



EARTHDATA

Artificial Intelligence in Earth Science Data Systems

ESDS AI Strategy FY27 – FY30

April 28, 2026

Revision Table

Document Revision	Effective Date	Description
	April 28, 2026	Baseline

Table of Contents

- Background 1
- The Problems We Are Solving 2
- Why This Will Work 2
- How We Work: Operating Principles and Positioning 3
- The Four Pillars of Our AI Strategy 4
- Cross-Cutting Capabilities 5
- Boundaries: What We Do Not Do 7
- How We Execute: Five Tactics 7
- How We Know If It Is Working 9
- What This May Look Like in Practice 9

Background

NASA has collected nearly 180 petabytes of Earth science data over the past five decades of satellite observation. That number is accelerating toward 600 petabytes by the early 2030s. If our systems cannot scale to meet that growth, we risk becoming invisible in the workplaces that matter most: where researchers are doing science. This strategy defines how we use AI to close that gap before it closes us out.

The Earth Science Data Systems (ESDS) program exists to make NASA's free and open data interactive, interoperable, and accessible for research and societal benefit. Our primary goal since 1994 has been to maximize the scientific return from NASA's missions and experiments. We process instrument data into Earth System Data Records, actively manage the archive as a national asset, set the standard for efficient production and stewardship of science-quality data, and lead the research and technology development that keeps our data systems ahead of mission needs.¹

That mission has not changed. What has changed is the scale of the challenge. The processing pipelines, discovery interfaces, and analytical workflows that serve our community were built for a different era.

This strategy defines how ESDS applies artificial intelligence (AI) to close that gap. We aim to apply AI across the full data lifecycle from production and curation, through discovery and access, to analysis and insight. At every stage, the goal is the same: shrink the time between data collection and scientific understanding.

This document is an internal organizing framework as much as it is a public strategy. All ESDS AI-focused work should find its home in one of the four pillars of this strategy: infrastructure, production, access, and analysis. Teams should be able to articulate how their efforts advance both this AI strategy and broader program goals. Our choices about where to invest, partner, and govern should reflect these priorities.

AI infrastructure and service providers will shift. Costs, alliances, and market dynamics are unpredictable. This strategy is designed to be durable precisely because it is not tied to any specific vendor or platform. We are building capabilities and practices that will scale as our system's architecture evolves, while building fluency and institutional knowledge within the current team.

¹ [About the ESDS Program | NASA Earthdata](#)

The Problems We Are Solving

ESDS's program goals include advancing open science data systems for the next generation of missions, leading innovation for managing complex Earth science data, and setting the standard for efficient production and stewardship of science-quality data. These goals were achievable at previous archive scales, but with more than 170PB now in our archives, this has already become difficult to achieve.

Our archive grows faster than our community's ability to navigate it. A researcher looking for the right data for their science question may spend considerable time identifying the right datasets, understanding their quality and coverage, and getting them into a usable form. That time is not science. It is friction between our data and the scientific return.

The same friction exists on the production side. Processing pipelines are typically built for a single mission, with scientific decisions embedded in code that is rarely documented and difficult to transfer. Over time these systems become monolithic: tightly coupled to specific hardware and software environments and to institutional knowledge that is hard to pass on. As technology evolves, maintaining legacy processing pipelines grows more expensive and adapting them for new missions or new science questions becomes increasingly difficult. Reproducibility suffers because the processing history was never fully captured. Reuse is limited because the architecture was never designed for it. This is not a scale problem; it is a transparency and portability problem, and it compounds with every new mission that inherits the same patterns.

Engaging the Earth science community in the evolution of our data systems also requires that we meet that community where it is headed. Researchers increasingly work through AI-assisted tools.² If NASA data is not findable and usable in those environments, we cede our role as the authoritative source for Earth science data.

AI does not solve all of these problems. But applied thoughtfully, it can make each stage of the data lifecycle significantly faster, cheaper, and more reliable, freeing our teams to focus on the work that demands human expertise.

Why This Will Work

When ESDS applies AI across the data lifecycle, the people who create and manage NASA data can spend more time on work that requires human judgment. They can do more with the same resources, move faster, and proactively address problems earlier.

² [Survey of researchers shows active AI adoption for core scientific tasks](#)

When production and stewardship systems work better, the data that reaches researchers is richer, more reliable, and easier to use. Researchers spend less time navigating, reformatting, and validating, and more time doing science.

This is how ESDS turns AI investment into scientific returns. Every pillar and tactic in this strategy connects back to that lifecycle.

How We Work: Operating Principles and Positioning

The principles below are not aspirational; they are constraints defining what ESDS does not do as much as what it will. They also align with Executive Order 14303, Restoring Gold Standard Science,³ which directs federal agencies to conduct scientific activities in a manner that is reproducible, transparent, and rigorous. Our strategy is designed to strengthen those qualities in NASA Earth science data.

How We Operate

These commitments define how ESDS conducts its AI work. They apply to every project, every partnership, and every decision about where to invest time, effort, and funding. The test is simple: does this accelerate the path from data to scientific discovery?

Balancing stewardship with acceleration

AI moves fast, yet quality and provenance are non-negotiable. We do not sacrifice accuracy or traceability for speed. We would rather be slower and right than faster and wrong.

Shared infrastructure, not proprietary capability

We build infrastructure and access layers; we do not own the science. When we develop AI capabilities, we build them as shared infrastructure: open, documented, and designed for the community to adopt, extend, and build upon.

We leverage existing models and frameworks rather than building our own. Our comparative advantage lies in the data and the systems that make it accessible. AI capabilities, algorithms, and workflows developed with ESDS funding are expected to be openly available and consistent with NASA open science policy. Funded agreements will specify intellectual property (IP) terms explicitly, as needed.

³ Executive Order 14303 of May 23, 2025: Restoring Gold Standard Science. Office of the Federal Register.
<https://www.whitehouse.gov/presidential-actions/2025/05/restoring-gold-standard-science/>

Embedded collaboration

ESDS intends to work alongside mission teams and the research community as a technical partner, with shared accountability for the path to science. This approach may take longer to show early results, but it builds trust, strengthens community, and produces outcomes that last.

How We Position

These stances define where ESDS stands relative to adjacent domains and communities. They tell partners, mission teams, and our staff what to expect from us and what not to.

GeoAI and adjacent domains

We enable domain-specific applications through excellent data infrastructure. We do not validate, endorse, or take responsibility for domain-specific scientific outputs.

Model data: federation and co-hosting

We want to expand model data integration with our system, bringing observational and model data together through operational experience and community engagement, rather than top-down mandates.

Gold Standard Science alignment

Our AI strategy directly supports EO 14303, Restoring Gold Standard Science. Provenance-first outputs, reproducibility scaffolding, transparent metadata, and open interfaces advance the order's tenets of reproducibility, transparency, and rigor. AI in our systems strengthens these qualities. It does not shortcut them.

The Four Pillars of Our AI Strategy

Our AI strategy is organized around four pillars that map to the data lifecycle. Every ESDS AI activity should connect to at least one pillar. The pillars are not rigid, and work can span more than one. But every effort should have a clear primary home.

Pillar	What it covers
Infrastructure	Cloud infrastructure, metrics, metadata systems, cloud-native formats, AI-readiness standards, and open machine-readable interfaces. This is the foundation that makes all other pillars possible, including the ability of external AI tools to reach the archive directly.
Production	AI for the teams and systems that enable the creation of data. Pipeline operations, quality assurance at scale, metadata generation at ingest, knowledge surfacing across the production community. Projects and mission delivery obligations.
Access	Semantic search, application programming interfaces (APIs), discovery interfaces, and the standards that make NASA data findable by both humans and AI agents. GeoAI and the expanding model data ecosystem.
Analysis	<p>AI workflow augmentation for researchers. Reproducibility scaffolding, cross-mission synthesis, and embedded collaboration with mission Data Applications Research Technology (DART) teams. ESDS enables these capabilities; we do not own the science that results.</p> <p>Looking ahead, we also want to enable workflows that span observational and model-based data: reanalysis products, projections, and similar tools where AI can reduce time and cost without generating new science products. This is a natural area for closer collaboration with the Integrated Modeling Virtual Institute (IMVI) and similar efforts.</p> <p>This pillar represents the highest-leverage but longest-horizon investment. Near-term activity will be targeted and opportunistic; broader enablement will scale as resources allow.</p>

Cross-Cutting Capabilities

Two capabilities run across all four pillars and are not owned by any single one. They apply to everything ESDS builds and operates, and every team has a role in maintaining them.

Governance and standards

Policies, review processes, metadata requirements, and interoperability standards ensure consistency and trust across everything we build.

Operational intelligence

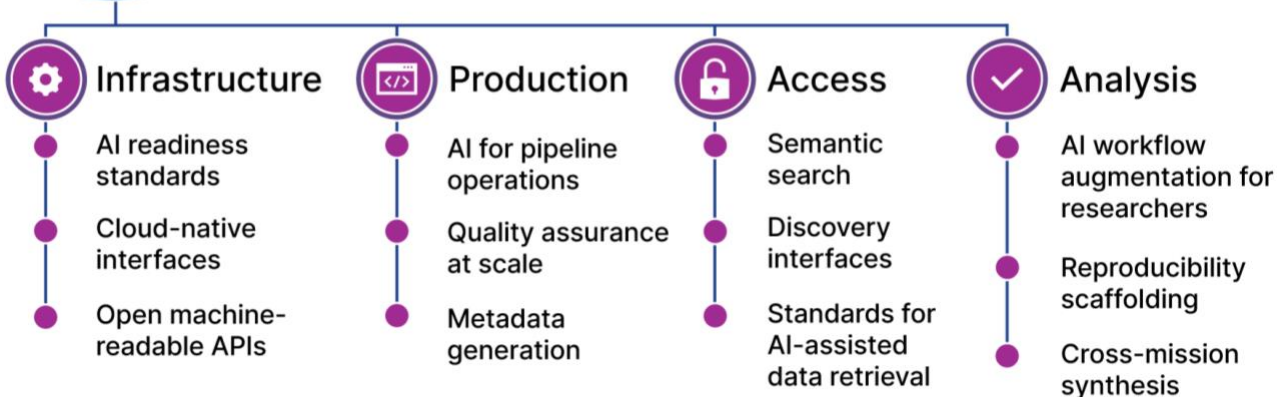
AI-assisted monitoring, anomaly detection, and performance metrics analysis are applied continuously across the full data lifecycle. From pipeline health to infrastructure performance to access patterns, operational intelligence is how we know the system is working and where it needs attention.

It is also how we understand our community. Synthesizing user needs across feedback channels, usage patterns, help requests, and direct engagement is part of operational intelligence. This includes emerging AI-assisted approaches to aggregating and interpreting that signal at scale. The goal is a coherent picture of how Earth science data is being used and where the gaps are, so that picture can actively inform priorities across the other pillars.



EARTHDATA Intelligence Strategy

AI-powered infrastructure and workflows for Earth observation data



Cross-Cutting Capabilities

Foundational capabilities that span all four pillars



Governance & Standards

The policies, review process, metadata requirements, and interoperability standards that ensure consistency and trust across everything we build.



Operational Intelligence

AI-assisted monitoring, anomaly detection, and performance metrics analysis applied continuously across the full data lifecycle. From pipeline health to infrastructure performance to access patterns, operational intelligence is how we know the system is working and where it needs attention.

Boundaries: What We Do Not Do

A strategy is also a set of choices about what not to pursue. These boundaries exist to protect our focus, manage expectations, and ensure we are accountable for the right things.

- We do not validate or endorse domain-specific AI outputs. GeoAI models, climate projections, and other domain applications that use our data are not ours to evaluate scientifically. We enable them through infrastructure; we do not certify them. We are happy to partner in such cases if it aligns with our goals.
- We do not default to building new portals, websites, or user-facing interfaces. Our AI capabilities are designed to reach users through the tools they already use.
- We do not build AI capabilities for specific science domains. Where there is overlap with our mission, we aim to engage early and technically. We expect to contribute and to learn.
- We do not pursue theoretical AI research. Our work is grounded in funded missions and operational delivery. The publishable insights that result from our AI work are a byproduct, not the primary goal.
- We do not deploy AI capabilities without clear provenance, transparency, and accountability. AI-assisted outputs in our systems must be traceable. We follow applicable NASA and U.S. government ethical AI guidance and expect our partners and funded teams to do the same.
- We default to open access and open source. AI capabilities, algorithms, and workflows developed with ESDS funding are expected to be openly available consistent with NASA open science policy. Funded agreements will specify IP terms explicitly as needed.

How We Execute: Five Tactics

To successfully execute the strategy, we need some concrete tactics. Those listed below are how we move from principles and pillars to real work. They are complementary, not sequential, and we pursue them in parallel.

TACTIC 01 AI Innovation Call Q3 FY26

Issue an open call inviting funded members of the ESDS ecosystem to propose AI-powered solutions to real problems in Earth science data systems. Enhancements to existing projects are encouraged and are our preferred starting point. If a team is already doing something and AI can make it faster, smarter, or more useful, that is exactly the kind of proposal we want to see.

Eligibility extends beyond core data systems work to any funded, directed effort within the Earth Science Division where there is a clear connection to data access, production, or use.

The call runs in two phases: a short, written proposal, then funding for selected teams to build a working prototype or integration. Awards are expected to be in the range of \$150K to \$200K. We offer use-case examples as inspiration, not prescription. The best ideas will come from people closest to the work.

Detailed guidance on specific timelines, proposal requirements, evaluation criteria, and prototype-to-operational transition expectations will be provided in the forthcoming call documentation.

TACTIC 02 Embedded collaboration with mission teams IN DEVELOPMENT

ESDS already assigns data systems leads directly to missions at the point where data products are being defined, designed, and produced. In the AI context, this means ensuring ESDS works alongside mission teams to evaluate AI-powered pipeline segments, develop AI-ready data products, and build the metadata and provenance scaffolding that makes mission data useful at archive scale. We infuse funds as appropriate to advance desired outcomes.

Doing this well requires dedicated support and focus. How that capacity gets resourced within the program is an active question we will address as this strategy matures.

TACTIC 03 Governance and standards setting IN DEVELOPMENT

We already work within our community to establish and maintain standards. We need to push further into defining what AI-ready data looks like, what provenance requirements apply to AI-generated outputs, and how AI tools interact with NASA's archive. These standards apply internally to ESDS programs and externally to missions and community partners who want their data in the ecosystem. It is the minimum set of rules that allows the community to trust what they get from us, and what we get from them.

TACTIC 04 Partnerships: commercial and interagency IN DEVELOPMENT

No single organization will build the AI-powered Earth science data ecosystem alone. We pursue deliberate partnerships with commercial AI providers, cloud platforms, other federal data agencies, and international partners such as the European Space Agency. These partnerships extend our reach, reduce duplication, and position NASA data as a first-class participant in the broader AI ecosystem.

A critical emerging challenge in this space is data rights. As NASA increasingly acquires data from commercial satellite operators rather than operating its own fleets, licensing restrictions on commercially acquired data complicate open access and AI ecosystem participation. We engage this challenge proactively at the acquisition stage, not after the fact.

TACTIC 05 Training and workforce development IN DEVELOPMENT

AI tools are only as good as the people using them. We invest in building AI literacy across the ESDS workforce and the broader Earth science data community. Not deep technical specialization for everyone, but enough shared understanding that teams can evaluate AI capabilities critically, identify where they apply, and recognize where they fail.

This includes developing internal training pathways for ESDS staff, contributing to community training through user support and partner programs, and ensuring that the embedded collaboration model transfers knowledge rather than creating dependency on ESDS expertise.

How We Know If It Is Working

We still need to develop a full measurement framework for this strategy, but some leading indicators worth monitoring across the pillars include:

- Reduction in the manual burden on production pipelines and operational maintenance in user facing systems, measured as staff hours;
- Search session success rate: the proportion of discovery sessions that result in a data download or access event;
- Researcher time-to-first-use: time from project start to first meaningful data access, as reported through community surveys;
- Number of AI-powered prototypes moved into operational status, tracked through the Innovation Call and embedded collaboration program;
- Compliance rate of new mission data products against AI-readiness standards at ingest.

These are internal key performance indicators (KPIs) used by the program to learn, course correct and demonstrate responsible stewardship of resources. They are not performance targets for individuals or teams. The real measure of success is whether our community is doing more science with NASA data faster than they were before and whether it is in alignment with Gold Standard Science.

What This May Look Like in Practice

The strategy pillars above align ESDS work. The use cases below illustrate the range of problems AI can help us solve in concrete terms. This is not intended as a complete list, as the Innovation Call is explicitly designed to surface more.

Production

- AI-powered pipeline segments for computationally expensive processing steps, with ESDS owning reliability and the mission science team owning scientific validation.
- Mission data and metadata development: AI-assisted compliance, richness, and AI-readiness built into a mission's data product lifecycle from the start.

Infrastructure

- Open, machine-readable interfaces that position NASA data as a native participant in external AI ecosystems: the archive is reachable directly from the AI tools researchers already use.
- Intelligent user support triage so staff spend time on genuinely novel problems and science enablement rather than repeated common questions.

Access

- Dataset matchmaking delivered through external AI assistants: the right combination of primary, ancillary, and model data surfaced when a researcher describes their science question. No NASA portal is required.
- Semantic discovery capability exposed as a service, enabling AI agents to understand science intent and retrieve relevant datasets on a researcher's behalf.

Analysis

- Earth science notebook extension embedded in JupyterLab, with knowledge of NASA data formats and common scientific Python tools.
- Reproducibility scaffolding that auto-generates provenance records and methods section drafts from analytical sessions.

Operational Intelligence

- Collection health monitoring: a continuous AI watch over archive holdings for anomalies, gaps, and quality signals before users encounter them.
- Operational performance metrics: AI-assisted analysis of system behavior trends to inform infrastructure and pipeline investment decisions.

Artificial Intelligence in Earth Science Data Systems



EARTHDATA