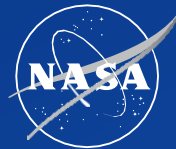




OCSDO

OFFICE OF THE CHIEF
SCIENCE DATA OFFICER



NASA Core Data and Computing Services for Science

SMD Divisional Data Workshop

May 7, 2025

GSFC



SMD by the Numbers

 **15 M +**
FY24 users

 **140 PB**
FY23 Total volume

 **50 PB/year**
Ingest Rate

 **>700 PB**
Projected FY29
archive volume

 **1.5 B**
Total files

 Efficient data management and computing
are essential for NASA's mission.

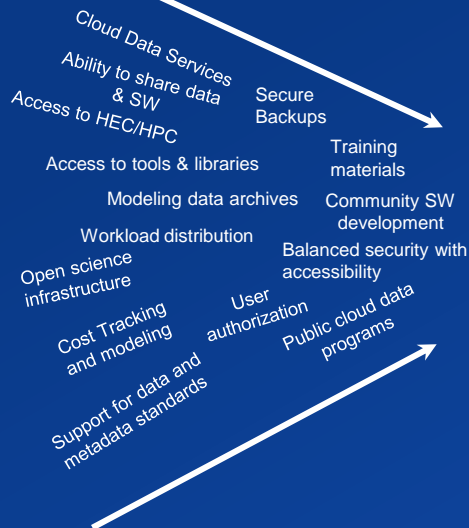
Data and Computing Architecture Study Report

17
Workshops

70 RFI
Responses

5 Technical
Interchange
Meetings

Past Studies



Synthesis of input within Science Data Lifecycle

Find Data

- Metadata catalog
- Image server
- Data search client

Access Data

- Data Services
- APIs
- Open access

Analysis

- Tools (access)
- Data services
- Questions

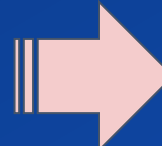
Publication

- Mission data production system
- Std. Product Generation

Research

- Documentation

284 Inputs



User Interface
Services

Analysis Services

Data Services

IT Support
Services

Cloud and HEC
Infrastructure
Services

Reference Architecture

Data Science Innovation

Search and Discovery Tools



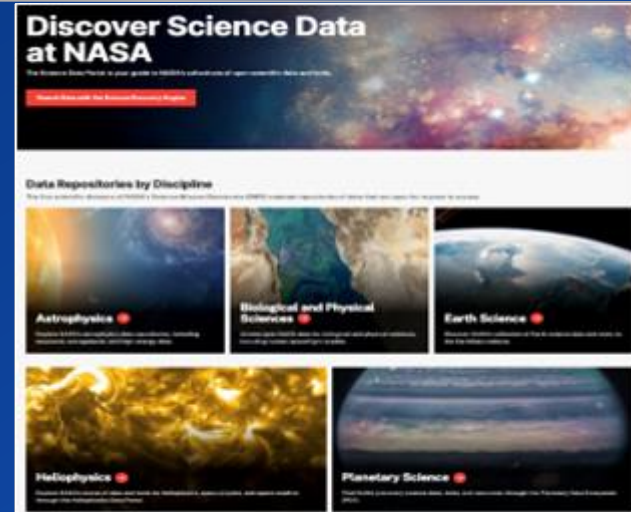
SciX

Literature-based digital library designed to revolutionize how science resources are found, accessed and utilized. It's an expansion of the Astrophysics Data System.



Science Discovery Engine (SDE)

The Science Discovery Engine is a search platform that enables discovery of over 95% of NASA's open science data and documentation across disciplinary areas.



Science Data Portal

The Science Data Portal provides a list of freely accessible science data. These data are used by scientists, researchers, and the public to explore valuable information about our universe and beyond.

User Interface Services

User interface services for person and machine-to-machine access and use of analysis services

Interactive Analytics and Visualization

Data Visualization and Collaboration

Analysis Services

Analysis services that are accessed by the user interface services to support Open-Source Science

Data Analytics and Data Science Platforms

AI/ML Tools and Frameworks

MLOps Platforms

Ontology Service

Data Services

Data access and management services that support the analysis services

Data Service Management

Data Protection

Data Monitoring and Tracking

Data Virtualization and Federation

Enterprise Data Repositories

Data Visualization and Business Intelligence

Data Discovery

Data Acquisition and Conditioning

Data Access and Delivery

IT Support Services

IT Support Services for developing and running the user interface, analysis, data, and IT support services

Application and Resource Monitoring

Software Factory

Edge Service Routing and Access Management

Identity and Access Management

Workflow Management

Cloud and HEC Infrastructure Services

Infrastructure services (e.g. computing, networking, communication, etc.) to provision the user interface, analysis, data, and IT support services

Computing (including High-End and Cloud)

Person/Machine Entity Client Software

Container Orchestration

Storage Management

Cluster Management

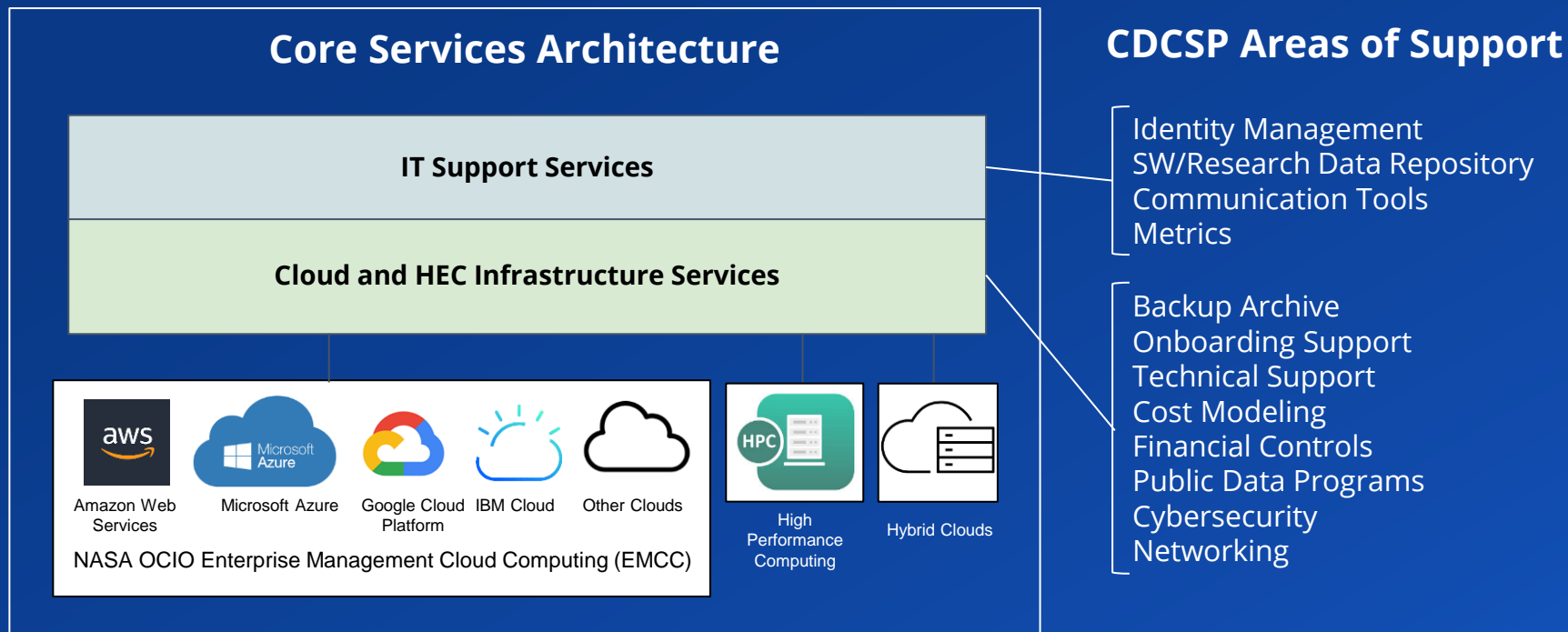
Networking and Connectivity

Load Balancing

Cybersecurity

Core Data Computing and Services Program (CDCSP)

The Core Data and Computing Services Program (CDCSP) will provides **IT Support Services** and **Cloud and HEC Infrastructure Services** to manage accounts



HECC Hardware Assets and Related Services

5 Compute Clusters

- Aitken 13.1 PF
- Electra 8.32 PF
- Cabeus 1,106 GPUs
- Endeavour 155 TF
- Next Gen 20.84 PF

1 Visualization Cluster

13 Lustre File Systems 118.1 PB

3 NFS File Systems 3.3 PB

2 BeeGFS File Systems 1.5 PB

1 VAST File System 2.5 PB

Archive System 1,572 PB

NAS Facility Extension

- One-acre 30 MW site to house HPC systems in modules.
- Average PUE 1.05

