decisions about resource allocation and identify high-value datasets and software that may benefit from enhanced services and support.

The Data and Computing Architecture Study (2024) noted that inconsistent implementation of data standards creates barriers to data discovery and reuse across disciplines. The study advocated for a core architecture that preserves domain expertise while enabling more efficient integration of services through standardized approaches.

Software Recognition and Support

Scientific software faces many of the same challenges as scientific data. The 2024 Software for NASA SMD Workshop Report ([2024](#)) emphasized that software has become a foundational component of NASA's scientific enterprise but lacks appropriate recognition and support infrastructure.

NASEM reports, including *Accelerating and Deepening Approaches to FAIR Data Sharing* ([2024](#)), have called for improved software preservation strategies and enhanced recognition of software as a critical scientific asset. Additional challenges were identified, including significant issues with software catalogs, release processes, and lifecycle management that impede NASA's ability to leverage software as a critical scientific asset.

The 2024 Software for NASA SMD Workshop Report ([2024](#)) emphasized the need for better "infrastructure for software lifecycle management" and improved support for community collaboration. The report characterized software as having an "organic nature not properly supported by current approaches."

Across multiple workshops and reports, there is consistent recognition that scientific software developed for NASA missions or research often lacks clear pathways for maintenance, documentation, and long-term preservation. Unlike data, which has established archival processes, software frequently exists without coordinated stewardship strategies, leading to duplication of effort and lost capabilities when projects end.

The Planetary Data Ecosystem Independent Review Board (IRB) Final Report ([2021](#)) noted that analysis-ready data products should be accessible through well-defined Application Programming Interfaces (APIs), highlighting the critical intersection between software development and data access.

Workforce Skills Gap

NASA's greatest value lies in tackling "high-risk, long-lead science and technology challenges" (NASA at a Crossroads, [2024](#)). Cultural change is hard and transforming scientific discovery requires both advanced technological capabilities and a prepared scientific community that can leverage these innovations effectively.

Modern scientific workflows increasingly require expertise spanning multiple domains: advanced computational methods, cloud infrastructure, data science, and AI. Rather than expecting every scientist to become an expert in all these areas, effective collaboration between domain scientists and computational experts is essential for identifying efficient approaches to complex problems and maximizing both scientific productivity and efficient resource utilization.

The expanding volume and complexity of NASA's data creates unprecedented opportunities for discovery through advanced computational methods, including AI. By co-creating tools with domain experts, NASA can ensure they address real scientific needs while maintaining the integrity of the scientific process. However, this requires mechanisms for collaboration between scientific domain experts and those with specialized expertise in advanced computational methods.

NASA scientists are already using AI and other complex computational approaches, creating an immediate need to establish frameworks and practices that ensure research maintains the rigor and transparency essential to the scientific method. This requires not just individual training, but also accessible pathways for scientists to consult with experts in computational optimization and advanced methods.

The 2024 Software for NASA SMD Workshop Report ([2024](#)) identified issues with "training" as one of the significant challenges facing NASA's software ecosystem. Enhancing computational literacy across disciplines—while facilitating expert collaboration—will ensure researchers can leverage advanced tools and methodologies efficiently.

Multiple NASEM decadal surveys and reports have emphasized the importance of training opportunities on modern data science techniques and complex software. The Planetary Data Ecosystem IRB Final Report ([2021](#)) specifically highlighted training as a critical need alongside technological infrastructure.

Fostering communities of practice that span traditional disciplinary boundaries, creating forums for knowledge sharing and collaboration, and establishing mechanisms that connect domain scientists with computational experts will accelerate the adoption of efficient computational approaches. This requires deliberate investment in workforce development strategies that emphasize collaboration and knowledge transfer rather than expecting every scientist to master every computational method.

APPENDIX C: DETAILED FINDINGS AND RECOMMENDATIONS

Area 1: Infrastructure and Management

Finding 1a (Common software frameworks): Maintaining ad-hoc repository software across science divisions is costly and inefficient, hindered by outdated systems and restrictive licenses. Modernizing these systems for ingest, organization, and management of data with AI tools and modular software solutions can facilitate collaboration, ease of use for contributors and users, and reduce maintenance burdens of time and cost. For instance, standardizing data submission and organization schemas for repository systems in the Biological and Physical Science (BPS) division resulted in increased efficiencies.

Recommendation 1a: SMD should encourage divisions to develop and implement common software frameworks that standardize repository services for data ingest, organization, and management while preserving science-specific functionality.

Finding 1b (Standardized interface designs): NASA scientists interact with multiple decentralized systems across 10 repositories and numerous data providers with inconsistent interfaces. While domain-specific customizations benefit specialized use cases, the proliferation of redundant implementations of common functions (e.g., search, access, visualization) has created challenges. This fragmentation creates cognitive burden, impedes interdisciplinary research, and hinders efficient data discovery, while creating maintenance challenges for redundant implementations of common functions. These challenges highlight the urgent need for consistent interface designs and consolidated implementation of common functions while preserving domain-specific capabilities where scientifically necessary.

Recommendation 1b: SMD should work with divisions to develop and implement standardized interface designs and interaction patterns for common repository functions (e.g., search, access, visualization - for both human and machine) and document criteria for determining which functions require customization for specialized scientific needs.

Finding 1c (Trusted data sources): NASA's data volume is projected to grow from 150 PB to 530 PB by 2030. Users and AI systems struggle to locate authoritative data sources due to fragmentation and inconsistent implementations across repositories and data providers, limiting accurate citation and use of data, efficient exploration, and commercial applications that require reliable, standardized access. By creating authoritative data sources with standardized ingest processes, analytics capabilities, and AI-friendly access patterns, SMD can improve data discoverability, citation accuracy, accurate use of data, and overall accessibility while facilitating scientific exploration and commercial innovation.

Recommendation 1c: SMD should work with divisions to establish an authoritative source for each dataset, with a documented plan for consolidating redundant repositories and migrating priority data to the SMD Science Cloud.

Finding 1d (Model and simulation output preservation): The current lack of systematic archival and preservation of SMD-funded computational models and their simulation outputs hampers transparency, reproducibility, reuse, and accessibility in scientific research. Although not all model and simulation outputs require archiving, NASA's absence of a structured archival system limits the potential for AI and other complex techniques to be applied to these outputs for advancements and adds an additional burden to NASA researchers to create transparent and reproducible results. Additionally, the inability to cite sizable and complex simulation outputs and accurately track their provenance further restricts their utility and academic recognition. An archival framework for model and simulation outputs would support future scientific advances but also boost the credibility and impact of NASA-funded research through improved data stewardship and efficient resource management.

Recommendation 1d: SMD should implement an archival framework for model and simulation outputs with clear preservation criteria and domain-specific lifecycle management policies.

Finding 1e (Software preservation catalog): Reusing existing software shortens development cycles and minimizes effort, but users are struggling to find and use NASA-funded software. NASA's software catalogs need improvement, consolidation, and a user-friendly search interface. There is no comprehensive centralized repository or search engine for scientific software, therefore researchers rely mainly on word of mouth or institutional knowledge to find new software and machines struggle to find NASA software at all. Additionally, it isn't clear to users what software is being maintained, so a basic level of curation should be required, possibly with additional higher levels for prioritized software.

Recommendation 1e: SMD should create a comprehensive catalog of SMD-funded scientific software source code with standardized metadata, maintenance status indicators, and search functionality.

Finding 1f (Single sign-on service): All SMD data is publicly available, but access to some data currently requires a user login (e.g., Earth data and other data within its proprietary period). Users find the current implementation of login requirements cumbersome, intrusive, and not secure. As SMD continues to migrate its data to the cloud and expand its cloud computing offerings, the need for cyber security and efficient compute resource allocation requires a way to identify and authorize users. Modernizing the user login experience with a combined single sign-on (SSO) with Phishing-Resistant Multifactor Authentication (MFA), where personal identifying information (PII) is managed according to gov requirements would address both the needs of the NASA science community and NASA data repositories.

Recommendation 1f: SMD should implement a single sign-on service using industry-standard solutions that can manage digital identities, authenticating users, including authorized users without NASA Personal Identity Verification (PIV) badges, and, when necessary, authorizing access to data and compute resources.

Area 2: Standards and Accessibility

Finding 2a (Levels of Service framework): Data archival is an indispensable element supporting both SMD-funded research and global scientific initiatives. However, it is neither practical nor resource-efficient for all SMD data to achieve uniformly high levels of quality, accessibility, and supporting services. Currently, there is a lack of a unified approach among SMD repositories to define and determine appropriate service tiers for each dataset and software based on their scientific value, usage patterns, and preservation requirements. Levels of Service refers to a tiered framework that defines different categories of support and capabilities provided to datasets based on their scientific value and usage patterns. For example, a top-tier "Platinum" service level might guarantee 99.9% uptime, multiple access methods (including cloud-optimized formats), comprehensive documentation, visualization tools, and rapid user support. In contrast, a baseline "Bronze" service level might ensure only basic access, minimal documentation, and standard preservation. By implementing such differentiated service tiers, repositories can make strategic decisions about where to invest limited resources. Without such a structured framework, repositories struggle to effectively allocate limited resources, communicate service expectations to users, and prioritize enhancements to high-impact datasets. A comprehensive Levels of Service approach will enable more strategic resource allocation, clearer communication with users about service expectations, and prioritization of enhancements for the most scientifically valuable assets.

Recommendation 2a: SMD should implement a common Levels of Service framework with clearly defined tiers, documented criteria for dataset classification, and specified requirements for each tier.

Finding 2b (Data migration prioritization): The migration to the cloud has been inconsistent across SMD divisions, creating disparities across both internal capabilities and external (user) experiences. Divisions that have migrated appropriate datasets to the cloud have found substantial efficiency gains through simpler, centralized data access, elimination of redundant services, and standardized discovery and visualization tools. However, not all data is optimal for cloud deployment. Large, infrequently accessed datasets may be more cost-effective on-premises, while computationally intensive workflows may benefit from HEC environments. A detailed assessment is needed to determine the optimal placement strategy for each dataset, considering factors such as access patterns, computational requirements, data transfer costs, and long-term storage economics. For researchers working with cloud-hosted data, the transformation from download-centric to compute-adjacent paradigms enables new possibilities—particularly for AI and agentic workflows requiring direct access to large datasets. The key is developing a hybrid approach that strategically places data where it can be most effectively accessed and analyzed, whether that's cloud, on-premises, or HEC environments. The capabilities being developed by OCSDO projects should support this flexibility through standardized access methods and interoperable resources that work across different infrastructure approaches, allowing researchers to seamlessly work with data regardless of its physical location.

Recommendation 2b: SMD should complete an infrastructure assessment of all scientific data assets and develop a prioritized migration roadmap based on scientific value, usage patterns, and potential for enabling new research capabilities.

Finding 2c (Optimized Data Access and Formats): Modern scientific workflows—particularly AI applications—require data formats and access patterns optimized for high-throughput, low-latency operations across diverse computing environments. The evolution from traditional download-centric models to direct, indexed access offers numerous advantages: chunked organization, parallel processing capabilities, efficient HTTP(S) access with range requests, seamless integration with object storage, rapid sub-setting through built-in indexes, and improved compression. However, NASA has not yet implemented these optimizations for most historical datasets, nor established requirements for future missions to address these emerging access patterns. This gap creates significant performance bottlenecks for data-intensive applications, forces inefficient data duplication by users requiring high-performance access, and limits NASA's flexibility to deploy data across different infrastructure environments based on evolving research needs.

Recommendation 2c: SMD should assess which datasets would benefit most from enhanced access methods and data formats to support modern scientific workflows, then develop a prioritized plan for implementing these improvements. Future missions should be encouraged to consider anticipated user workflows when designing data products and access methods.

Finding 2d (Minimum metadata standards): Users and machine-to-machine processes struggle with accurate use and citation of NASA science data, primarily due to inconsistent metadata application across divisions and classes. This leads to misinterpretation or erroneous results and a lack of proper recognition, which can decrease trust in science. Establishing a minimum set of metadata standards is essential to ensure citable, coherent and understandable resources across all divisions. This will allow for more predictable results, would enhance data integration and usability, meeting the foundational needs of the National Archives and Records Administration (NARA) while supporting advanced metadata systems. Establishing consistent metadata standards fulfills requirements under the [Foundations for Evidence-Based Policymaking Act](#), which mandates federal agencies to create comprehensive data inventories with appropriate metadata to facilitate data discovery and use.

Recommendation 2d: Working with the NASA Chief Data Officer (CDO), SMD and divisions should establish minimum metadata standards that apply to all scientific data, with compliance verification mechanisms and implementation guidance.

Finding 2e (Metadata tools): Creating high-quality metadata is labor-intensive, requiring significant expertise and time. This burden often leads to inconsistent implementation or minimal compliance rather than comprehensive description. While existing pockets of metadata excellence demonstrate the potential scientific value of well-described datasets, scaling these practices across all NASA data requires reducing the manual effort involved. Automated or semi-automated approaches that assist scientists and data managers in creating rich, structured metadata would significantly increase adoption while improving quality and consistency. Such tools are particularly important for enabling AI applications that depend on machine-readable contextual information about data provenance, quality, uncertainty, and appropriate use cases.

Recommendation 2e: SMD should develop metadata creation tools that automate or semi-automate the generation of standardized metadata for scientific datasets making it easier to implement metadata standards and capture the rich contextual information needed for both human and AI users.

Area 3: Computation and Analysis Environments

Finding 3a (Seamless collaboration): The NASA science community has access to multiple disconnected systems of data, software and compute that they use for their science, though access to computational resources is not equally available to all systems for all users. This fragmentation creates significant barriers to collaboration across computational environments, forcing scientists to work within silos determined by their access privileges and prior experience rather than by scientific need. Scientists need the ability to seamlessly collaborate and integrate computation, software and data across the full spectrum of computational resources—from low-computation personal systems to local clusters, HPCs, HEC systems, and cloud computing environments—into one fluid scientific exploration and collaboration workflow. The lack of such interoperable infrastructure impedes cross-disciplinary collaboration, limits scientific productivity, and forces researchers to spend valuable time on infrastructure management rather than focusing on science.

Recommendation 3a: SMD should implement an open-source integration layer that enables authorized users to seamlessly work across data and computational resources, with documented interoperability standards and authentication protocols.

Finding 3b (Compute capabilities process): NASA's HEC resources are oversubscribed and overburdened, while SMD lacks a comprehensive strategy for allocating these limited resources. The current first-come, first-served approach fails to align computational resources with NASA's highest-priority science objectives. Additionally, researchers face challenges identifying and accessing appropriate compute resources, particularly when deciding between on-premises HEC and cloud computing options, with limited understanding of the performance and cost implications of these choices.

Recommendation 3b: SMD should establish a priority-based compute allocation process with transparent criteria, application procedures, and resource distribution metrics.

Finding 3c (HEC/cloud resource optimization): Even when researchers gain access to appropriate computational resources, they often lack the specialized expertise needed to optimize their code and workflows for different computing environments. This inefficiency leads to wasted computational resources, increased costs, and longer processing times. Substantial performance improvements and cost savings could be achieved through targeted optimization of scientific codes for specific infrastructure environments, but this requires specialized knowledge that most domain scientists do not possess.

Recommendation 3c: SMD should create a computational efficiency project that provides expertise and tools for optimizing scientific code across HEC and cloud platforms.

Finding 3d (Data and analytics platforms): NASA researchers are struggling to access compute resources adjacent to remote data. Current practices are not aligned with future norms requiring collocated data and analysis platforms to enable advanced scientific research. Scientists need integrated environments where they can explore, analyze, visualize, and share data without transferring massive datasets. Open-source solutions that enable these workflows already exist in the broader scientific community but have not been systematically deployed across NASA's data ecosystem. Without leveraging these existing open-source platforms, NASA scientists face unnecessary barriers to working efficiently with cloud and HEC resources, particularly for data-intensive applications like AI/ML workflows.

Recommendation 3d: SMD should deploy integrated data and analytics platforms on both cloud and HEC resources that co-locate data with computational tools.

Area 4: Artificial Intelligence and Quantum Information Processing

Finding 4a (Co-develop AI models): NASA's scientific data assets represent an unparalleled resource for advancing artificial intelligence capabilities that can secure U.S. leadership in critical technologies. The development of AI models across SMD currently occurs through multiple independent efforts, creating potential redundancies and inefficiencies. NASA's unique position—having both data experts who understand dataset nuances and domain scientists who can verify models against scientific principles—creates an opportunity to develop gold-standard foundation models following principles of reproducibility and transparency. However, the lack of coordination between these independent efforts limits the development of large-scale, reusable foundation models that could be efficiently adapted across diverse scientific applications. The expansion to multimodal foundation models, physics-informed AI, reinforcement learning for scientific exploration, and data-driven digital twins presents transformative opportunities for scientific discovery while maintaining U.S. technological leadership and creating commercial applications that benefit the American economy.

Recommendation 4a: SMD should establish an enterprise AI program that develops open-source foundation models with documented scientific validation protocols and cross-divisional applications.

Finding 4b (Explainable AI models): Scientific advancement depends on understanding not just what conclusions are reached, but how they are derived. AI systems functioning as "black boxes" undermine this core principle, regardless of their accuracy. The scientific method demands reproducibility, peer review, and critical analysis—all impossible without visibility into how results are generated. When AI systems make predictions or classifications without revealing their reasoning pathway, they create a fundamental tension with scientific epistemology. This opacity prevents proper validation of findings, limits the ability to identify potential biases or errors, and creates barriers to scientific consensus-building. Additionally, comprehensive documentation should accompany each model, detailing methodologies, training data limitations, potential biases, and appropriate use cases to foster scientific reproducibility and maintain the integrity of research that is fundamental to NASA's scientific mission.

Recommendation 4b: SMD should publish AI transparency and reproducibility standards for NASA-funded research, including documentation requirements and verification protocols.

Finding 4c (AI workflow best practices): As scientists increasingly rely on AI and agentic workflows, the fundamental process of generating results is transforming through non-deterministic algorithms, complex computational dependencies, and autonomous decision-making pathways. This transformation introduces a critical distinction between reproducibility (the ability of independent researchers to test a hypothesis through multiple methods and consistently achieve results) and replicability (perform the same experiment or study using the same methods and conditions to achieve the same result)¹⁸. While Earth and space sciences have historically navigated this nuance, AI systems amplify these challenges through their inherent complexity and potential opacity. Currently, SMD lacks standardized documentation requirements for AI model architectures, training methodologies, and decision processes. There are no established science-specific guardrails or factuality-checking mechanisms to ensure AI outputs conform to physical laws, nor technical standards for computational environments that facilitate verification of AI-driven results. Without clear protocols for documenting AI decision-making processes, detecting biases in AI systems, and distinguishing between reproducibility and replicability in non-deterministic workflows, researchers face mounting difficulties in verifying, validating, and building upon each other's work, potentially compromising scientific integrity as AI becomes integral to research processes.

Recommendation 4c: SMD should establish best practices for AI-driven scientific workflows that address transparency, bias mitigation, and the distinction between reproducibility and replicability in non-deterministic AI systems.

Finding 4d (Vendor-agnostic AI infrastructure): Science increasingly requires sophisticated AI capabilities, yet commercial AI solutions lack the specific guardrails essential for scientific applications. While general AI systems incorporate safeguards against hate speech and misinformation, science requires fundamentally different validation mechanisms—ensuring citations are accurate, data is used correctly, and models adhere to physical laws such as conservation of mass and energy. Dependence on proprietary, vendor-specific AI tools creates risks of technological lock-in and escalating costs while failing to address science-specific requirements for trustworthiness, reproducibility, transparent error quantification, and falsifiability. Current commercial solutions lack citation verification capabilities, data usage validation, and factuality checks that ensure AI outputs conform to established scientific principles. Without open-source alternatives that incorporate these unique validation criteria, NASA scientists face increasing barriers to maintaining scientific integrity while leveraging advanced AI capabilities, potentially limiting scientific progress and creating unsustainable dependencies on commercial vendors.

Recommendation 4d: SMD should develop open-source AI frameworks with science-specific validation mechanisms to avoid vendor lock-in and ensure scientific integrity.

¹⁸ Executive Order No. 14303. (2025, May 28). "[Restoring Gold Standard Science](#)". Federal Register, 90 FR 22601.

Finding 4e (Quantum Computing): Quantum computing presents unique opportunities but requires specialized expertise. The fundamentally different computational model of quantum systems and the rapidly evolving quantum technology landscape makes it difficult for non-specialists to identify which scientific problems are best solved by quantum methods versus those better suited for classical approaches. Currently, SMD lacks a coordinated strategy for identifying priority scientific applications where quantum computing offers practical advantages over classical approaches. There is insufficient quantum expertise embedded within science teams across NASA centers, limiting the ability to evaluate quantum solutions for complex computational challenges. Without dedicated expertise centers and strategic partnerships with industry, NASA risks missing transformative quantum applications while potentially investing resources in quantum approaches that may not yield practical scientific benefits. The absence of systematic evaluation processes makes it challenging to distinguish grand-challenge problems that truly require quantum solutions from those better addressed through classical or hybrid approaches.

Recommendation 4e: SMD should implement a quantum computing strategy that identifies priority scientific applications and establishes expertise centers to support implementation.

Area 5: Documentation and Training

Finding 5a (Standardized educational materials): NASA SMD repositories have evolved from basic README files to comprehensive tutorials, notebooks, and hackathons for data exploration and analysis. However, this expansion has occurred without standardized best practices, resulting in inconsistent quality, outdated code, maintenance gaps, and divergent pedagogical approaches across repositories. Currently, there are no defined quality requirements for data access and computational tutorials, no clear assignment of maintenance responsibilities for materials that help users navigate NASA's data and computing resources, and no systematic review procedures to ensure these materials remain current and useful. The lack of templates, quality assurance protocols, and automated testing means that materials teaching data access, computational workflows, and platform usage vary widely in their pedagogical effectiveness and technical reliability. As materials become increasingly complex, sustainability challenges mount, while the rise of Large Language Models (LLMs) has reduced the need for elaborate analysis examples, shifting priorities toward reliable data access guidance and metadata interpretation. Without centralized standards and consistent review processes for data and computational educational resources, these materials risk becoming obsolete or misleading, undermining their value to the scientific community.

Recommendation 5a: SMD should establish standards for data and computational educational materials that define quality requirements, maintenance responsibilities, and review procedures.

Finding 5b (Strategic training program): Despite significant investments in advanced AI capabilities, foundation models, complex software, science platforms, and data and computing infrastructure, NASA lacks a comprehensive strategy to train scientists on effectively utilizing these resources. While various training approaches exist—including workshops, hackathons, and

ad hoc mentorship—these efforts are not coordinated under a documented curriculum with clear learning objectives. There is significant variation in delivery methods across divisions without established metrics to evaluate effectiveness or track participant skill development. Several divisions have demonstrated that Communities of Practice dramatically accelerate community adoption of new technologies and scientific progress, highlighting the need for both structured training and ongoing collaborative learning opportunities. Without a more effective and efficient approach to education on applying complex methods to scientific problems and leveraging new computational environments, these technological investments risk underutilization and diminished scientific return. The technical complexity of these tools creates a substantial knowledge barrier for domain scientists whose expertise lies in their scientific discipline rather than in AI implementation, cloud and HEC computing, or other modern advances.

Recommendation 5b: SMD should create a comprehensive training program for scientific data and compute users with documented curriculum, delivery mechanisms, and participant metrics.

Area 6: Governance and Strategy

Finding 6a (Comprehensive metrics framework): Current metrics collection for SMD repositories is limited to basic information, such as volumes and downloads, failing to capture broader scientific contributions, open science activities, and economic impacts of data and software. Enhanced metrics incorporating aspects like publications and open science efforts can readily be mapped to standards, improving FAIR alignment and metadata assessment. Being able to assess the value of data is essential for allocating resources effectively and demonstrating both scientific impact and economic benefits.

Recommendation 6a: SMD should implement a comprehensive metrics framework to evaluate and track data and software throughout its lifecycle, FAIR criteria, openness, scientific impact, and economic benefits of data and software resources.

Finding 6b (Supporting scientific software): Some divisions have a long history of supporting open-source scientific software tools and sharing scientific software with their published papers, thus increasing their transparency and reproducibility. While some widely used, mission critical, open-source projects are well supported, many are not. Divisional support for software development and maintenance positively impacts reproducibility of results when properly prioritized. There are meaningful ways to support open-source software projects, in addition to funding.

Recommendation 6b: SMD should develop a strategy for identifying and supporting scientific open-source software through its full lifecycle.

Finding 6c (Modern development tools): Inconsistent access and education in the use of modern tools has hindered advances and efficiency in software engineering, lifecycle management, cyber-compliance, and the large-scale use of complex scientific software.

Recommendation 6c: SMD should ensure that its workforce has increased access to modern integrated development environments (including AI-assisted development), software package management, version control, analytic platforms to improve efficiency

and cyber-security, and access to appropriate training and educational resources including Communities of Practice in the proper and effective use of these tools and complex scientific software.

Finding 6d (Streamlined software release process): The NASA Software Release Process ([NPR 2210.1E](#)) needs significant streamlining, clarity, and improvement, including simplified processes for the release of open-source software and legacy codes and improved uniformity across centers.

Recommendation 6d: SMD should continue to identify areas for improvement and advocate for improvements to NASA’s software release process to improve efficiencies.

Finding 6e (Mid-cycle workshops): The development of this strategy was greatly accelerated and enhanced by NASA's systematic approach to gathering NASA scientific community input through multiple channels: community workshop and commissioned National Academies reports (see [References](#)). These structured engagements allowed us to synthesize diverse perspectives and emerging requirements efficiently without requiring extensive new community input collection. This continuous engagement model provided an invaluable foundation of recent, relevant community feedback that made this strategy development responsive and well-informed. Without this ongoing dialogue, updating the strategy would have required a significantly longer timeframe and more resource-intensive process of community consultation. However, the current approach of relying primarily on periodic, 5-year comprehensive updates does not provide formal mechanisms for mid-cycle assessment, adjustment of implementation priorities, or integration of feedback about emerging technologies and research needs. There are no documented procedures for systematically gathering community input during the implementation period, nor are there established channels for incorporating insights about rapidly evolving fields like AI into ongoing strategic decisions. By not proactively identifying unknown unknowns between major strategy updates, NASA risks missing opportunities to align resources with emerging trends and novel scientific use cases.

Recommendation 6e: SMD should establish a mid-cycle strategy assessment process with documented procedures for gathering community input and updating implementation priorities.

APPENDIX D: ALIGNMENT WITH OCSDO AND NASA

The strategy was designed with alignment across NASA and Office of the Chief Science Data Officer (OCSDO) through four key mappings ([Table D. 1-Table D. 3](#)): OCSDO's goals, OCSDO requirements, and NASA Data Strategy goals. This practical approach accelerates implementation by leveraging established capabilities, minimizing duplicative efforts, and enabling rapid progress toward NASA's vision for data and for AI-ready data ecosystems that accelerate discovery.

Table D. 1. *SDMC30 strategies and recommendations map directly to the OCSDO Goals, demonstrating strong alignment between the strategy and the office responsible for its implementation. This mapping shows how all 11 strategies are distributed across OCSDO's focus areas of evolving data systems, advancing scientific discovery and productivity through innovative applications, and supporting collaborative, transparent and reproducible science.*

OCSDO Goals	SMD Strategy 2025–2030 Strategies
<p>Goal 1: Continuously evolve data and computing systems for efficiency, sustainability, security, and scientific integrity</p>	<p>Strategy 1.1: Standardized data and software management approaches</p> <p>Strategy 1.3: Standardized interfaces and authoritative data pathways</p> <p>Strategy 2.1: Computing resource selection frameworks</p> <p>Strategy 2.2: Integrated science platforms</p>
<p>Goal 2: Advance scientific discovery through innovative applications and partnerships in data science, advanced analytics, and artificial intelligence (AI)</p>	<p>Strategy 2.3: Seamless computation</p> <p>Strategy 3.1: Open-source AI models for scientific data</p> <p>Strategy 3.3: Cross-disciplinary communities of practice</p> <p>Strategy 3.4: Scientific reproducibility in AI research</p>
<p>Goal 3: Support transparent, trustworthy, and reproducible science through open science tools, practices, and guidance</p>	<p>Strategy 1.2: Consistent metadata standards and service frameworks</p> <p>Strategy 1.4: Transparent governance frameworks for scientific assets</p> <p>Strategy 3.2: Training programs</p>

Table D. 2. *SDMC30 strategies map directly to existing OCSDO Program requirements. Rather than requiring new organizational structures or programs, the strategy can be implemented through established operational vehicles. This approach accelerates implementation by leveraging current capabilities and expertise while minimizing administrative overhead.*

OCSDO Program Requirements	SMD Strategy 2025–2030 Strategies
<p>2.1.1 Develop and oversee activities and initiatives to ensure NASA's science data are accessible to all and produce practical benefits to society.</p>	<p>Strategy 1.2: Consistent metadata standards and service frameworks</p> <p>Strategy 1.3: Standardized interfaces and authoritative data pathways</p> <p>Strategy 2.2: Integrated science platforms</p>
<p>2.1.2 Enable adoption, monitor effectiveness, and periodically update the Scientific Information Policy for the Science Mission Directorate (SMD Policy Document SPD-41a).</p>	<p>Strategy 1.1: Standardized data and software management approaches</p> <p>Strategy 1.4: Governance frameworks for data and software lifecycles</p>
<p>2.1.3 Evaluate, develop, and evolve SMD data and computing capabilities to meet strategic needs of science divisions and research communities.</p>	<p>Strategy 2.1: Computing resource selection frameworks</p> <p>Strategy 2.3: Seamless computation</p>
<p>2.1.4 Develop strategies and activities to support the use of advanced data science techniques, including AI/ML, to meet SMD's science objectives.</p>	<p>Strategy 3.1: Open-source AI models for scientific data</p> <p>Strategy 3.4: Scientific reproducibility in AI research</p>
<p>2.1.5 Maintain and update the SMD Strategic Plan for Data Management and Computing every five years.</p>	<p>Strategy 1.4: Transparent governance frameworks for scientific assets</p>
<p>2.1.6 Develop and maintain strategic partnerships to increase the adoption of open science and the use of advanced data science tools and techniques to drive scientific discovery.</p>	<p>Strategy 3.2: Training programs</p> <p>Strategy 3.3: Cross-disciplinary communities of practice</p>

Table D. 3. *SDMC30’s alignment with NASA’s agency-wide 2021 Data Strategy goals demonstrate how SMD’s science-specific approach builds upon and extends NASA’s enterprise data governance framework while addressing the unique requirements of scientific data and computing. The mapping shows how SMD’s implementation efforts will contribute to agency-wide data maturity objectives while focusing on science-specific needs.*

NASA Data Strategy (2021) Goals	SMD Strategy 2025–2030 Strategies
Goal 1: Strengthen Data Governance	Strategy 1.2: Consistent metadata standards and service frameworks Strategy 1.4: Transparent governance frameworks for scientific assets
Goal 2: Advance a Data-Driven Culture	Strategy 3.2: Training programs Strategy 3.3: Cross-disciplinary communities of practice
Goal 3: Improve Data Products & Data Principles	Strategy 1.1: Standardized data and software management approaches Strategy 2.2: Integrated science platforms Strategy 3.4: Scientific reproducibility in AI research
Goal 4: Build an Enterprise Data Architecture	Strategy 1.3: Standardized interfaces and authoritative data pathways Strategy 2.1: Computing resource selection frameworks Strategy 2.3: Seamless computation Strategy 3.1: Open-source AI models for scientific data

APPENDIX E: DEFINITIONS AND ACRONYMS

Definitions

Dataset: A collection of organized data that has been uploaded to a data repository and assigned (or could be assigned) a Digital Object Identifier (DOI).

Repository: An organized storage location that makes data findable and accessible. For NASA, SPD-41a states that repositories should follow the OSTP Desirable Characteristics of Data Repositories for Federally Funded Research. This location should be the authoritative source for NASA data and the long-term archive.

Data Node: A specialized data collection that is part of a larger data repository.

Artificial Intelligence (AI): This strategy adopts the definition of artificial intelligence established in [15 U.S.C. § 9401](#), as referenced in Executive Order 14179 (2025). For the purposes of this strategy, “artificial intelligence” (AI) refers to a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems use machine- and human-based inputs to: (A) perceive real and virtual environments; (B) abstract such perceptions into models through automated analysis; and (C) use model inference to formulate options for information or action.

Reproducibility: The ability of independent researchers to test a hypothesis through multiple methods and consistently achieve results that confirm or refute it, ensuring findings are generalizable and robust across different approaches.

Replicability: The ability to perform the same experiment or study using the same methods and conditions to achieve the same result.

Scientific software: Software that provides users some degree of scientific utility or produces a scientific result or service.

Acronyms

Acronym	Full Term
AI	Artificial Intelligence
APD	Astrophysics Division
API	Application Programming Interface
ARC	Ames Research Center
ATM	PDS Atmospheres
AWS	Amazon Web Services
BPS	Biological and Physical Sciences
CDCS	Core Data and Computing Services
CDO	Chief Data Officer
CISTO	Computational and Information Science and Technology Office
CPU	Central Processing Unit

CSDO	Chief Science Data Officer
DAAC	Distributed Active Archive Center
DASP	Data Analysis Services Program
DOI	Digital Object Identifier
DSI	Data Science and Innovation
EAR	Export Administration Regulations
EOSDIS	Earth Observation System Data and Information System
ESD	Earth Science Division
ESDIS	Earth Science Data and Information System
Evidence Act	Foundations for Evidence-Based Policymaking Act of 2018
FAIR	Findable, Accessible, Interoperable, and Reusable
FAIRUST	Findable, Accessible, Interoperable, Reusable, Understandability, Security, and Trustworthiness
FTPS	File Transfer Protocol Secure
GEO	PDS Geosciences
GES DISC	Goddard Earth Sciences Data and Information Services Center
GIBS	Global Imagery Browse Services
GIS	Geographic Information System
GMAO	Global Modeling and Assimilation Office
GPU	Graphics Processing Unit
GSFC	Goddard Space Flight Center
HDRL	Heliophysics Data Repository and Library
HEC	High-End Computing
HEASARC	High Energy Astrophysics Science Archive Research Center
HLSP	High-Level Science Products
HPC	High-Performance Computing
HPD	Heliophysics Division
HTTP	HyperText Transfer Protocol
HTTPS	HyperText Transfer Protocol Secure
IBM	International Business Machines
IDE	Integrated Development Environment
IMG	PDS Cartography and Imaging Sciences
IMPACT	Interagency Implementation and Advanced Concepts Team
IPAC	Infrared Processing & Analysis Center
IRB	Independent Review Board
IRSA	Infrared Processing & Analysis Center (IPAC) Infrared Science Archive
ITAR	International Traffic in Arms Regulations

JPL	Jet Propulsion Laboratory
LLM	Large Language Models
M-25-21	OMB 2025 Memo 21 "Accelerating Federal Use of AI through Innovation, Governance, and Public Trust"
M-25-22	OMB 202 Memo 22 "Driving Efficient Acquisition of Artificial Intelligence in Government"
MAAP	Multi-Mission Algorithm and Analysis Platform
MAST	Mikulski Archive for Space Telescopes
MFA	Multifactor Authentication
ML	Machine Learning
MSFC	Marshall Space Flight Center
NAIF	PDS Navigation & Ancillary Information Facility
NARA	National Archives and Records Administration
NAS	NASA Advanced Supercomputing
NASEM	National Academies of Sciences, Engineering and Mathematics
NCCS	NASA Center for Climate Simulation
NSF	National Science Foundation
NSSDCA	National Space Science Data Coordinated Archive
OCSDO	Office of the Chief Science Data Officer
OIG	Office of Inspector General
OMB	Office of Management and Budget
OPEN Act	Open, Public, Electronic, and Necessary Government Data Act
OSDR	Open Science Data Repository
OSDMP	Open Science Data Management Plan
OSI	Open Science Implementation
OSS	Open-Source Software
OSTP	Office of Science and Technology Policy
PB	Petabytes
PDS	Planetary Data System
PID	Persistent Identifier
PI	Principal Investigator
PID	Persistent Identifier
PII	Personal Identifying Information
PIV	Personal Identity Verification
PLRA	Preliminary Life-Cycle Review Assessment
PPBE	Planning, Programming, Budgeting, and Execution
PPI	PDS Planetary Plasma Interactions
PSD	Planetary Science Division

PSI	Physical Sciences Informatics
RAM	Random Access Memory
RMS	PDS Ring Moon Systems
ROSES	Research Opportunities in Space and Earth Sciences
S3	Simple Storage Service
SALMON	Stand Alone Missions of Opportunity Notice
SBN	PDS Small Bodies
SBU	Standard Billing Unit
SciX	Science Explorer
SDAC	Solar Data Analysis Center
SDE	Science Discovery Engine
SDMC24	Strategy for Data Management and Computing for Groundbreaking Science 2019–2024
SDMC30	Strategy for Data Management and Computing to Accelerate Scientific Discovery 2025–2030
SIPS	Science Investigator-led Processing Systems
SMCE	Science Managed Cloud Environment
SMD	Science Mission Directorate
SOSS	Supplements for Open-Source Science
SPD-41a	Science Mission Directorate Policy Document-41a
SPDF	Space Physics Data Facility
SSO	Single Sign-On
STSci	Space Telescope Science Institute
TB	Terabytes
TOPS	Transform to Open Science
UMB	University of Maryland
URL	Uniform Resource Locator
U.S.	United States of America
VEDA	Visualization, Exploration, and Data Analysis

The use of trademarks or names of manufacturers in this report is for accurate reporting and does not constitute an official endorsement, either expressed or implied, of such products or manufacturers by the National Aeronautics and Space Administration.

Available from:

NASA STI Program / Mail Stop 050
NASA Langley Research Center
Hampton, VA 23681-2199