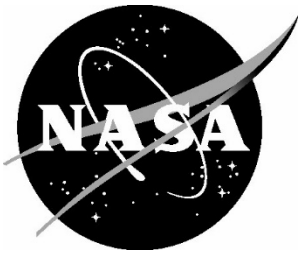
NASA/SP-20260001965



NASA Science Mission Directorate's FY2024 Science Data Repository Metrics Report

Kevin J. Murphy
NASA Science Mission Directorate, Office of the Chief Science Data Officer, NASA Headquarters, Washington, D.C.

Suggested Citation: NASA (2026) NASA Science Mission Directorate's FY2024 Science Data Repository Metrics Report, https://doi.org/10.64631/BJDW5471.

March 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.

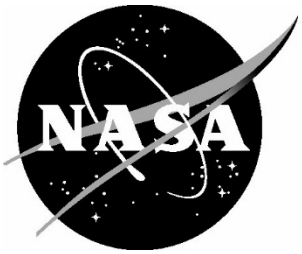
Specialized services also include organizing and publishing research results, distributing specialized research announcements and feeds, providing information desk and personal search support, and enabling data exchange services.

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-20260001965



NASA Science Mission Directorate's FY2024 Science Data Repository Metrics Report

Kevin J. Murphy

NASA Science Mission Directorate, Office of the Chief Science Data Officer, NASA Headquarters, Washington, D.C.

Suggested Citation: NASA (2026) NASA Science Mission Directorate's FY2024 Science Data Repository Metrics Report, https://doi.org/10.64631/BJDW5471.

National Aeronautics and
Space Administration

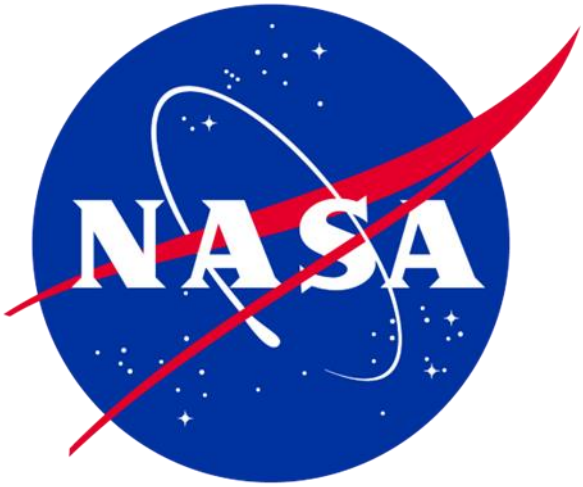
Headquarters
Washington, D.C.

March 2026

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.

Available from:

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



Science Mission Directorate
Office of the Chief Science Data Officer
FY2024 Science Data Repository Metrics Report

Prepared by the Office of the Chief Science Data Officer

Approved by:

Kevin Murphy Digitally signed by Kevin Murphy
Date: 2026.03.17 16:58:33 -04'00'

_____ Date: _____

Kevin Murphy
Chief Science Data Officer
Science Mission Directorate

INTRODUCTION

NASA's Science Mission Directorate (SMD) manages over 150 Petabytes (PB) of scientific data across 54,532 datasets distributed among 10 repositories, serving more than 53 million unique users annually. This report presents Fiscal Year (FY) 2024[1] metrics for the scientific data repositories, establishing quantitative baselines for measuring infrastructure health, growth trajectories, and utilization patterns. Annual publication of these metrics serves multiple critical functions: transparency through documenting stewardship of federally funded scientific data; strategic planning by establishing baselines for measuring progress toward strategic goals; resource allocation through providing data-driven insights for investment decisions; and accountability by demonstrating responsible management of public resources.

The five divisions within NASA's Science Mission Directorate, Astrophysics (APD), Biological and Physical Sciences (BPS), Earth Science (ESD), Heliophysics (HPD), and Planetary Science (PSD), collected comprehensive metrics on their data management practices. These metrics included data volumes, growth rates and projections through 2030, user engagement and access patterns, infrastructure distribution across on-premises and cloud environments, standards compliance, and management approaches and policies.

The Office of the Chief Science Data Officer (OCSDO) enhances the value of NASA's science data through the evolution of data and computing systems, accelerating discoveries through innovative data science techniques, and fulfilling NASA's strategic goal of ensuring scientific data are readily available to all. Within SMD, the OCSDO ensures that the vast amounts of scientific data generated by NASA's missions are accessible and preserved for future analysis and ready to support future exploration.

Note on methodology: All metrics reflect FY2024 unless otherwise noted. Future annual editions will attempt to maintain consistent measurement approaches enabling trend analysis.

[1] FY2024: October 1, 2023 – September 30, 2024

BACKGROUND

The infrastructure that supports NASA's scientific holdings reflects both the historical evolution of discipline-specific data management practices and the fundamental differences in scientific requirements across domains. Each science division has adopted organizational structures that reflect their scientific communities' needs, mission portfolios, and historical evolution.

The metrics presented in this report represent the holdings of NASA SMD's scientific and mission data repositories rather than the complete universe of NASA scientific data. These 10 repositories (Table 1) contain the publicly accessible, long-term preserved NASA data for the different NASA divisions:

- APD: Three independent repositories mainly focused on distinct wavelength regimes
- BPS: Two focused repositories with centralized quality control
- ESD: Single repository with a unified search layer, supplied by data provided by 11 distributed active archive centers and 11 science investigator-led processing systems.
- HPD: Two independent repositories each serving distinct physical domains.
- PSD: Two independent repositories, including one repository serving planetary mission data through 6 science data and 2 support nodes

In addition, NASA also supports a larger ecosystem of domain specific data repositories that have risen through organic processes to support scientific discovery. At this time, thematic repositories that are not part of the primary data repositories are beyond the scope of this report but may be included in future evaluations.

SMD DATA METRICS

Table 2 presents the overall scale of NASA SMD's data holdings as of FY2024, revealing significant differences in volume, user engagement, and growth trajectories across the five science divisions. These differences reflect the mission portfolios, data characteristics, and scientific community sizes of the individual SMD divisions.

SMD data includes 150 PB across 54,532 datasets serving 53 million users annually, with a 31.4 PB/year current ingest rate that will increase holdings to 530 PB by 2030.

DATA REPOSITORY EVOLUTION

In addition to the metrics for each repository, we also report on several areas that represent evolution of the data repositories. This includes on the transition to using cloud capabilities, standards implementation, user accesses, and the integration with High End Computing. In some cases, the specific experience within a division is presented.

On-Premises to Cloud Transition

Tables 3 and 4 document NASA's major infrastructure transition: from storing and delivering data on-premises to storing and distributing data from a cloud environment.

Current distribution: 58% on-premises (87 PB) / 42% cloud (90 PB)
Projected 2030: 16% on-premises (83 PB) / 84% cloud (458 PB)

This change represents a transformation from primarily operating data centers to primarily managing cloud services, requiring different skill sets for both NASA repository staff and researchers who access the data. The data storage and distribution architecture for each division are:

- BPS: Completely cloud-native (100% in AWS S3): Demonstrates viability of cloud-only for modest volumes (240 TB), eliminates on-premises infrastructure costs entirely, and serves 50,000 users from fully cloud-managed systems.
- ESD: Aggressive migration (82 PB in cloud, targeting 395 PB by 2030): Cloud increasingly serves as primary storage, not just as a mirror, generates 147,650 PB/year cloud access, >99.9% of all SMD cloud access, and a successful user community transition to cloud-native workflows (21M cloud users).
- APD: Conservative selective migration (4.1 PB cloud, 74% remains on-premises): Prioritizes high-demand datasets for cloud, maintains primary copies on-premises, developing Fornax platform with resource management tools to control cloud costs.
- HPD: Balanced hybrid (40% duplicated across both environments): Uniquely, 57% of data copied to NCCS for HEC access, 100% API coverage in cloud vs. 44% on-premises, and a multi-platform strategy serving different computational use cases.
- PSD: Gradual progressive (54% in both environments): Balancing traditional download workflows with emerging cloud capabilities, 60% API coverage.

On-premises usage (14,292 PB/year data access, 64 PB/year downloads) is less than cloud usage (147,654 PB/year data access, 649 PB/year egress), primarily due to ESD usage. The rapid adoption of cloud-based data was supported by ESD's provision of cloud-optimized formats, enabling efficient partial access and their extensive community support, hackathons, training programs, and provision of cloud analysis platforms.

ESD User Community Adaptation

ESD's successful transition of user community to cloud-native workflows demonstrates large-scale migration is achievable:

- 21 million cloud users
- 147,654 PB/year cloud access (>99.9% of all SMD cloud access)
- 228× access-to-storage ratio

Much of the success in ESD is due to a strategic approach towards the cloud transition, including:

- Sustained training and documentation: ESD provides hosted Jupyter environments (Openscapes, Cyrocloud, Open Synthetic Aperture Radar (SAR) Lab), comprehensive platforms (Multi-Mission Algorithm and Analysis Platform - MAAP), and extensive workshops
- Platform development: Managed user-friendly analysis environments reduce barriers by allowing productivity without cloud architecture expertise
- Maintaining optionality: Traditional download workflows remain available alongside cloud-native options during transition
- Community engagement: Regular workshops and hackathons identify barriers and develop solutions

STANDARDS IMPLEMENTATION

Tables 3-4 reveal substantial disparities in Digital Object Identifiers (DOI) and Application Programming Interface (API) implementation. Datasets without DOIs are harder to cite in publications, track for impact, and integrate into broader data ecosystems.

Datasets without APIs create barriers for automated workflows, Artificial Intelligence (AI) applications, and integration with external tools. As AI becomes more important for scientific discovery, lack of API access prevents machine interaction with data.

Table 5 documents that three divisions (ESD, HPD, PSD) have implemented division-level standardized data management plans providing consistent guidance across missions. APD and BPS maintain repository-specific or mission-specific approaches, trading consistency for flexibility.

USER AUTHENTICATION AND ACCESS FRICTION

ESD requires user login for all data access, enabling usage tracking and user support but creating friction (Table 5). User feedback characterizes current implementation as "cumbersome and intrusive." As cloud computing offerings expand, authentication becomes increasingly important for cybersecurity and resource allocation but must be modernized to balance security needs with user experience.

APD requires login only for embargoed data during proprietary periods. BPS, HPD, and PSD provide fully open data access without authentication, simplifying user experience but limiting ability to track users or manage computational resources.

Cloud-based computational environments will likely require more divisions to implement authentication, not to restrict data access, but to authorize compute usage and ensure security.

COMPUTATIONAL RESOURCES

Scientists have access to multiple disconnected systems (on-premises repositories, cloud storage, High-End Computing (HEC) at NASA Advanced Supercomputing (NAS), NASA Center for Climate Simulation (NCCS)) creating barriers to collaboration and seamless workflows. HPD uniquely integrates HEC by copying 57% of data (1,800 TB) to NCCS, demonstrating value of bringing data to computation when moving computation to data is impractical.

Most divisions provide limited computational resources adjacent to data:

- APD: Developing platforms (Time series Integrated Knowledge Engine (TIKE), Roman Research Nexus (RRN), Fornax) to facilitate cloud usage, but still maintaining download-centric workflows
- BPS: Provides analysis tools (Multi-study Data Visualization, Environmental Data Application, RadLab) on Science Managed Cloud Environment (SMCE)
- ESD: Leads in platform development (MAAP) and in-cloud services

- HPD: Provides analysis tools, and requires login for cloud compute access, also developing a uniform hybrid computing platform on top of HEC and cloud resources
- PSD: Traditional download dominates, cloud platforms under development

Co-locating data with computation and analytics enables workflows impossible with traditional download approaches, particularly for AI requiring direct data access at scale.

CONCLUSION

The FY2024 science data repository metrics report provides the status of the primary repositories for scientific and mission data for SMD. Annual publication of these metrics will enable tracking progress toward strategic goals, provide early warning of challenges, and ensure accountability for responsible stewardship of NASA's invaluable scientific data assets.

The challenge for NASA's data management evolution lies in preserving the benefits of specialized expertise and domain-specific optimization while reducing unnecessary fragmentation, increasing cross-disciplinary interoperability, and positioning the data enterprise to support the next generation of scientific discovery through 2030 and beyond.

Table 1. SMD Repository Ecosystem

SMD Repository Ecosystem				
Repository Name	**Data Discovery**	**Short Name**	**Division**	**Host Institution**
High Energy Astrophysics Science Archive Research Center	HEASARC	HEASARC	Astrophysics	GSFC
Mikulski Archive for Space Telescopes	MAST	MAST	Astrophysics	STScI
Infrared Processing & Analysis Center (IPAC) Infrared Science Archive	IRSA	IRSA	Astrophysics	IPAC
Open Science Data Repository	OSDR	OSDR	Biological & Physical Sciences	ARC
Physical Sciences Informatics	PSI	PSI	Biological & Physical Sciences	MSFC
Earth Observation System Data and Information System	Earthdata	EOSDIS	Earth Science	GSFC
Solar Data Analysis Center	Heliodata	SDAC	Heliophysics	GSFC
Space Physics Data Facility	Heliodata	SPDF	Heliophysics	GSFC
Astromaterials Data System	Astromat search	Astromat	Planetary Science	Columbia University
Planetary Data System	PDS search	PDS	Planetary Science	GSFC

Table 2. *Data assets of the NASA Science Mission Directorate (SMD) repositories for 2024 - 2030.*

Data Assets of the NASA SMD Repositories for 2024-2030						
	APD	**BPS**	**ESD**	**HPD**	**PSD**	**Totals**
Unique User IPs	8,800,000	50,000	28,400,000	8,000,000	7,990,000	**53,240,000**
Data Repositories	3	2	1	2	2	**10**
Datasets	46	1,068	13,500	11,500	27,460	**54,532**
Dataset Volume	15,700 TB	240 TB	127,665 TB	3,151 TB	3,860 TB	**150,616 TB**
Growth	2,100 TB/yr	92 TB/yr	28,000 TB/yr	470 TB/yr	785 TB/yr	**31,447 TB/yr**
2030 Projected Volume	70,000 TB	750 TB	400,000 TB	35,000 TB	24,000 TB	**529,750 TB**

Accompanying Information for Table 2:

- Unique User IPs (Number): Unique IP addresses accessing a repository or NASA website to download or perform computation each year. For APD, this is a ballparked number, as not all data distribution points reported this information and because of contamination by bots, crawlers, etc. For ESD, there were 8.4M unique Earthdata.gov (7.4M on-prem/~1M cloud) and 20M unique GIBS/Worldview.
- Data Repositories (Number): Number of repositories for each Division. For APD, there are only three data repositories with mission data. Within ESD there are 11 distributed active archive centers and 11 science investigator-led processing systems. Within PDS, there are 6 science nodes and the Navigation and Ancillary Information Facility (NAIF) that host data.
- Datasets (Number): Number of datasets. For APD, survey and observatory missions are considered one dataset.
- Dataset Volume (TB): Volume of data holdings, excluding any duplicative data, in Terabytes (TB).
- Growth (TB/year): Total volume of new data expected, in TB per year, as of 2024. This is expected to significantly increase with future launches of high-volume missions.
- 2030 Projected Volume (TB): Expected 2030 volume of data holdings, in TB, excluding any duplicative data.

Table 3. *Current state assessment of the NASA SMD repositories - On-premises metrics.*

On-premises Metrics							
		APD	**BPS**	**ESD**	**HPD**	**PSD**	**Totals**
Storage	NASA Mission Dataset Volume	7,700 TB	0 TB	50,000 TB	3,050 TB	860 TB	61,610 TB
	NASA Investigator Dataset Volume	900 TB	0 TB	6,000 TB	35 TB	930 TB	7,865 TB
	Non-NASA Dataset Volume	6,600 TB	0 TB	11,121 TB	0 TB	0 TB	17,721 TB
	NCCS Dataset Volume (copied)	0 TB	0 TB	0 TB	1,800 TB	0 TB	1,800 TB
	Total Dataset Volume	15,200 TB	0 TB	67,121 TB	3,085 TB	1,790 TB	87,196 TB
	Growth	2,100 TB/yr	0 TB/yr	15,000 TB/yr	370 TB/yr	515 TB/yr	17,985 TB/yr
	2030 Projected Volume	37,000 TB	0 PB	5,000 TB	35,000 TB	6,000 TB	83,000 TB
Access	Unique User IPs	8,800,000	0	7,400,000	8,000,000	7,990,000	32,190,000
	Data Access Volume	5,000 TB/yr	0 TB/yr	14,287,000 TB/yr	Not Available	2 TB/yr	14,292,002 TB/yr
	Data Download Volume	5,000 TB/yr	0 TB/yr	56,000 TB/yr	750 TB/yr	2,210 TB/yr	63,960 TB/yr
Dataset	Total Datasets	958	0	7,800	8,000	13,620	30,378
	Datasets with DOIs	890	0	7,800	2,000	1,565	12,255
	Datasets with API Access	958	0	7,800	3,500	640	12,898

Accompanying Information for Table 3:

- NASA Mission Dataset Volume: Volume of datasets produced by NASA missions or guest observer programs, excluding any duplicative data, in TB.
- NASA Investigator Dataset Volume: Volume of datasets produced by NASA-funded Principal Investigators (PIs), excluding any duplicative data, in TB. For ESD, this includes model, aircraft, and field measurements. For APD, this includes contributed high-level science products (HLSPs).
- Non-NASA Dataset Volume: Volume of datasets that are not from a mission or NASA-funded PI. This data includes ancillary data necessary for processing mission and PI data, partner data where we have an agreement to hold the data in a repository, and other data where NASA has traditionally acted as the community repository.
- Total Dataset Volume: NASA Mission, NASA PI, and Non-NASA data, excluding any duplicative data, in TB.
- NCCS Dataset Volume (copied): Data copied to NASA Center for Climate Simulation for use on HEC, in TB.
- Growth: Total volume of new data expected, in TB per year, as of 2024. This is expected to significantly increase with future launches of high-volume missions, in TB/yr.
- 2030 Projected Volume: Expected 2030 volume of data holdings, in TB, excluding any duplicative data.
- Unique User IPs: Unique IP addresses accessing a repository to download or perform computation each year. For APD, this is a ballparked number, as not all data distribution points reported this information and because of contamination by bots, crawlers, etc.
- Data Access Volume: Total volume of data accessed in TB/yr. ESD values include ESDIS archive and visualization tools. For HPD, the Data Access Volume is not a tracked metric, but one that can be added in the future.
- Data Download Volume: Total volume of data downloaded, in TB/yr.
- Total Datasets: Number of datasets. For APD, there is one dataset per survey/observatory mission and HLSP.
- Datasets with DOIs: Number of datasets that have an assigned DOI.
- Datasets with API Access: Number of datasets that are accessible through an API.

Table 4. *Current state assessment of the NASA SMD repositories - Cloud metrics.*

Cloud Metrics							
		APD	**BPS**	**ESD**	**HPD**	**PSD**	**Totals**
Storage	NASA Mission Dataset Volume	3,400 TB	6 TB	56,500 TB	1,200 TB	1,970 TB	6,576 TB
	NASA Investigator Dataset Volume	700 TB	166 TB	15,000 TB	55 TB	100 TB	16,021 TB
	Non-NASA Dataset Volume	0 TB	68 TB	10,544 TB	66 TB	0 TB	10,678 TB
	Total Dataset Volume	4,100 TB	240 TB	82,044 TB	1,321 TB	2,070 TB	89,775 TB
	Growth	500 TB/yr	92 TB/yr	39,000 TB/yr	350 TB/yr	270 TB/yr	40,212 TB/yr
	2030 Projected Volume	42,000 TB	750 TB	395,000 TB	2,000 TB	18,000 TB	458,000 TB
Access	Unique User IPs	10,500	50,000	21,000,000	Not Available	30,000	21,090,500
	Data Access Volume	4,200 TB/yr	Not Available	147,650,000 TB/yr	Not Available	1 TB/yr	147,654,201 TB/yr
	Data Egress Volume	Not Available	Not Available	648,000 TB/yr	Not Available	1,120 TB/yr	649,120 TB/yr
Datasets	Total Datasets	46	1,068	5,700	3,500	13,840	24,154
	Datasets with DOIs	12	704	5,700	2,000	3,970	12,386
	Datasets with API Access	43	600	5,700	3,500	8,350	18,193

Accompanying Information for Table 4:

- NASA Mission Dataset Volume: Volume of datasets produced by NASA missions or guest observer programs, excluding any duplicative data, in TB.
- NASA Investigator Dataset Volume: Volume of datasets produced by NASA-funded principal investigators (PIs), excluding any duplicative data, in TB. For ESD, this includes model, aircraft, and field measurements. For APD, this includes contributed high-level science products (HLSPs).
- Non-NASA Dataset Volume: Volume of datasets, in TB, that are not from a mission or NASA-funded PI. This data includes ancillary data necessary for processing mission and PI data, partner data where we have an agreement to hold the data in a repository, and other data where NASA has traditionally acted as the community repository.
- Total Dataset Volume: NASA Mission, NASA PI, and Non-NASA data, excluding any duplicative data, in TB.
- Growth: Total volume of new data expected, in TB/year, as of 2024. This is expected to significantly increase with future launches of high-volume missions.
- 2030 Projected Volume: Expected 2030 volume of data holdings, in TB, excluding any duplicative data.
- Unique User IPs: Unique IP addresses accessing a repository to download or perform computation each year. For APD, this is a lower limit, as not all repositories reported this information. HPD does not track this metric currently.
- Data Access Volume: Total volume of data accessed for collocated analysis, in TB/yr. For APD this is a lower limit, as not all data repositories track this metric. ESD values include ESDIS archive and visualization tools. For HPD, as with on premises compute and storage, Data Access Volume is not currently tracked.
- Data Egress Volume: Total volume of data egressed from cloud storage, including egress of both NASA data and subsequent analysis results in TB/yr. For HPD, data egress data is not tracked in a manner consistent with other divisions. APD does not track this information.
- Total Datasets: Number of datasets. HPD does not currently track this metric. For APD, there is one dataset per survey/observatory mission and HLSP.
- Datasets with DOIs: Number of datasets that have an assigned DOI.
- Datasets with API Access: Number of datasets that are accessible through an A

Table 5. *Management and policies for the NASA SMD repositories and data they hold. This cross-divisional comparison shows both common patterns and significant variations in how each science division approaches key aspects of data management.*

5.a	Data Management Approach: How data flows through systems from creation to distribution
APD	• Three primary repositories: HEASARC, IRSA, and MAST • Repositories operate independently and report to NASA HQ Archive leadership • Mission teams and PIs submit data on agreed cadence using established standards • Repositories manage full lifecycle: ingest, preservation, curation, discovery, access, and distribution • Provide value-added services, documentation, tutorials, and user support • Serve as authoritative entry points with cross-archive discovery capabilities
BPS	• Data submitted via OSDR and PSI submission portals • Repository staff review, normalize, and validate metadata/data standards • Data associated with appropriate repository for management • Repositories manage long-term storage and preservation • OSDR and PSI serve as primary user entry points for discovery and access
ESD	• Managed through the Earth Science Data and Information System (ESDIS) enterprise • Data providers (SIPS, missions, PIs, partners) submit through DAACs to EOSDIS • Transitioning to centralized, cloud-based ingest and archive architecture • EOSDIS manages ingest, preservation, access, distribution, discovery, and transformation • DAACs provide user services, curation, and training • Single authoritative point of entry for Earth science data
HPD	• Federated partnership of two independent repositories: Heliophysics Data and Resource Library (HDRL) • Two distinct repositories operating independently of flight missions • Data submitted based on mission-specific cadence and standards • Each repository responsible for preservation, curation, access, and user services • HDRL provides standardized high-level metadata for unified search • Specialized applications maintained for domain-specific needs • Central node redistributes user-requested subsets (not long-term archive)
PSD	• Two complementary systems: Astromat and Planetary Data System (PDS) • Astromat and individual PDS nodes manage curation, documentation, user services, and preservation coordination • Astromat serves as long-term archive and single point of entry for designated datasets • PDS operates as a federation of discipline-specific nodes • Data submitted to appropriate PDS node in consultation with PDS • Within PDS, multi-node search interface plus specialized node-level tools • Long-term archival preservation coordinated with national archive partners

5.b	Data Storage Approach: Percentage of data on-premises, cloud, and HEC (by the data repository team). This does not include data copied to NCCS for HEC access by individual HEC users.	Division-Level Standardized Data Management Plan:	User Login required?
APD	74% data are only stored on-premises. 3% data are only stored in AWS S3. 23% of data are both on-premises and cloud. 0% of data has been copied to NCCS for HEC access.	No	Yes, for embargoed data
BPS	0% data are only stored on-premises. 100% data are only stored in AWS S3. 0% of data are both on-premises and cloud 0% of data has been copied to NCCS for HEC access.	No	No
ESD	45% data are only stored on-premises. 41% data are only stored in AWS S3. 14% of data are both on-premises and cloud 0% of data has been copied to NCCS for HEC access.	Yes	Yes
HPD	58% data are only stored on-premises. 2% data are only stored in AWS s3. 40% data are both on-premises and cloud. 57% of data have been copied to NCCS for HEC access.	Yes	No
PSD	46% data are only stored on-premises. 0% data are only stored in AWS S3. 54% of data are both on-premises and cloud 0% of data has been copied to NCCS for HEC access.	Yes	No

5.c	Mission data transfer schedule to repository: When is data moved to a repository?	Data Access Approach: How are the Division's data accessed and/or retrieved?
APD	Transferred on a variety of fixed schedules, including continuously or at the end of the mission.	https, s3, API, web services, value-added services
BPS	Transferred continuously.	https, s3, API
ESD	Transferred continuously.	https, s3, API, value-added services, web services
HPD	Transferred on a variety of fixed schedules, including continuously or at the end of the mission.	https, ftps, s3, API, web services, value-added services
PSD	Transferred on a variety of fixed schedules, including continuously or at the end of the mission.	https, API

5.d	Data Computing Approach	Cloud Storage Approach
APD	• Users download data from on-premises or cloud repositories for analysis on systems ranging from laptops to high-end computing (HEC). • Increasing emphasis on cloud compute co-located with data to reduce movement and improve efficiency. • Science platforms (Fornax, NEXUS and TIKE) enable scalable, data-proximate analysis.	• Data stored in AWS us-east-1 via NASA or AURA Open Data Registry with free egress. • Primary authoritative copies maintained on-premises. • Fornax platform implementing tiered storage aligned to usage. • Resource management tools under development to monitor compute and egress costs in paid cloud environments.
BPS	• Users download data for local analysis. • Access datasets directly from public S3 buckets for use in their own cloud infrastructure. • Utilize division-provided cloud analysis platforms when appropriate.	• All data stored in AWS S3. • Architecture supports both direct download and cloud-native workflows.
ESD	• Most users download data for local analysis. • Expanding “in-place” access to retrieve only required data elements, reducing unnecessary data movement. • Integrated GIS and transformation services provide preprocessing within the data system.	• Most data stored, or being migrated to, AWS us-west-2 as the primary authoritative copy. • Level 0 backups maintained to enable full data reconstitution. • Enterprise backup guidance continuing to mature.
HPD	• Users primarily perform visualization, filtering, and lightweight analytics. • Supported by on-premises computational services.	• Data stored in AWS Open Data Registry with free egress to promote broad access. • Additional cloud management services will be considered when HDRL covers the cost.
PSD	• Users download data from on-premises resources to perform analysis on local compute resources. • Transitioning toward cloud-enabled workflows. • Future users may analyze data within SMD and PDS cloud compute platforms or egress as needed.	• Data stored in AWS us-west-2. • Cloud management practices and data backup policies are currently under development.

5.e	Cloud Computing Approach	HEC Approach
APD	• Developing publicly accessible Jupyter environments and full-fledged science platforms (e.g. Fornax, NEXUX and TIKE) for data visualization and analysis in the AWS cloud. • Enables data visualization and analysis in proximity to data. • Users may write and run code using limited cloud CPU and RAM resources without egress.	• Repositories do not provide HEC resources to the users. • NASA HEC resources are available to users with selected ROSES proposals through the NASA Advanced Supercomputing facility.
BPS	• Data visualization and analysis tools provided through web applications hosted on SMCE (CPU). • Tools include Multi-Study Data Visualization, Environmental Data Application, and RadLab.	• GeneLab workflows made available on NASA HEC for BPS scientists to access and run
ESD	• In-cloud services such as GIS and transformation services are provided. • User workloads are the responsibility of the user. • ESD provides some hosted environments, workshops and other NASA affiliations including managed Jupyter environments (Openscapes, Open SAR Lab, etc.) and more full-fledged analysis platforms (MAAP).	• HEC operates outside the data system. • On-premises NASA HEC run in close collaboration with DAACs (e.g., GMAO <-> GES DISC) on a case-by-case basis, where observation data feed models and output is archived.
HPD	• Provides a cloud-based stack of software and computing resources alongside HDRL data. • Includes both CPU and GPU machines. • User login required for access	• On-premises NASA HEC provided through Heliophysics resources at NCCS • Cloud resources provided through HelioCloud.
PSD	• Cloud Software and computing resources leverage SMD-provided services. • Additional PSD-provided services are under development.	• Repositories do not provide HEC resources to users. • On-premises NASA HEC resources are provided through the NASA Advanced Supercomputing facility.

5.f	Data and Computing Vision for 2030 (Cloud and HEC)	Publication Discovery Status
APD	• Progressively migrating data to the cloud to address rapidly increasing data volumes, particularly from survey missions. • Visualization, analysis tools, and highly scalable programmatic services through NASA Science Cloud platforms. • Integrating HEC in proximity to cloud-hosted data to enable advanced analytics, including AI/ML for cross-mission data integration.	• Publications are discovered through ADS (soon to be in Sci-X) and through generic internet or DOI searches. • Existing interlinking between APD data and the publications and software they support. • Additional efforts are needed to normalize the connection between data DOI and publication discovery.
BPS	• Expanding data visualization and analysis capabilities to support a broader range of data types across BPS, including advancing tools for analyzing and visualizing diverse biological and physical sciences data hosted on OSDR and PSI (e.g. omics, physiological, behavioral, environmental, and other mission-derived datasets). • Integrating compute resources alongside the data to enable scalable analysis using tools such as AI/ML for cross-experiment data integration, enhancing access and discoverability, accelerating translation of data into applied scientific insight.	• Publications discovered through the NASA Taskbook, DOI-searches, internet search engines, and analysis working group networks.
ESD	• Advancing toward iterative, exploratory cloud-based analysis. • Emphasizing in-cloud data use to reduce unnecessary downloads and support access to only required data subsets • Modular compute services will be interoperable, reusable, and deployable across the repository, enabling agile, open, and collaborative infrastructure to support Earth science data and decision-making.	• Publications primarily discovered through APIs and search engines such as Google Scholar.
HPD	• Pursuing a multi-tiered strategy for computing on the cloud and on premises based on differing user requirements. • Reducing barriers to computing while managing costs.	• Publication discovery via searches of Sci-X and generally through generic internet searches. Efforts are underway to increase the interlinking between HDRL-hosted resources and the publications and software they support.
PSD	• Expanding cloud-based data availability and enabling in-cloud analysis without requiring full data downloads. • Balancing cost management with broad accessibility.	• Publications are discovered via searches of Sci-X and generic internet searches.

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 6).

Table 6. Metrics Report 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
Michael Allen	OCSDO	Report Formatting
Amanda Adams	OCSDO	Communications Lead

ACRONYMS

Acronym	Full Term	Context
ADS	Astrophysics Data System	Publication discovery service
AI	Artificial Intelligence	Technology for automated analysis
API	Application Programming Interface	Method for programmatic data access
APD	Astrophysics Division	One of five SMD science divisions
ARC	Ames Research Center	NASA center
AWS	Amazon Web Services	Cloud service provider
BPS	Biological and Physical Sciences	One of five SMD science divisions
CPU	Central Processing Unit	Computing resource
CUI	Controlled Unclassified Information	Data classification level
DAAC	Distributed Active Archive Center	ESD data management facility
DOI	Digital Object Identifiers	Standard for citing datasets
ESD	Earth Science Division	One of five SMD science divisions
EOSDIS	Earth Observation System Data and Information System	ESD repository at GSFC
Fornax		APD platform for tiered data access and cost management
FTPS	File Transfer Protocol Secure	Secure file transfer method
FY	Fiscal Year	Budget/reporting year (FY2024)
GIBS	Global Imagery Browse Services	ESD service
GES DISC	Goddard Earth Sciences Data and Information Services Center	ESD DAAC
GIS	Geographic Information System	Spatial data analysis tools
GMAO	Global Modeling and Assimilation Office	ESD collaboration partner
GPU	Graphics Processing Unit	Computing resource
GSFC	Goddard Space Flight Center	NASA center
HEASARC	High Energy Astrophysics Science Archive Research Center	APD repository at GSFC
HEC	High-End Computing	Advanced computational resources
HelioCloud	Heliophysics Cloud	HPD computing resource
HDRL	Heliophysics Data Retrieval Library	HPD federated repository system
HLSP	High-Level Science Products	APD contributed data products
HPD	Heliophysics Division	One of five SMD science divisions
HTTPS	Hypertext Transfer Protocol Secure	Web data access method
IP	Internet Protocol	Network addressing
IPAC	Infrared Processing & Analysis Center	Host institution for IRSA
IRSA	Infrared Processing & Analysis Center (IPAC) Infrared Science Archive	APD repository at CalTech/JPL
JPL	Jet Propulsion Laboratory	Host institution (part of CalTech)
MAAP	Multi-Mission Algorithm and Analysis Platform	ESD analysis platform

MAST	Mikulski Archive for Space Telescopes	APD repository at STScI
MSFC	Marshall Space Flight Center	NASA center
NAIF	Navigation and Ancillary Information Facility	PDS data hosting facility
NAS	NASA Advanced Supercomputing	NASA's supercomputing facility
NCCS	NASA Center for Climate Simulation	NASA facility for climate/weather modeling
NEXUS	Roman Research Nexus	APD cloud-based analysis platform
NSSDCA	National Space Science Data Center Archive	Long-term archival facility
OCSDO	Office of the Chief Science Data Officer	Data management
ODR	Open Data Registry	AWS service for free egress data
OSDR	Open Science Data Repository	BPS repository at ARC
PB	Petabytes	Unit of data storage
PDS	Planetary Data System	PSD repository at GSFC
PI	Principal Investigator	Research grant holder
PSI	Physical Sciences Informatics	BPS repository at MSFC
PSD	Planetary Science Division	One of five SMD science divisions
RAM	Random Access Memory	Computing resource
ROSES	Research Opportunities in Space and Earth Sciences	Research proposal program
RRN	Roman Research Nexus	APD analysis platform
S3	Simple Storage Service	AWS cloud storage service
SAR	Synthetic Aperture Radar	Type of sensor/data (Open SAR Lab)
Sci-X	Science Explorer	Publication discovery/search system
SDAC	Solar Data Analysis Center	HPD repository at GSFC
SIPS	Science Investigator-led Processing Systems	ESD data processing systems
SMCE	Science Managed Cloud Environment	BPS cloud computing environment
SMD	Science Mission Directorate	NASA's directorate managing scientific data
SPDF	Space Physics Data Facility	HPD repository at GSFC
STScI	Space Telescope Science Institute	Host institution
TB	Terabytes	Unit of data storage
TIKE	Time series Integrated Knowledge Engine	APD analysis platform