

# 2025 NASA OPEN SOURCE SCIENCE DATA REPOSITORIES WORKSHOP REPORT

Science Mission Directorate

December 5, 2025



# TABLE OF CONTENTS

<b>CONTRIBUTORS .....</b>	<b>1</b>
<b>EXECUTIVE SUMMARY .....</b>	<b>2</b>
Workshop Metrics .....	3
Workshop Structure .....	4
<b>DAY ONE .....</b>	<b>4</b>
Keynote Talks .....	4
Plenary Session.....	5
SMD Data Governance .....	6
Round Robin Activity .....	6
Posters & Lightning Talks.....	10
<b>DAY TWO.....</b>	<b>11</b>
Repositories' Latest/Greatest .....	11
Getting Ready for an AI World .....	12
Breakout Sessions.....	13
<b>DAY THREE .....</b>	<b>18</b>
Do More With Less .....	18
<b>CONCLUSIONS AND FUTURE DIRECTIONS .....</b>	<b>19</b>
Major Workshop Takeaways .....	20
Next Steps .....	20
Suggested Workshop Changes.....	21
<b>RESOURCE LINKS.....</b>	<b>23</b>



# CONTRIBUTORS

## Planning Committee

Kaylin Bugbee (DASP)  
Deborah Smith (DASP)  
Rebecca Ringuette (HP)

## Session Leads

Andrew Mitchell (OCSDO)  
Slesa Adhikari (ES)  
Subhayu Sen (BPS)  
Cassy Turner (BPS)  
Sara Lubkin (ES)  
Candida Dewes (ES)

## Support Staff

Tri-an Cao (DASP)  
Fred Kepner (DASP)  
Rachel Wyatt (ODSI/DASP)  
The DASP Governance team  
The DASP Training & Learning team



# EXECUTIVE SUMMARY

The 2025 NASA Open Source Science Data Repositories Workshop was held on August 25-27, 2025 in Huntsville, AL and brought together data stewards and data providers from the Science Mission Directorate (SMD) and the broader NASA data governance community. The workshop was hosted by the Data and Analysis Services Project (DASP) and supported by the NASA Office of Chief Science Data Officer (OCSDO) and the Core Data and Computing Services (CDCS) program element. The workshop was designed to solicit feedback on current governance initiatives and to highlight accomplishments, common challenges and best practices across SMD.

Workshop interactive sessions included an overview of the SMD governance efforts spearheaded by DASP, a showcase of tools and projects developed by the SMD repositories, and topical sessions such as the use of artificial intelligence (AI) in data governance, the impact of metrics, redeployable and reusable platforms, documentation, and the data lifecycle. This public hybrid meeting was well attended both in person and online, with 171 attendees on average during the 2.5-day workshop. All SMD divisions and multiple repositories were represented either in person and/or virtually, in addition to interested stakeholders and students from across the US.

## **The primary goals of the workshop were the following:**

1. Foster and strengthen a community of practice around SMD governance and science data stewardship.
2. Socialize and collect feedback on current SMD governance activities.
3. Identify common pain points and ways to maximize efficiencies across SMD data management.
4. Collectively determine what must be archived and how.
5. Identify community needs and how best to meet those needs to move closer to effective open science practices.

## **Several immediate key outcomes were identified by the end of the workshop:**

- The SMD data stewardship community accepts the governance processes implemented this past year but continues to express the need for input on the prioritization and direction of initiatives and derived solutions.
- The data stewardship community needs a better understanding of the purpose of SMD governance and the type of results expected.

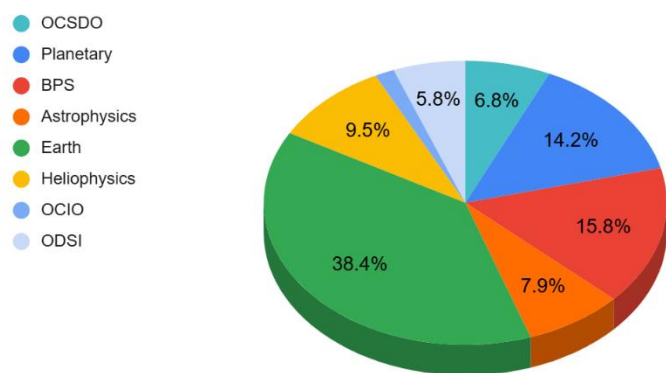
- There is a need across SMD divisions for more training and guidance to increase understanding of the potential applications of AI in data stewardship.
- Improved communication and sharing continues to be essential to the success of governance efforts. This includes giving more opportunities for sharing ideas, guiding solution development and reviewing initiative solutions before implementation.
- Attendees identified the following governance activities to be of high value:
  - Determine what is needed to make SMD data more discoverable by AI tools
  - Further develop guidance for best practices in data preservation, data submission and document delivery
  - Identify key metrics for SMD data repositories to track

Video recordings of the workshop and other resource links are provided at the end of this report.

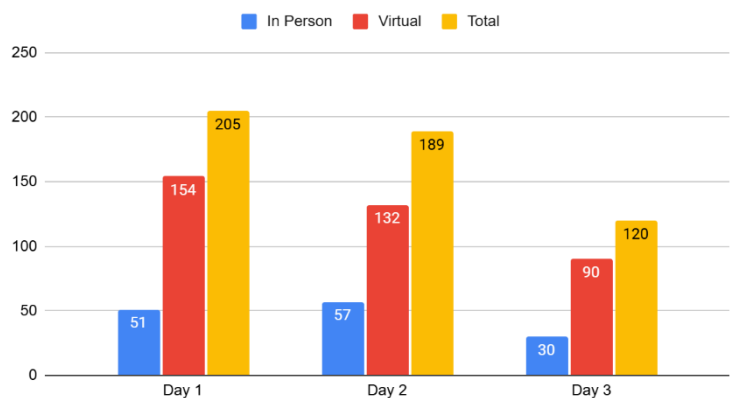
## WORKSHOP METRICS

The workshop was well attended with data officers, data stewards and data producers from each of the SMD divisions. The workshop received 446 registrations in total, with representation from NASA partners and contractors, NASA civil servants, external researchers, and college/university students and professors. In-person and virtual workshop attendance metrics are summarized below, categorized by day and by division. The final half day of the workshop had lower attendance, as is typical due to travel needs.

SMD Attendance Breakdown



Attendance Metrics



## WORKSHOP STRUCTURE

The workshop occurred over two and a half days, each day focusing on governance-related activities occurring across SMD and the NASA data repositories, including current data management accomplishments and challenges. The [meeting agenda](#) provides an overview of the workshop sessions.

- **Day 1** provided participants an opportunity to contribute feedback to existing and planned governance initiatives during a round robin activity highlighting various in-progress work. After the group activity, data stewards gave short poster presentations followed by poster viewing and discussion to close the day.
- **Day 2** explored the use and future of AI in data governance work, while afternoon breakout sessions focused on both repository and data-specific stewardship topics.
- **Day 3** included a session in which attendees generated ideas for more efficient repository operation in times of minimal resources. The workshop concluded with a summary of findings from the breakout groups and a chance for attendees to provide feedback on the workshop in general which will be useful for planning the next one.

## DAY ONE

### KEYNOTE TALKS

*Kevin Murphy*, the Chief Science Data Officer (CSDO), introduced the OCSDO, the office responsible for enabling groundbreaking discoveries by leading the SMD to expand the availability and utility of NASA scientific data. In the past year, the DASP and the Science Cloud Infrastructure Project (SCIP) project offices were established within OCSDO. DASP spearheaded initiatives such as the launch of Open Science 101 (version 2) and the new Open Science Essentials course to promote adoption of open science practices across SMD. In addition, three new foundation models (FMs) trained on scientific datasets from different SMD divisions were developed and released for use (Surya Heliophysics FM, Prithvi-HLS and Prithvi-WxC).

*Kaylin Bugbee*, the DASP technical lead, introduced the project which is responsible for providing SMD-level data, analysis, and user interface services to help scientists,



data producers and data stewards effectively utilize NASA scientific digital resources. She presented each of the sub-projects under the DASP umbrella: SMD Data Governance, Research Repository, Training & Learning, Science Analytics Platform, and the Science Discovery Engine (SDE). Each of these sub-projects focuses on different areas for increasing efficiency across SMD and ensuring NASA's scientific data are discoverable and accessible.

## PLENARY SESSION

*Lori Arnett*, acting Deputy Director for Aerosciences Evaluation and Test Capabilities (AETC) within the Aeronautics Research Mission Directorate (ARMD) was the plenary speaker for the 2025 Data Repositories Workshop. She shared the AETC data governance story outlining why robust data governance is critical for any organization. Arnett's advice to attendees: "think big, start small, and act fast". Her work is about enabling innovation and fostering data stewardship. She also shared that good data governance should focus on organizational outcomes by prioritizing use cases and involving every stakeholder in the process of data stewardship. Thinking differently about data is vital - cultivate a data culture where "data is not just what we deliver, it is how we operate." Tracking measurable progress and ensuring expediency in meeting agency goals, including ethical AI use, are also key. Ultimately, good data governance helps bring all the pieces together to "Get the data, Use the data, Secure the data".



*Lori Arnett presenting the Plenary Session.*

Arnett's message of advice and lessons learned was found, as evidenced by workshop feedback, to be highly applicable to NASA SMD by attendees. A primary take-away is to note that as we establish cross-divisional data governance, collaboration will be key to eliminating duplication and leveraging existing efforts, with implementation needing to be organization-specific. A shift in mindset to avoid a "what's in it for me" mentality is needed.

## SMD DATA GOVERNANCE

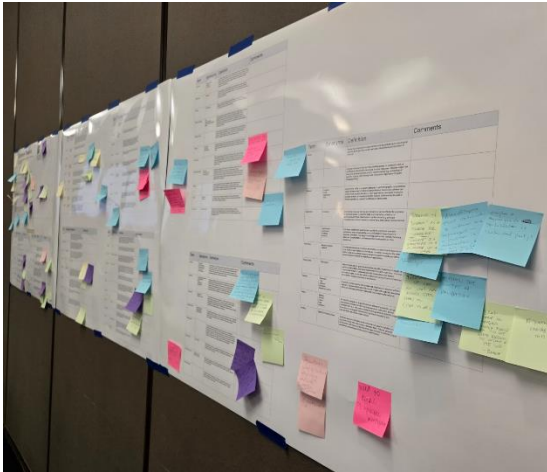
Following the plenary session, *Deborah Smith* (DASP) presented the governance workflow used to move individual SMD-level small-scale, fast-paced initiatives to completion. Each initiative is aimed at finding a solution to a particular governance need. A list of needs is maintained and regularly prioritized by the governing board from which new initiatives are selected as workload permits. Anyone with governance needs and concerns is encouraged to use the [Governance Initiative Proposal form](#) to add to the ongoing list. The status of the first four initiatives currently underway at the time of the workshop were shared with attendees via a round robin activity session after the break. Most of these initiatives resulted from the key takeaways of the [2024 Workshop](#) in Pasadena, CA.

Smith also shared the latest information and mockups of a new resource currently in development. The Governance for Understanding, Accessibility, and Collaboration (GUAC) Community Knowledge Hub is designed to meet the needs of NASA SMD data stewards, data officers, the OCSDO, and the DASP Governance team and will consist of searchable standards, guidelines, best practices and training resources; a document review space; a community user forum; a metrics tracking dashboard; event information, registration and recordings; and a co-pilot to aid users in locating and using needed resources. The expected release date of the GUAC beta version is early 2026.

## ROUND ROBIN ACTIVITY

This session aimed to provide time for both virtual and in-person attendees to share feedback on current SMD governance initiatives. Attendees were split into five groups that cycled through five activity stations. Online attendees used virtual Miro boards for feedback with a governance team member facilitated discussion. In-person attendees gathered around posters showing the status of the initiative solution and responded to questions aimed to collect feedback and spark discussion. Notetakers were also used at each station to capture important comments. All feedback was later gathered and reviewed post-workshop to inform and improve initiative solution development. The five stations are summarized in the following pages.





*Example of “Governance Terms” board.*

governance terms used within NASA’s five science divisions as well as other related entities outside of NASA. This activity generated lots of useful information for completing a draft thesaurus for community review. Release of the approved governance thesaurus is expected by mid-2026.

## Governance Terms and Definition

The Governance Terms and Definitions activity, led by *Emily Foshee* (DASP) and *Sam Berg* (DASP), addressed a previously identified need for development of a common data governance language that can improve both internal and cross-division communications. A draft high-level list was shared containing selected terms, derived definitions and synonyms. This past summer, the DASP governance team collected definitions and synonyms for

## Level of Service Model

A Level of Service Model is another initiative that was included in the round robin activity. As presented by *Deborah Smith* (DASP), this effort was at an early stage of development at the time of the workshop. The activity therefore focused on asking attendees pointed questions about what characteristics they use to distinguish the manner in which they serve different data, such as “How does your repository divide all data into various groups that align with service level needs?” and “What are the requirements and services for each service level?”. The feedback is guiding model development.

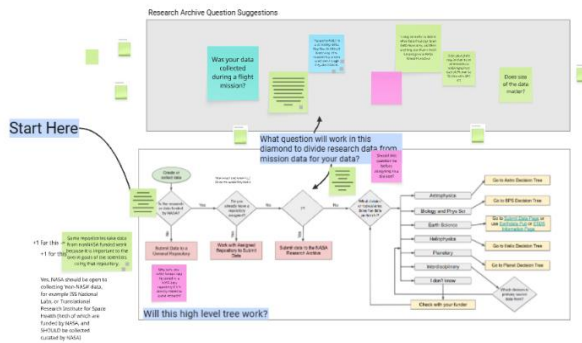


*Deborah Smith leading the Level of Service activity.*

Attendees had widely varying thoughts on what required activities were needed but they did agree that 3 levels were sufficient for the model with near universal agreement on

the use of “Basic”, “Standard”, and “Comprehensive” for the service level names. The governance team will next detail the different requirements for each service level, draft guidance for model use, and present the solution to the community to collect further feedback. The final model solution is expected in January 2026.

## Data Repository Decision Tree



*Example of online Miro board.*

A Data Repository Decision Tree was developed by DASP during the summer of 2025 that provides clarity for data producers ready to submit data to NASA SMD data repositories. *Kaylin Bugbee* (DASP) shared the high-level decision tree and individual division decision tree drafts for feedback and discussion about how repository data placement decisions are made. Through discussions, it was

determined that the high-level decision tree can identify the correct SMD division into which data should be archived, but the actual repository decisions are up to the divisions.

Workshop attendees were prompted to review the applicability and order of existing decision steps and possible ways to differentiate NASA repository holdings for both mission and research data. The feedback collected was used to refine the high-level decision tree and develop a division guidance document that aids divisions in creating their own decision trees. This initiative solution was delivered Sep 24, 2025 for board review and further community input.

## Minimum Metadata

The need for identification of the minimum metadata necessary for data originated from discussions in two previous sessions during the 2024 data repositories workshop. This year, a round robin activity, led by *Rebecca Ringuette* (HPD)



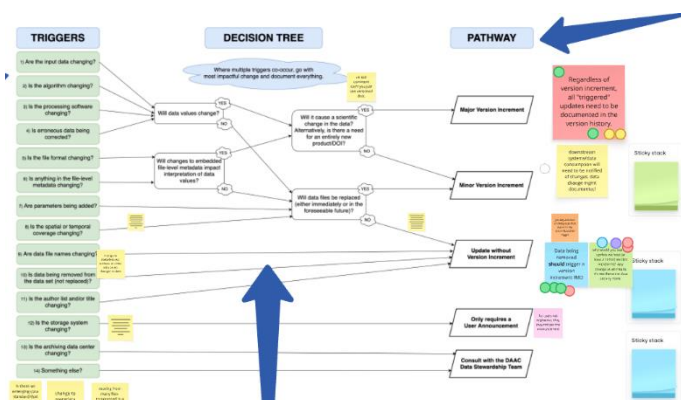
*Rebecca Ringuette leading the Minimum Metadata activity.*

presented a potential minimum metadata list for both mission and research data. Ringuette asked attendees to vote on the proposed metadata fields and to determine requirement levels for each. Suggestions for additional metadata fields were also collected.

Attendees indicated DataCite metadata fields should be mandatory for mission data. For research data, however, an “affiliations” field for all NASA authors was thought to also be mandatory by a small number of votes. All other metadata fields were voted as either recommended or optional. Full voting results and contributed comments are described in Ringuette (2025a).

To determine the final minimum metadata list that will be included in this initiative solution, the governance team will next draft and share an updated minimum metadata list with the community to collect any further comments prior to full release. Guidance materials will also be drafted for use in developing division-wide requirements for specific metadata fields.

## Versioning Decision Tool Concept



Example of online Miro board.

The National Snow and Ice Data Center (NSIDC) within NASA's Earth Science Division (ESD) was also included in the round robin session to share their new Versioning Decision Tool concept and collect feedback.

This tool is one example of a governance activity carried out by an ESD repository. The framework helps data stewards

consistently version published data sets. *Candida Dewes* (ES) guided attendees in reviewing a list of potential triggers that require data updates, such as “Is the algorithm changing?” or “Are data file names changing?”. These triggers result in a sequential versioning decision tree that points users towards the proper documentation and labeling of the highest pathway level identified for that dataset update.

Participants provided feedback on the update scenarios for examples and added comments on unrepresented triggers, on the interpretation of the language used in

labeling data updates, and on the appropriateness of the questions utilized in the decision tree pathways for different use cases. Collected feedback includes a large volume of potential triggers to review, suggestions for how to account for data removals, and expression of the potential for confusion from the use of ‘major and minor version increment’. Next steps include refining terminology used across the tool and reordering decision stages to better represent the optimal pathways.

## Summary of Round Robin Activities

Governance Topic	Primary Finding	Next Steps	Expected Delivery Date
Governance Terms/Definitions	Continued opportunities for term/definition reviews are needed. Advise developing a guidance document.	Share drafts with the community for review.	Dec 1, 2025
Level of Service Model	Identifying division terminology and data differences are essential to model development.	Develop an adaptable model and provide guidance for use.	Jan 15, 2026
Data Repository Decision Tree	SMD governance should not devise division decision trees. However, guidance is needed to assist with understanding SMD expectations.	Deliver final decision tree and guidance.	Oct 1, 2025
Minimum Metadata	A uniform set of required metadata used across SMD repositories improves data discovery and accessibility for all NASA data. However, the metadata list developed must be consistently applied.	Share metadata list with the SMD community.	Mar 1, 2026
Data Versioning Tool	This NSIDC / ESD work could potentially be more broadly applied to SMD divisions.	Add to SMD governance needs list.	N/A

## POSTERS & LIGHTNING TALKS

Day 1 closed with a poster session consisting of a series of two-minute oral presentations followed by physical or virtual poster viewing. The posters covered various topics relevant to governance, data stewardship or data tools. The following posters were shared:

- ESDIS Mission Lifecycle Process: *Beth Huffer (ES)*
- Scaling up the Solar Data Analysis Center: *Jack Ireland (HP)*
- The Importance of Quality Metadata in PDS4 Product Labels: *Tom Stein (PS)*
- Improving Open Science for Heliophysics: *Rebecca Ringuette (HP)*
- Adoption of Open Science Best Practices: *Steve Crawford (OCSDO)*
- FAIR Research Data Archive Needs, Outreach, and Data Literacy: *Sabiya Jamadar (AP)*
- Building Partnerships to Promote Responsible and Transparent Research: *Virginia Pow (ES)*
- Atmospheric Science Data Center (ASDC) Suborbital Data Ingest and Curation: *Megan Buzanowicz (ES)*
- DUTC Open Science: Key Highlights: *Dorra Saidi (BPS)*
- Communities Yielding Power by Harnessing Emergent Resources (CYPHER) Using Open Science: *Amy Quarkume (ES)*
- The Life Detection Knowledge Base: A Community Repository for Life Detection Science: *Svetlana Shkolyar (PS)*
- Starting Small, Thinking Open: Building a Grassroot Open Science Community for ECRs: *Aswathi Surendran (PS)*
- Physics AI Data Lake Proposition: *Yan Barros (AP)*
- The Heliophysics Software Search Interface: *Julie Barnum (HP)*
- Open Science and Public Universities (Mexico): *Ángel Escobedo (ES)*
- Vymanika Aerospace: From Ancient Inspiration to Modern Innovation: *Pinkey Bartake (AP)*

# DAY TWO

## REPOSITORIES' LATEST/GREATEST

Day 2 started with presentations by SMD division representatives sharing two-minute Lightning Talks. Each talk featured divisional projects, specialized tools, and cloud data solutions.

The session contained the following talks:

- Dataset Curation (data models, ontologies, etc): *Kris Peach (BPS)*
- PO.DAAC Virtualized Datasets: *Mike Gangl (ES)*
- Microgravity Analytics Platform: *Subhayu Sen (BPS)*
- Metadata standards and data curation: *Lan Jian (HP)*
- NISAR and Vertex at ASF DAAC: *Franz Meyer (ES)*
- Community Engagement - Analysis Working Groups: *Ryan Scott (BPS)*



*Main conference hall.*



- The Helio.data.nasa.gov website: *Brian Thomas (HP)*
- The Roman Research Nexus Cloud Platform: *Jonathan Hargis (AP)*
- Tips for Unifying Your Web Presence: *Andi Brinn Thomas (ES)*
- SciX: *Alberto Accomazzi (cross-division)*
- PDS Atmospheric Modeling Annex: *Lynn Neakrase (PS)*
- A data visualization solution for solar physics: *Jack Ireland (HP)*
- AI User Assistant: *Subhayu Sen (BPS)*
- Astromat: *Kerstin Lehnert (PS)*
- mEditor: Reimagining Documents, Metadata, Configuration, and Workflow Automation: *Jon Carlson (ES)*

After all talks, each attendee voted for their favorite by distributing 100 points as desired. The highest voted talks included: Alberto Accomazzi's "SciX" talk, Ryan Scott's "Community Engagement - Analysis Working Groups" talk and Kris Peach's "Dataset Curation" talk. There was significant workshop feedback indicating this session approach was favored by many and it should be used again next year.



*Session speakers (left to right): Andrew Mitchell, Kevin Murphy, Rahul Ramachandran, Kaylin Bugbee, Jonathan Galazka.*

## GETTING READY FOR AN AI WORLD

*Andrew Mitchell* (OCSDO) moderated this session, which opened with remarks from four panelists: *Kevin Murphy* (CSDO), *Kaylin Bugbee* (DASP), *Rahul Ramachandran* (ODSI), and *Jonathan Galazka* (BPS). Each panelist shared their envisioned future of AI tools in SMD data governance and the

importance of AI advancements for their specific work and for data science innovation in general.

Panelists stressed the importance of data science innovation, highlighting how AI has become a fundamental part of doing science. More and more, AI functions as a partner with scientists to analyze large amounts of data, speed research and assist with day-to-day processes. AI allows scientists and data producers to think strategically about their work and cut down on time spent discovering data or information for research.

For instance, Ramachandran described the development of Accelerated Knowledge Discovery (AKD), a modular, agentic AI framework in development that will enhance the



scientific research process. By streamlining literature search and data discovery, AKD will be one of the many AI tools that can be used by SMD division scientists in their research. However, the researchers themselves are still crucial in the validation of agentic tools. Guardrails and fact-checking remain vital components in the research process. The panelists fielded audience questions and encouraged everyone to look forward and identify ways in which AI can improve data stewardship and science.

## BREAKOUT SESSIONS

The afternoon of Day 2 consisted of two rounds of concurrent sessions focused on topical group discussions. The first round was related to broader data repository concerns: impact metrics measurement and the challenges of serving cloud data. The second round of sessions focused more specifically on data and data management: understanding the data lifecycle and its relationship to data stewardship tasks and the need for data documentation standardization. All session notes and content can be found in the [workshop Zenodo repository](#).

### Session 1A: Metrics that Measure Impact

*Rebecca Ringuette* (HP) continued last year's discussion on ways to use available metrics to assess the impact of hosted NASA data. This year, a presentation and demonstration of the capabilities of the Science Explorer ([SciX](#)) for data citation metrics was provided by *Alberto Accomazzi* (cross-division). SciX offers data citation metrics with enhanced accuracy over that of existing publisher services for NASA SMD science area publications.

#### Metrics considered to be automatically trackable:

- Direct citations (either to the data/software or the publication)
- Secondary citations (e.g. citations to the publication(s) the data/software is cited in)
- Number of authors/contributors
- Number of accesses and downloads
- Original data (PI) use vs. archival use (follow-up studies)
- Distance between original proposal and published studies (e.g. overlap of keywords)
- Number of times the data/page shows up in federated search
- Engagement with or of the user community
- Mentions of datasets (i.e. in the caption of the figure) but not a formal citation

This was followed by a discussion examining the issue of measuring data impact. Metrics-based impact assessments can only provide an imperfect and incomplete picture, particularly for some important data impact aspects that are not measurable by

metrics. The attendees worked together to produce a list of metrics that could be used to measure impact, of which many could be automatically trackable given existing capabilities on dataset landing pages, SciX services, and DataCite DOI metadata. A few metrics on the list were deemed important but difficult to measure (see [Ringuette 2025b](#) for details).

Based on these discussions, the major findings and outcomes of the session were the following:

- Automatically trackable impact metrics give an imperfect and incomplete portrayal of the actual impact of a dataset.
- A large sampling of metrics can be accumulated based on the existing infrastructure, including dataset landing page monitoring, although some improvements are needed.
- Other important, but more difficult to obtain, metrics need further development or infrastructure to enable automated tracking, while still others are likely to remain only obtainable by human assessment.

#### **Suggested next steps for improving impact metrics tracking:**

- Gather more community input and classify each metric as currently automatically trackable, manually trackable, or non-trackable.
- Encourage cross-repository collaboration to increase automated reporting for trackable metrics based on existing infrastructure.
- Prioritize the effort and support needed to make important manual and non-trackable metrics automatically trackable.

### **Session 1B: Challenges of Serving Data on the Cloud**

*Slesa Adhikari* (ODSI) chaired this breakout session focused on the technical and cultural challenges of delivering NASA science data to users via cloud platforms. Four lightning talks were given that described the VEDA (*Brian Freitag*, ODSI), CryoCloud (*Tasha Snow*, ES), Fornax (*Jonathan Hargis*, AP), and HelioCloud (*Brian Thomas*, HP) tools which actively serve NASA data on the cloud. Shared experiences across the Earth science, cryosphere, astrophysics, and heliophysics communities were identified and found that NASA's science divisions are independently solving similar challenges.

Rather than converging on a single platform, the community is aligned on a shared approach: using the cloud as a collaborative research environment, enabled by common standards, shared tooling, and continuous community engagement.

The major findings and outcomes from this session were the following:

- The biggest barrier is not just “Big Data,” but “Big Variety.”
- Users struggle to locate and work across fragmented datasets, formats, and metadata standards.
- STAC and open data formats (COG, Zarr, FITS, HDF5) are enabling interoperability, but legacy datasets need streamable/cloud-optimized conversions.
- Cost efficiency is as much a *training problem* as an infrastructure problem. Most wasted costs come from users forgetting to clean up storage or misusing compute power.
- JupyterHub-based cloud platforms are becoming collaboration hubs, not just compute portals. Shared software images, tutorials, and communication channels (e.g., Slack) are essential.
- Early-career users and community hack weeks have been the strongest cloud adoption drivers.

Recurring Divisional Cloud Data Needs	
Topic Area	Priority Needs
Interoperability	Standardized metadata and tooling
Cost Management	Built-in cleanup tools plus quota, dashboard, and user monitoring
Usability	Low-barrier entry points such as visualization- first and code snippets co-located with datasets
Collaboration	Cross-division coordination on JupyterHub operations and cloud tooling

#### The next steps to address identified cloud data challenges:

- Continue monthly cross-Hub coordination meetings (currently led by CryoCloud).
- Develop shared cost-control and onboarding playbooks for new cloud users.
- Advance virtualized/subset-based data services (e.g., VEDA virtual cubes, Fornax postage-stamp API).
- Collect use cases to use in brainstorming ideas for a unified platform that can be reused across divisions.

## Session 2A: Standardizing Data Documentation

In this session led by *Subhayu Sen* and *Cassy Turner* (BPS), attendees explored ongoing efforts to standardize data documentation practices across the five divisions. Brief presentations by *Candida Dewes* (ES), *Lan Jian* (HP), and *Subhayu Sen* (BPS)

gave examples of the types of documentation currently required within projects for both mission and data product-level documentation. These presentations uncovered both the diversity in existing documentation practices across NASA SMD and some common practices. Feedback from attendees explored what is considered required and supplemental documentation and what challenges are faced when it comes to data documentation.

Some of the commonalities, although called by different names across divisions, include experimental or parameter tables, readme documents for explaining both experimental and software/modeling data and publications and reports to name a few. The biggest challenge identified was expediently obtaining a complete set of required documentation and information from the science team. The discussions pivoted to the potential role of AI in documentation standardization and how AI tools could assist in generating documentation elements thereby improving consistency and reducing manual effort. The conversation also explored how investigation products need to be generated in such a way that they are actionable by AI tools.

The common issues across divisions are identified as follows:

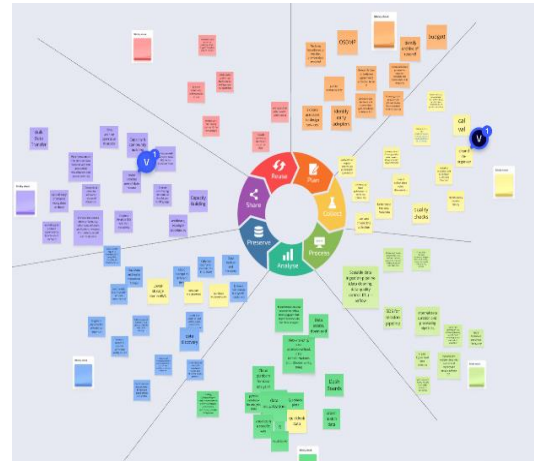
- There is a need for clear guidelines on the use of consistent terminology.
- Additional funding is needed for repositories to meet new documentation requirements.
- Overall, a shared understanding of the role of standardized documentation was appreciated, and attendees showed interest in continuing collaboration to explore how AI tools can be used to address shared documentation challenges.

### **Suggested next steps for document standardization:**

- Continue research on how tools, specifically the use of AI, can be used to reduce data ingestion and curation efforts and thereby speed data delivery.
- Ensure a common set of terms and guidance are developed and incorporated into division efforts.

## Session 2B: Data Lifecycle & Governance

*Deborah Smith* (DASP) and *Andy Mitchell* (OCSDO) facilitated this session aimed at increasing the understanding of the concept of a data lifecycle and how it can best be used in data governance and stewardship. Following an introduction to various types of data lifecycle diagrams containing varying stages and terms, attendees added to a Miro board the activities they thought were typically associated with each lifecycle stage.



*Example of online Miro board.*

Discussions followed exploring the divisional differences with respect to what activities are completed when and which activities are most important to good governance and strong data stewardship.

*Bruce Berriman* (AP) closed the session with a presentation titled “The Role of Virtual Observatory Protocols in Astronomy (meta)Data Governance”. This talk shared the growth of the virtual observatory alliance and how the use of standards enhanced data access and visualization across not just NASA missions, but also worldwide.

The primary findings of the session included the following:

- Attendees gained an appreciation of the differences and similarities in various tasks performed throughout the data lifecycle across divisions.
- There is a need for a common language and to reference a common lifecycle across SMD.
- Determination of what users need to use/reuse data is essential to providing valuable data services.

### **Some important next steps for data lifecycle processes:**

- Increase cross-division coordination and collaboration with respect to data lifecycle stewardship activities.
- Improve the use of standards to yield better data products and data management outcomes.
- Identify who is responsible for ensuring that data tasks are properly completed in order to provide high quality open data to users.

# DAY THREE

## DO MORE WITH LESS

*Sara Lubkin* (ES) designed this session to address the growing challenges faced by data repositories, specifically, the need to manage and archive increasing volumes of data and services while operating under reduced budgets. Presentations led by *Beth Huffer* (Earthdata Pub, ES) and *Walter Alvarado* (AI Curation Assistant, BPS) sparked a high level of engagement with attendees. Attendees then began brainstorming ideas for better managing data repositories, regardless of feasibility. Attendees broke into five groups to evaluate and refine the collected ideas. Discussions emerged around automation, collaboration, and simplification of data processing and analysis. Ultimately, the session surfaced a wide range of creative and actionable ideas to help data repositories remain effective and resilient amid growing demands and shrinking resources.

### Some of the “Do More with Less” ideas generated:

- Automating data checks and curation workflows.
- Using high-quality training data to enable AI tools to support metadata compliance, though human quality checks remain essential.
- Enhancing cross-division communication, collaborating on open-source solutions.
- Engaging with successful science communities like SciPy or Jupyter to learn from their practices.
- Simplifying and minimizing reporting requirements to focus only on meaningful metrics.
- Standardizing reporting across data centers.
- Reducing unnecessary administrative burden by communicating transparently.



# CONCLUSIONS AND FUTURE DIRECTIONS

The workshop concluded with brief breakout session reports and use of a Miro board to collect initial workshop feedback. Attendees were then given links to a post-workshop survey for more substantial feedback afterwards. The feedback received indicated that the meeting was successful in achieving the goals of building community, sharing SMD governance progress, learning of new tools and services, and discovering ways to maximize efficiencies. Many in-person and virtual participants expressed a desire to return to the workshop next year to further discussions and learn more from those in other divisions. The new Coffee Chats hosted by the DASP governance team will ensure communication, discussions and community interaction continues throughout the year. These events will occur bi-monthly and consist of governance progress for the first monthly meeting and division sharing for the second monthly meeting.

## MAJOR WORKSHOP TAKEAWAYS

Despite the large online attendance and reduced number of in-person attendees, the workshop was still productive and led to collection of significant feedback for ongoing governance initiatives.

The following workshop major takeaways were identified:

- Those in attendance agreed that SMD-level governance is needed and felt the process presented for addressing needs was satisfactory. However, concern remained about whether data stewards would have enough input to the prioritization of initiatives that DASP will work on.
- Improved communication and sharing will be essential to the success of governance efforts. Those in attendance wanted more opportunity to share their ideas, guide solution development, and review initiative solutions before required implementation. The workshop is only one opportunity. The DASP team started hosting biweekly virtual Coffee Chats in September. These meetings plus quarterly updates from OCSDO and data officers will provide improved communications.

- A complete understanding is needed by all stakeholders as to what an initiative is and what the overall governance effort entails. Some of the feedback received at the workshop was either off-topic, too detailed or outside the scope of the work. This indicates the need for greater clarity on the purpose of governance at the SMD level and the type of results to expect.
- The questions and comments received during the AI session indicate that more training, guidance and help with envisioning ways to utilize AI in vital stewardship efforts is needed to improve productivity and efficiency. DASP governance is looking toward the OCSDO to provide the overall vision.

## NEXT STEPS

The following next steps are devised to address some of the communication and governance issues identified:

Next Steps
The DASP governance team will proactively engage the community on initiative solution development and also regularly collect feedback on the governance process. Process improvements will be made based on the feedback where possible.
The DASP team will continue to host the annual workshop utilizing a planning committee of volunteers from each division. The DASP team will also ensure Coffee Chats and quarterly updates take place regularly and all data stewards are notified of upcoming presentations. Every other Coffee Chat will highlight data steward activities within divisions/repositories so long as volunteers offer to present.
We will use the Coffee Chats not only for sharing the DASP initiative progress and solutions, but also for educating the community on what governance is and the level and scope expected in DASP governance work. This may help improve the community feedback to better match what is needed and that can be effectively used.
The DASP team looks to OCSDO to provide vision and guidance to the divisions regarding AI use and expected outcomes. The governance team will incorporate AI wherever appropriate and possible in ongoing work and new initiatives.

## SUGGESTED WORKSHOP CHANGES

While many suggestions for next year came up during the workshop, the following points are recommended for consideration for next year. A full summary of attendee feedback is available in the resources section below.

The following points should be taken into consideration for the 2026 workshop:

- The 2024 workshop used a planning committee to organize the workshop after which it was decided to plan this year without a committee. Upon completing the 2025 workshop, the governance team is recommending that planning for 2026 take place with the use of a committee of division representatives. This will help to ensure more community involvement and buy-in.
- Much of the attendee feedback focused on the timing and pace of the workshop. We suggest the 2026 workshop be extended from 2.5 to 3.5 or 4 days. This will address pacing concerns while maintaining the much appreciated longer breaks and lunches at which more informal and important interactions take place. As to workshop timing, a survey should be sent early in 2026 to find the most advantageous time.
- There was significant feedback from attendees to improve the agenda clarity. This would include adding time zone, room numbers or names, wi-fi information, and other needed logistics. This advice will be followed for 2026.
- Feedback indicated the need for continued sessions on AI implementation and use in data stewardship with an expectation presentation from the CSDO to kick it off.
- The poster session in the 2025 workshop did not work well due to lack of presentation space, little time for poster discussion and the large number of online attendees that had no effective way to participate with in-person attendees. Use of a poster session should be carefully considered for 2026 and, if used, improved to yield greater interaction for all attendees.
- Concern was again expressed over the difficulties of online attendees during session question and answer periods despite efforts to improve over last year. While the audio was better this year and in-person attendees were requested to use the online service for questions, in-person discussions took over. Online attendees still felt their questions or comments were not given priority. We

recommend using an online facilitator that will direct the question/comment period in the workshop.

- Due to unforeseen circumstances, there were problems with in-person participation this year resulting from providing a fully open meeting.
- This year, we added training for Miro Board use on the first day of the workshop. Miro boards are an effective means to collect feedback from both in-person and online attendees, leveling the means for participation. We recommend adding the training to a coffee chat prior to the 2026 workshop and providing the recording of the training to those that can not attend so all are more comfortable with this type of tool before the workshop begins.

The insights and findings from the workshop will continue to support SMD efforts in data governance and stimulate partnerships for sharing of best practices among the divisions. Continued collaboration, innovation, and standardization will more broadly promote open science and produce interoperable data and information governance initiatives. DASP recommends the continuation of this important yearly workshop as it builds the community needed to successfully implement governance solutions.

# Resource links

[Workshop Agenda](#)

[Zenodo Repository](#)

[Breakout Sessions](#)

[Recordings](#)

[Posters](#)

[Round Robin Materials](#)

## ACRONYMS

Description	Acronyms
Accelerated Knowledge Discovery	AKD
Aeronautics Research Mission Directorate	ARMD
Aerosciences Evaluation and Test Capabilities	AETC
Artificial Intelligence	AI
Astrophysics Division	AP
Atmospheric Science Data Center	ASDC
Biological & Physical Sciences Division	BPS
Communities Yielding Power by Harnessing Emergent Resources	CYPHER
Core Data and Computing Services	CDCS
Data and Analysis Services Project	DASP
Digital Object Identifier	DOI
Don't Use This Code	DUTC
Earth Science Division	ES
Findable, Accessible, Interoperable, and Reusable	FAIR
Foundation Models	FMs

Governance for Understanding, Accessibility, and Collaboration	GUAC
Heliophysics Division	HP
NASA Earth Science Data and Information System	ESDIS
NASA-ISRO Synthetic Aperture Radar	NISAR
NASA's Alaska Satellite Facility Distributed Active Archive Center	ASF DAAC
National Aeronautics and Space Administration	NASA
National Snow and Ice Data Center	NSIDC
Office of Data Science and Informatics	ODSI
Office of the Chief Science Data Officer	OCSDO
Physical Oceanography Distributed Active Archive Center	PO.DAAC
Planetary Data System (Version 4)	PDS4
Planetary Science Division	PS
Science Cloud Infrastructure Project	SCIP
Science Discovery Engine	SDE
Science Explorer	SciX
Science Mission Directorate	SMD
Visualization, Exploration, and Data Analysis	VEDA