NASA
Agency Report

Andrew Mitchell

WGISS-53

Virtual

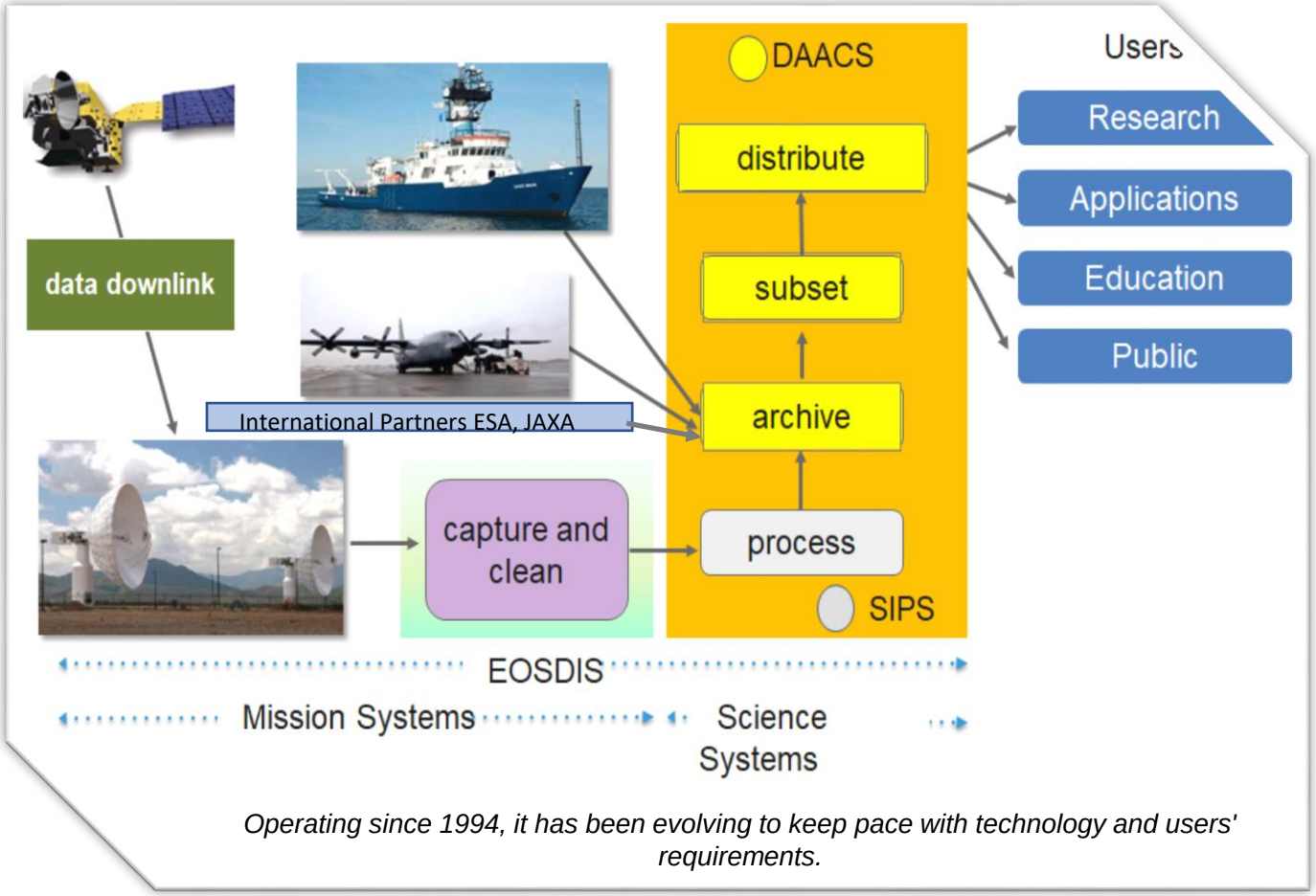
24 March 2022

NASA Agency Report Executive Summary

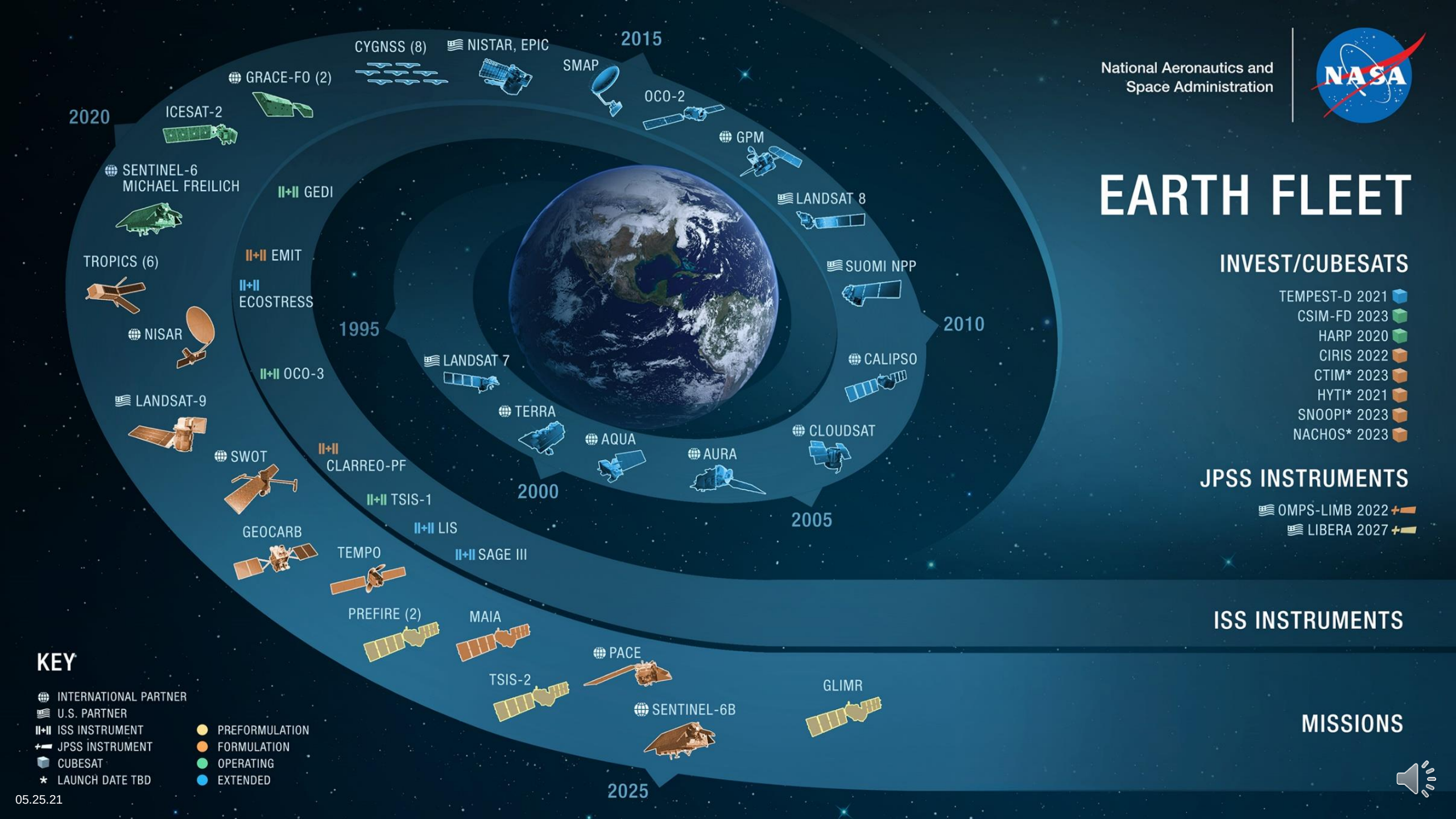
- **Delivered nearly billion products to over 4 million users in FY20 ((Oct. 1, 2020 to Sept. 30, 2021) to every country in the world representing an increase of 100 million products from FY20. 1,502 new and reprocessed datasets were added to the EOSDIS collection in FY21.**
- **NASA Earth Observing System Data and Information System (EOSDIS) data have been distributed to 195 independent countries**
- **Provided data stewardship to almost 13,000 unique data sets for which we have minted over 9700 digital object identifiers that enable users to track data sets through publications and documentation.**
- **By the end of FY21, had archived over 59 Petabytes of Earth Science data at a rate of 53 terabytes per day. This was a 34% increase from FY20. Over 15 petabytes are available in Amazon Web Services (AWS).**
- **Scored 81 on the American Customer Satisfaction Index (ACSI) survey an increase of two points from last year's score of 79. This continues the trend of high scores for EOSDIS performance, despite the pandemic, a shift to working from home, and changes in our NASA missions and the research community schedules and plans.**
- **Undergoing an Open Sourced Science for Earth System Observatory (ESO) Mission Science Data Processing Study with the goal of: Identify and assess potential architectures that meet the ESO mission science data processing objectives, promote open science principles, enable data system efficiencies, and support earth system science and applications.**

Earth Science Data and Information System (ESDIS)

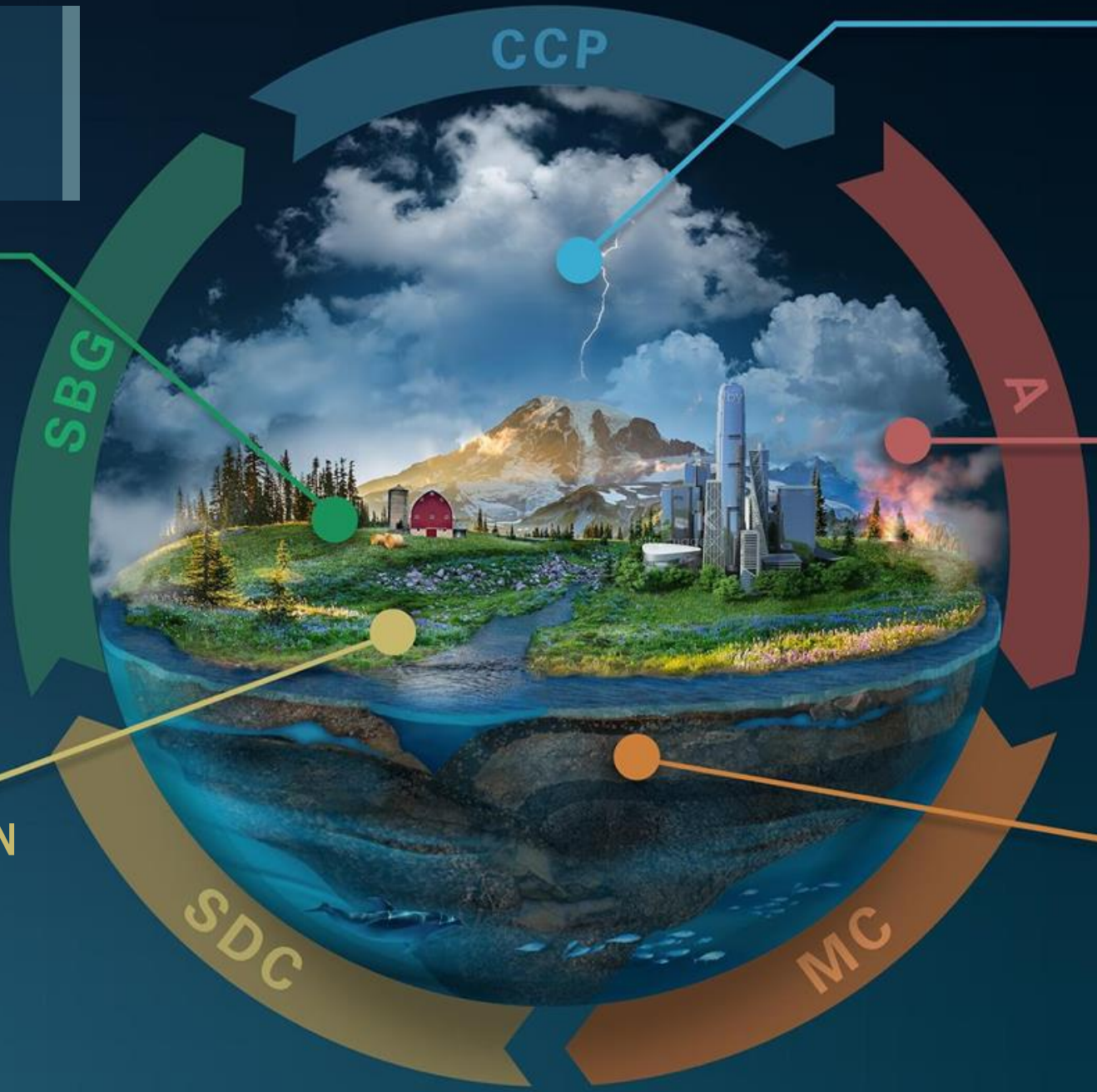
- The ESDIS Project manages the science systems of the Earth Observing System Data and Information System (EOSDIS).
- EOSDIS is a comprehensive distributed Earth science data and information system designed to support NASA's Earth science missions.
- EOSDIS is designed to ingest, archive, distribute, visualize, all types of Earth Science data which include:
  - *field campaign measurements, airborne data, in situ data, model data, ancillary products used for processing and other related datasets.*
- The ESDIS Project provides and controls all aspects of the effort including but not limited to requirements, design, acquisition, development, operations, maintenance and decommission.



Operating since 1994, it has been evolving to keep pace with technology and users' requirements.


National Aeronautics and Space Administration
NASA
EARTH FLEET
INVEST/CUBESATS
TEMPEST-D 2021
CSIM-FD 2023
HARP 2020
CIRIS 2022
CTIM* 2023
HYTI* 2021
SNOOPI* 2023
NACHOS* 2023
JPSS INSTRUMENTS
OMPS-LIMB 2022
LIBERA 2027
ISS INSTRUMENTS
MISSIONS
2015
2020
1995
2010
2000
2005
2025
CYGNSS (8)
NISTAR, EPIC
SMAP
OCO-2
GPM
LANDSAT 8
SUOMI NPP
CALIPSO
CLOUDSAT
AURA
AQUA
TERRA
LANDSAT 7
GRACE-FO (2)
ICESAT-2
SENTINEL-6 MICHAEL FREILICH
GEDI
EMIT
ECOSTRESS
TROPICS (6)
NISAR
OCO-3
LANDSAT-9
SWOT
CLARREO-PF
TSIS-1
LIS
SAGE III
GEOCARB
TEMPO
PREFIRE (2)
MAIA
PACE
TSIS-2
SENTINEL-6B
GLIMR
KEY
INTERNATIONAL PARTNER
U.S. PARTNER
ISS INSTRUMENT
JPSS INSTRUMENT
CUBESAT
LAUNCH DATE TBD
PREFORMULATION
FORMULATION
OPERATING
EXTENDED
05.25.21

EARTH SYSTEM OBSERVATORY



CLOUDS, CONVECTION AND PRECIPITATION

Water and Energy in the Atmosphere

AEROSOLS

Particles in the Atmosphere

MASS CHANGE

Large-scale Mass Redistribution

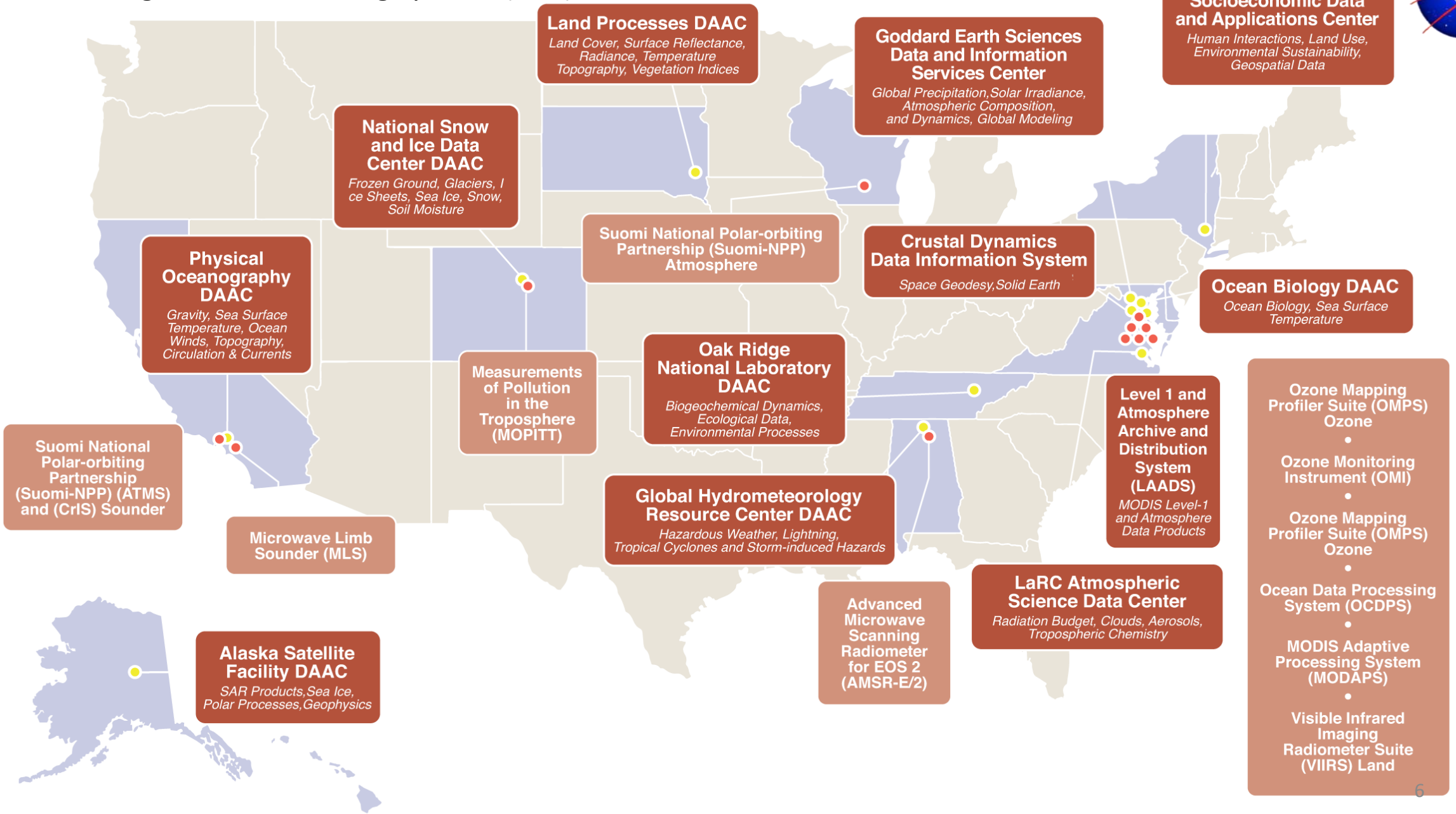
SURFACE BIOLOGY AND GEOLOGY

Earth Surface & Ecosystems

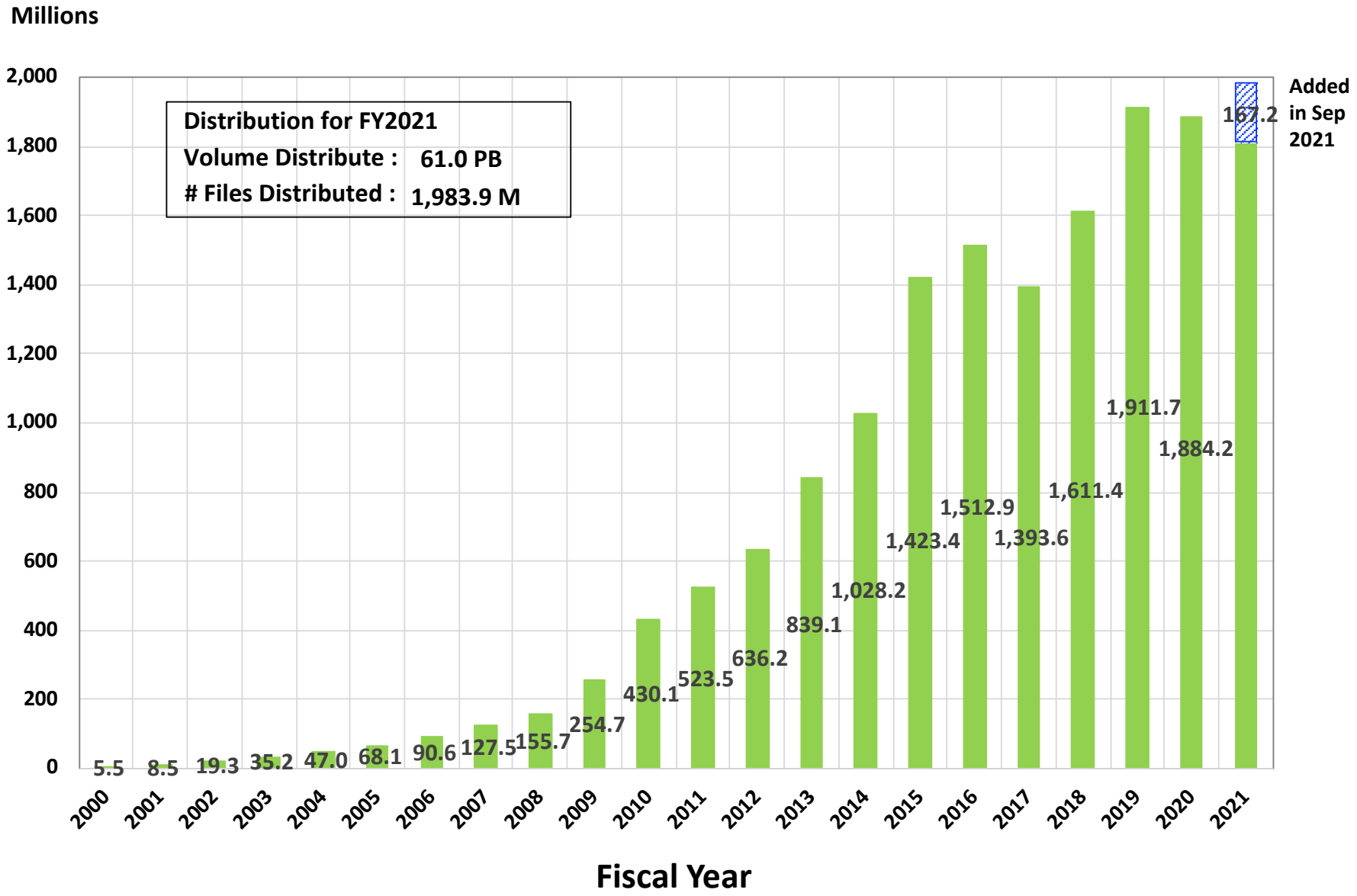
SURFACE DEFORMATION AND CHANGE

Earth Surface Dynamics

Discipline Specific, Distributed, Active Archive Centers (DAACs)
Science Investigator-led Processing Systems (SIPS)



EOSDIS Products Delivered: FY00 thru September 2021



Motivation for Commercial Cloud

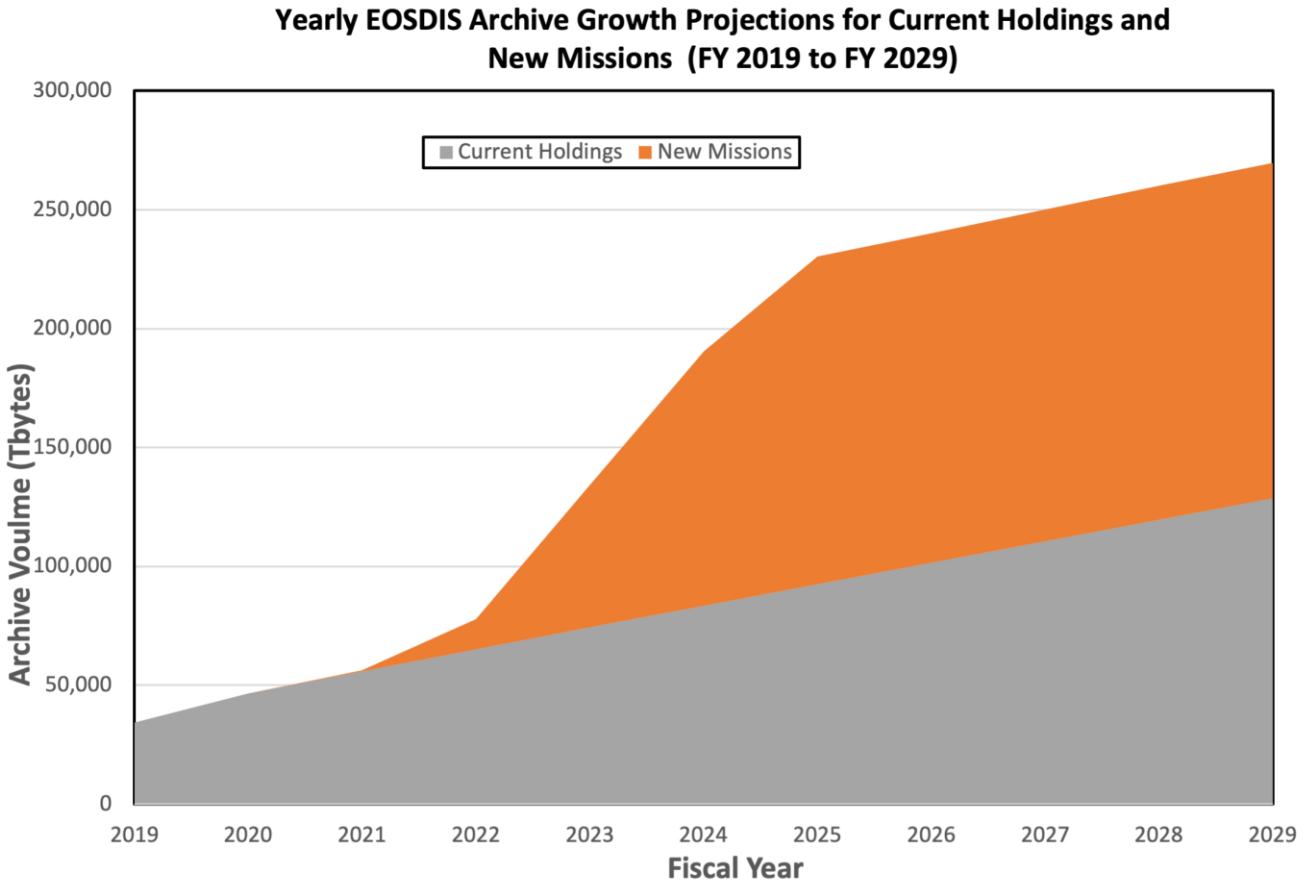
Motivation for Cloud

Growth of Mission Data & Processing: Projected rapid archive growth and the need to effectively process significantly larger volumes of new mission data requires **rethinking existing architectures**.

Data Systems: More cost-effective, flexible, and scalable data system ingest, archive, and distribution solutions are needed to **keep pace with new mission advancement**.

Science Users: Significantly larger data volumes requires **additional ways to access and utilize this data**, with “Data Close to Compute” or Data Lake”. **Bring Algorithms to the cloud.**

Projected Data Volumes

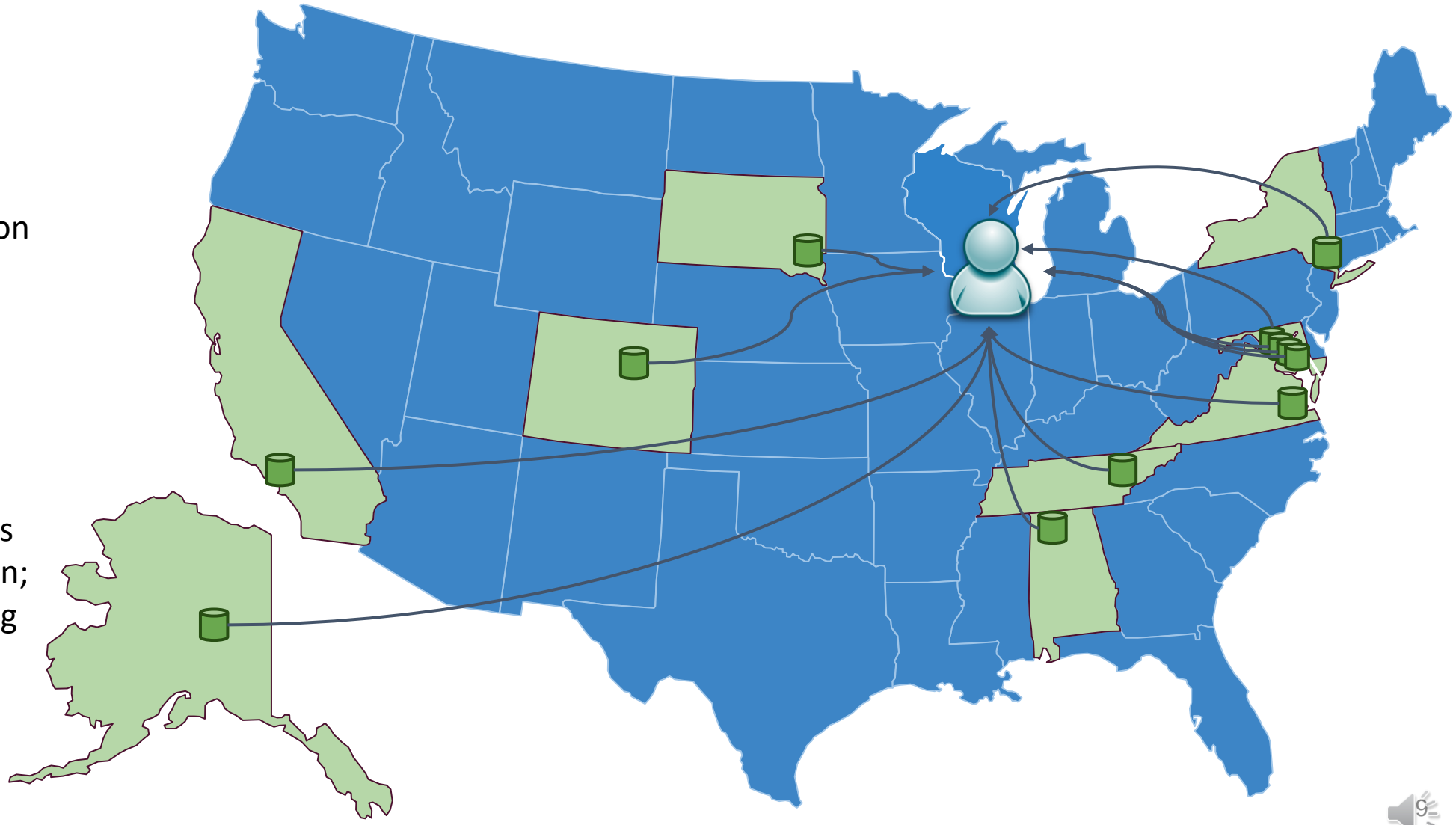


Current User Interaction

User must download data from centers across the US

What's Good: decentralization means that systems are independently available; minimize the decentralized view by common functions (Earthdata Search)

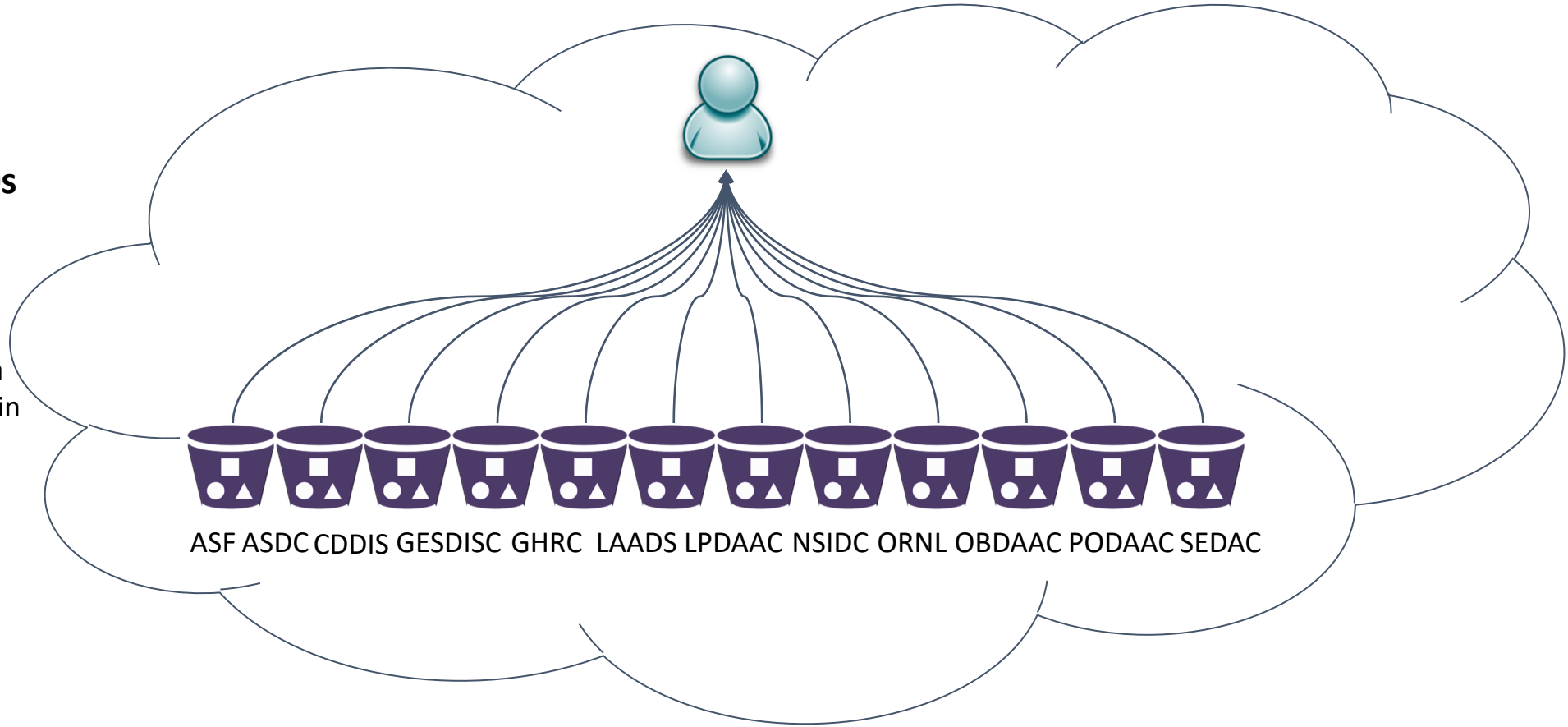
What's Not So Good: user has to get data from each location; user must invest in processing capacity; having many interfaces is confusing and complicated



Envisioning an EOSDIS "Data Lake"

Evolution of EOSDIS in the 2020s

- Move data into commercial cloud where all is accessible
- Make data versions in the Cloud analysis ready
- Provide common services on the data
- Enable easily understood access/use in the Cloud



User now goes to one location to get data

- **still has ability to download data,** but will be able to use even more services that will be available in a common environment

The EOSDIS Data Lake is part of a suite of EOSDIS Cloud capabilities:

- Core cloud compliance, security, reporting, cost control, and metrics capabilities
- EOSDIS application and service hosting capabilities
- EOSDIS data hosting capabilities

Our Elevator Pitch

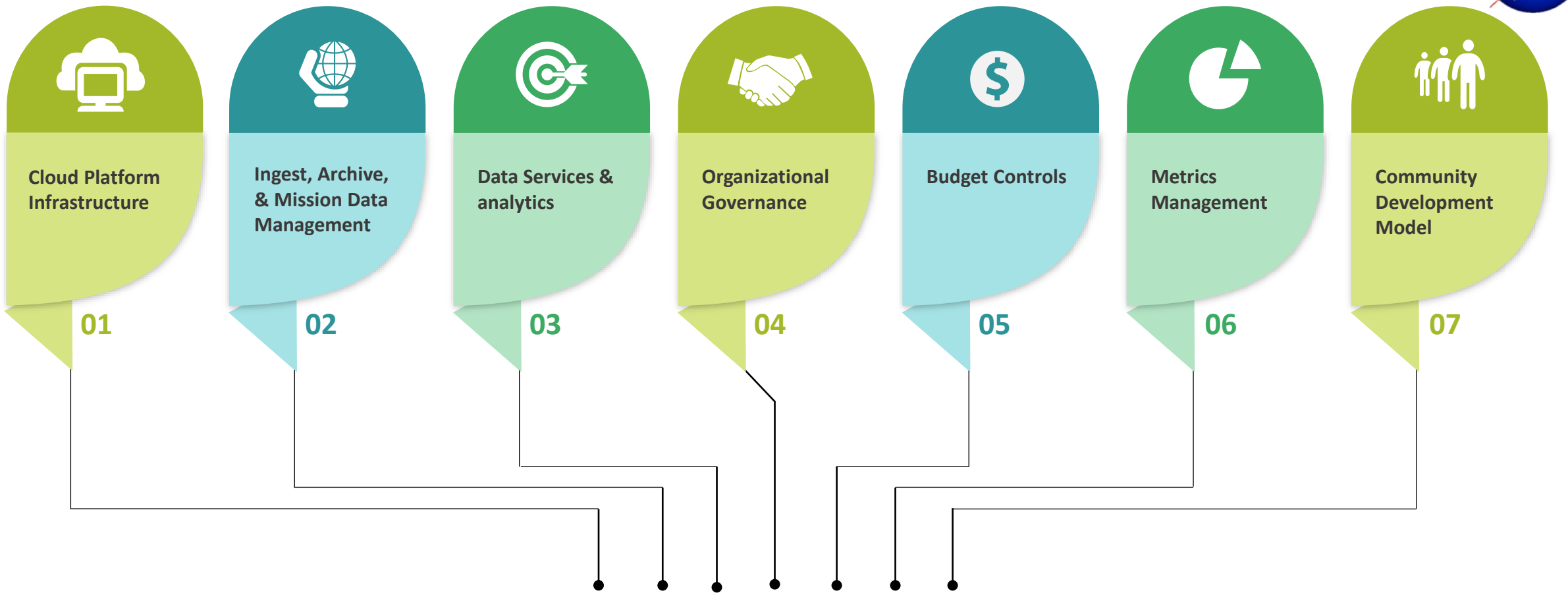
By hosting NASA's Earth science data in the cloud, EOSDIS is able to realize several end user benefits:

- **Power**: Any user can access big processing power "next to" Big Data.
- **Performance**: Data can be offered in a form enabling high-performance analysis.
- **Freedom from Data Transfers**: Users need not move Big Data.
- **Freedom from Data Management**: Users need not store and manage Big Data.
- **Data Co-location**: Users can easily work with multiple EOSDIS datasets together.
- **Choice**: Users can still download data if they prefer.

Free image from https://www.dreamstime.com/photos-images/elevator-pitch.html

Earthdata Cloud -EDC-

- Improve the efficiency of NASA's data systems operations – continues free and open access to data
- Prepare for planned high-data-rate missions
- Increase opportunity for researchers and commercial users to access/process PBs of data quickly without the need for data management
- Transparent/extendable open source processing framework



Components of the Earthdata Cloud

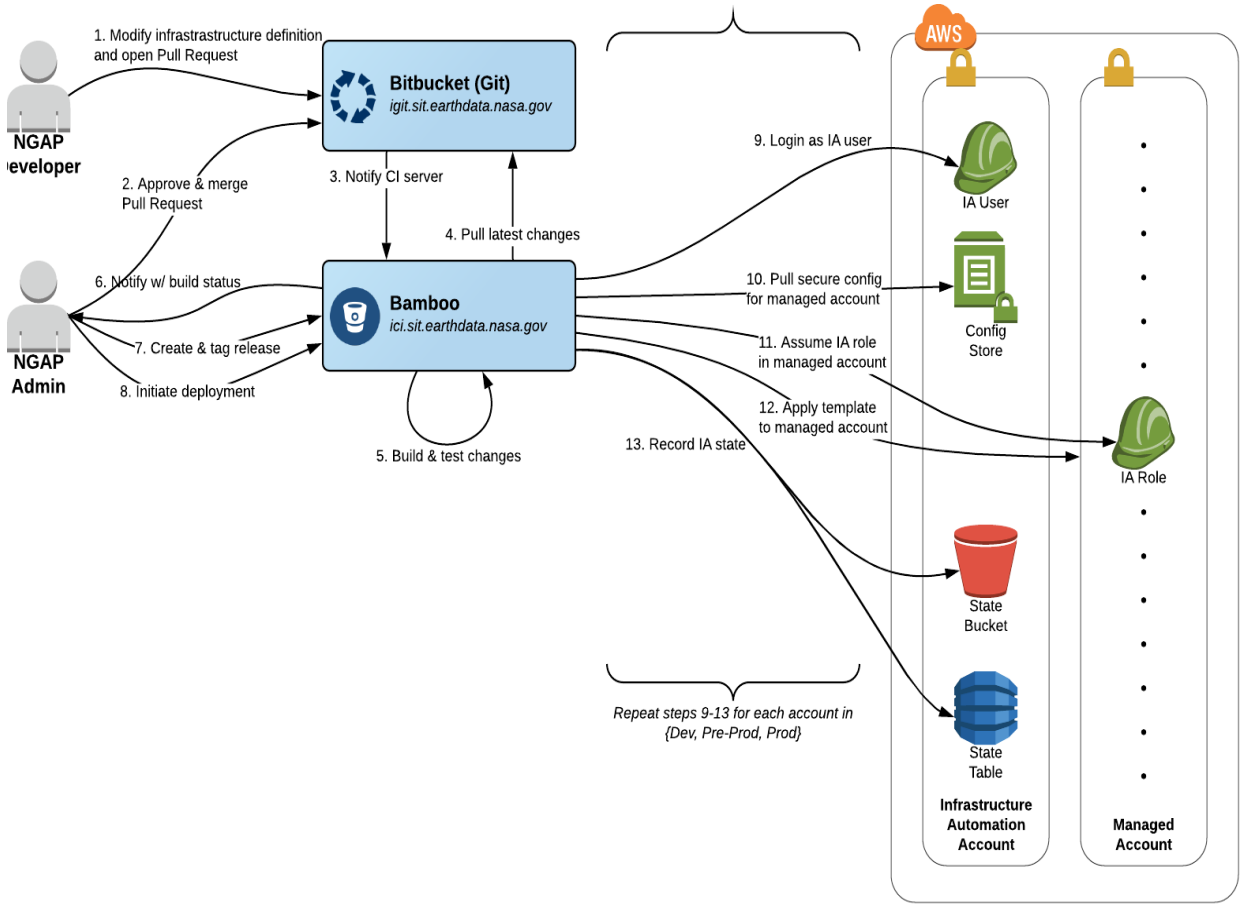
Cloud Platform Infrastructure

Common Services & Controls

What is the Earthdata Cloud Platform?

Earthdata Cloud Platform (NGAP - NASA General Application Platform) is a multi-account, Infrastructure-as-a-Service (IaaS) cloud platform operating on Amazon Web Services (AWS), providing shared cloud services and controls to EOSDIS.

1. **NASA-Approved Amazon Web Services (AWS):** vetted AWS and third-party Software-As-A-Service (SAAS) services and process to add new. Focus is on using AWS cloud-native services
2. **Code Deployment Services:** DevOps Continuous Integration Continuous Delivery (CICD) Pipeline to security scan, build, and deploy code
3. **Use of Infrastructure as Code:** including re-useable template to define a multi-account ecosystem
4. **Single System Security Plan (SSP) and Authority to Operate (ATO)**
5. **Single Identity and Access Management Solution (CloudTamer.io):**
   - Rotate AWS access keys
   - Apply session limits
   - Provide role-based access control
   - two-factor authentication



Unifying Ingest & Archive in the Cloud: Cumulus

What is Cumulus?

Custom built, open source, lightweight, cloud-native framework for data ingest, archive, distribution and management

A lightweight framework consisting of:

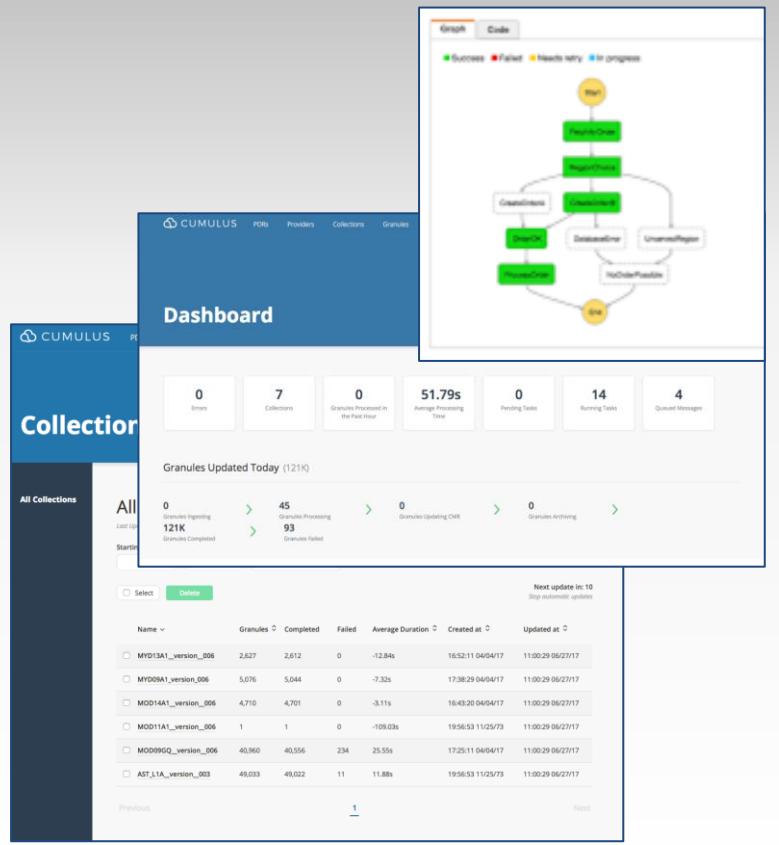
Tasks a discrete action in a workflow, invoked as a Lambda function or EC2 service, common protocol supports chaining

Orchestration engine (AWS Step Functions) that controls invocation of tasks in a workflow

Database store status, logs, and other system state information

Workflows(s) file(s) that define the ingest, processing, publication, and archive operations

Dashboard create and execute workflows, monitor system



Unifying Data Services in the Cloud: Harmony

Historically, EOSDIS DAACs have all provided their own tooling with diverse interaction patterns and APIs. Harmony is our ongoing effort to revisit these siloed capabilities in a more harmonized manner.

End-Users

Service Execution Framework(s)

Enterprise Integration: Login, Metrics, Egress, Metadata Catalog

Common Interface: Common API, Earthdata Search UI

Data staging, Transformation code execution

DAAC or Core-Team-Supplied

DAAC-Unique Transformations

e.g., SWOT water feature averaging

DAAC-Supplied

Common Transformations

e.g., Subsetting L3 NetCDF

DAAC or Core-Team-Supplied

EOSDIS Conceptual Cloud Based Architecture

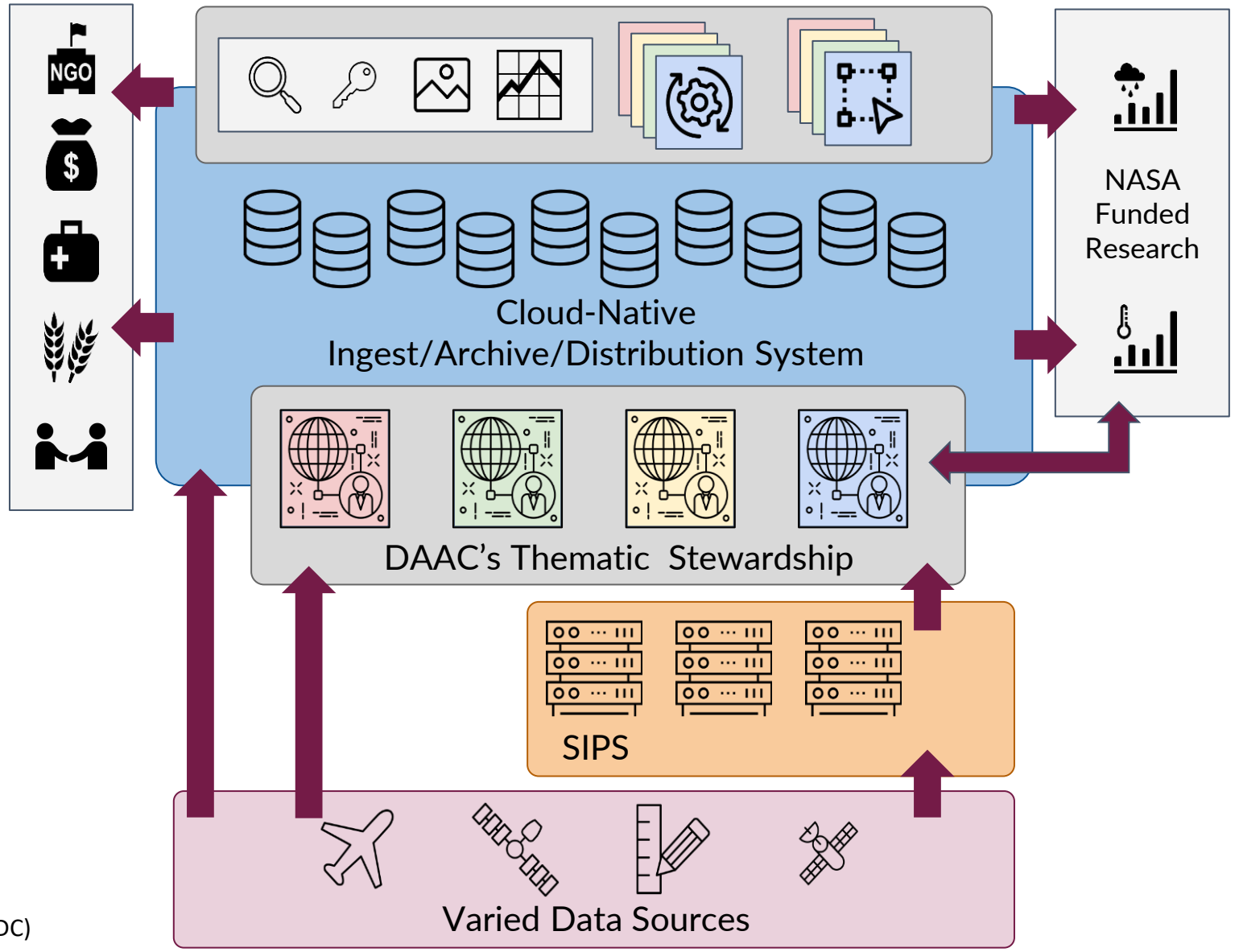
Benefits

Collocated, pay-as you-go processing for *anyone*

Expert user support

Streamlined product addition

Reduced duplication of tools and services



Challenges

Development coordination

Cost Management

Shifting Labor Needs

Security/Export Compliance

Vendor Lock In

Open Sourced Science for Earth System Observatory (ESO) Mission Science Data Processing Study

STUDY GOAL:

Identify and assess potential architectures that meet the ESO mission science data processing objectives,

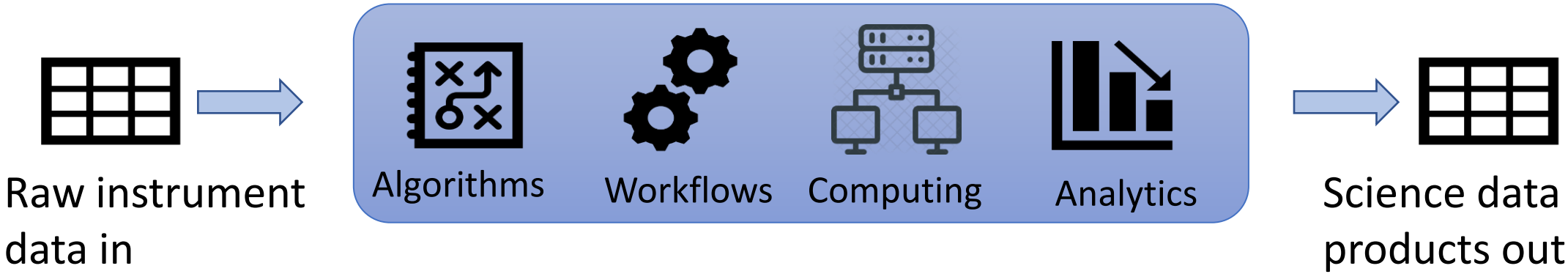
- promote open science principles,
- enable data system efficiencies,
- support earth system science and applications.

Aligns with the challenge set by NASA to create a single observatory that combines data from the ESO missions to understand the earth as a system and accelerate our ability to apply this understanding.

Definition of a Mission Science Processing System

The set of algorithms, software, compute infrastructure, operational procedures, and documentation to automatically process raw instrument data through to science quality data products.

This includes the software tools that support the development of the processing algorithms and validation and analysis of the processed data.



Study Approach

Workshop #1: October 19-20, 2021
Receive input from NASA Program Offices and ESO Missions on requirements, constraints, recommendations, and opportunities for science data processing. Report available on Study Website.

Workshop #2: March 1-4, 2022
Understand the current state of mission and science data processing and obtain community input.

Architecture Study: April – July, 2022
Analyze the architectural options and identify an optimal solution for the ESO missions.

Workshop #3: August, 2022
Report out on the study and make a recommendation to Kevin Murphy, SMD Chief Scientific Data Officer.

Workshop #2 Goals

- Understand the current state of mission science data processing:
  - NASA Earth missions (flagship and PI-led missions)
  - NASA Astrophysics missions
  - Other federal agency and international missions.
- Understand opportunities for collaboration with systems that interface with and enable mission processing systems.
- Learn from other big-data processing systems
  - Based on input received through a “Requestion For Information” solicitation
- Gain community insights and recommendation

Deliverable: Report of findings

Additional information is available at:

http://earthdata.nasa.gov/