The Evolution of NASA's Earth Observing System Data and Information System (EOSDIS)

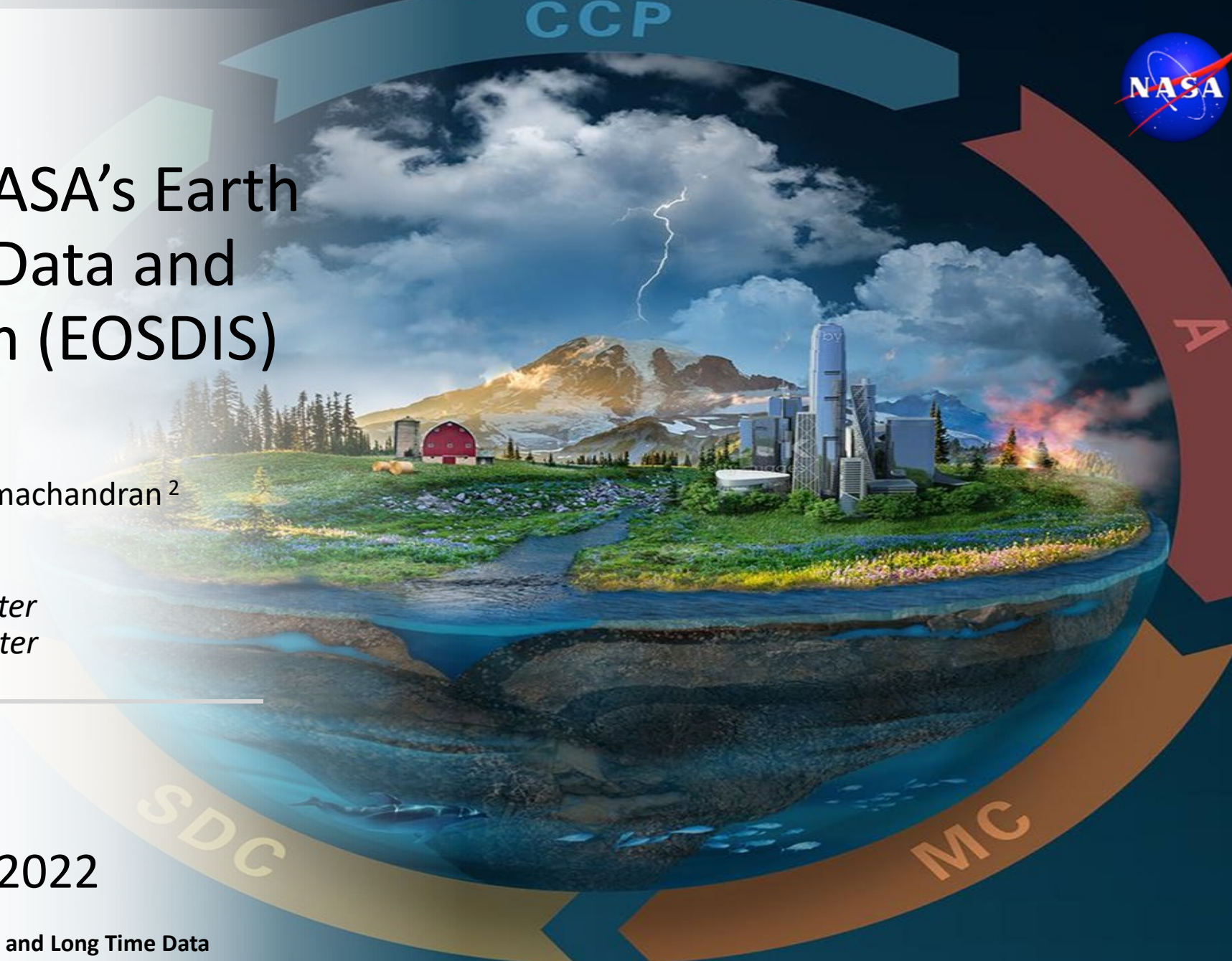**Authors:** Andrew Mitchell[1], Rahul Ramachandran [2] and Manil Maskey [2]
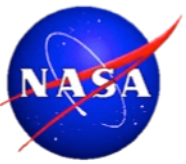
[1]NASA Goddard Space Flight Center
[2]NASA Marshall Space Flight Center
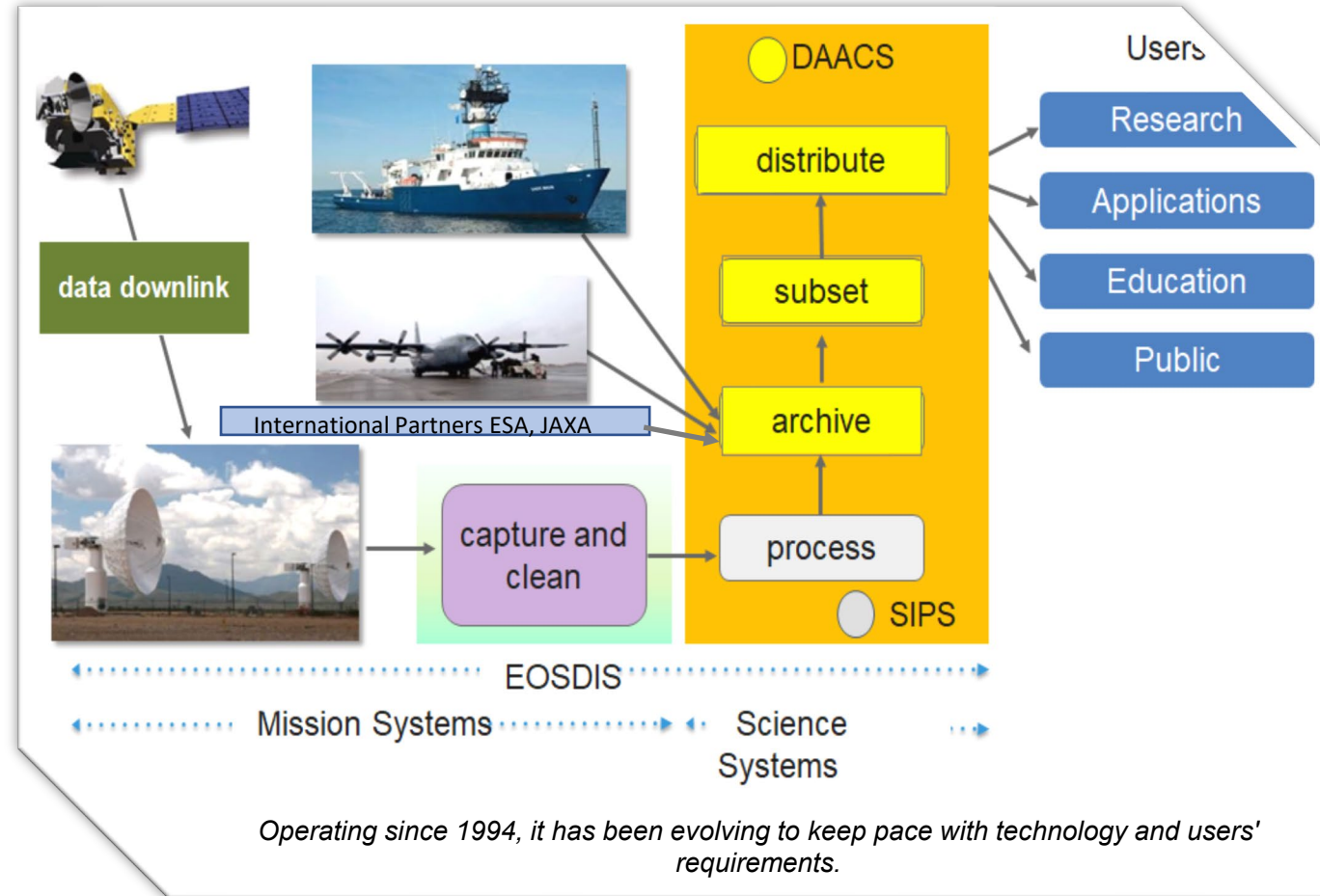
Tuesday 24.05.2022

Living Planet Symposium 2022

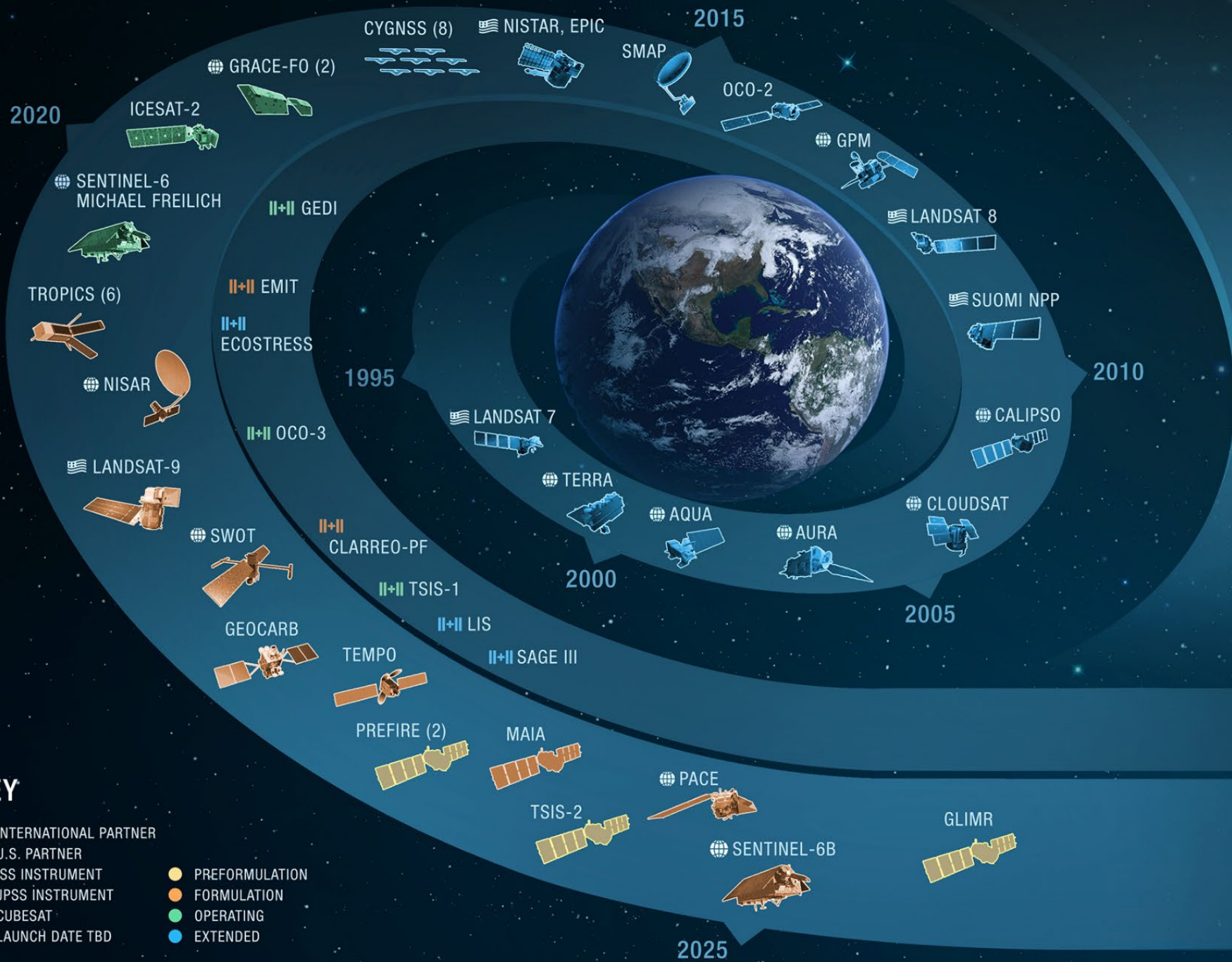**Session: B4.01** Heritage missions - **Heritage Missions and Long Time Data Series**

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.

2

National Aeronautics and
Space Administration

NASA

EARTH FLEET

2015

CYGNSS (8)

🌐 GRACE-FO (2)

NISTAR, EPIC

SMAP

2020

ICESAT-2

OCO-2

🌐 GPM

🌐 SENTINEL-6
MICHAEL FREILICH

II+II GEDI

🏴 LANDSAT 8

INVEST/CUBESATS

TEMPEST-D 2021	🟦
CSIM-FD 2023	🟩
HARP 2020	🟩
CIRIS 2022	🟧
CTIM* 2023	🟧
HYTI* 2021	🟧
SNOOPI* 2023	🟧
NACHOS* 2023	🟧

TROPICS (6)

II+II EMIT

II+II
ECOSTRESS

1995

🏴 SUOMI NPP

🌐 NISAR

II+II OCO-3

🏴 LANDSAT 7

2010

🌐 CALIPSO

🏴 LANDSAT-9

🌐 TERRA

🌐 AQUA

🌐 AURA

🌐 CLOUDSAT

JPSS INSTRUMENTS

🏴 OMPS-LIMB 2022	+▬
🏴 LIBERA 2027	+▬

🌐 SWOT

II+II
CLARREO-PF

II+II TSIS-1

2000

2005

GEOCARB

II+II LIS

TEMPO

II+II SAGE III

ISS INSTRUMENTS

PREFIRE (2)

MAIA

🌐 PACE

GLIMR

KEY

🌐 INTERNATIONAL PARTNER
🏴 U.S. PARTNER
II+II ISS INSTRUMENT
+▬ JPSS INSTRUMENT
⬜ CUBESAT
* LAUNCH DATE TBD

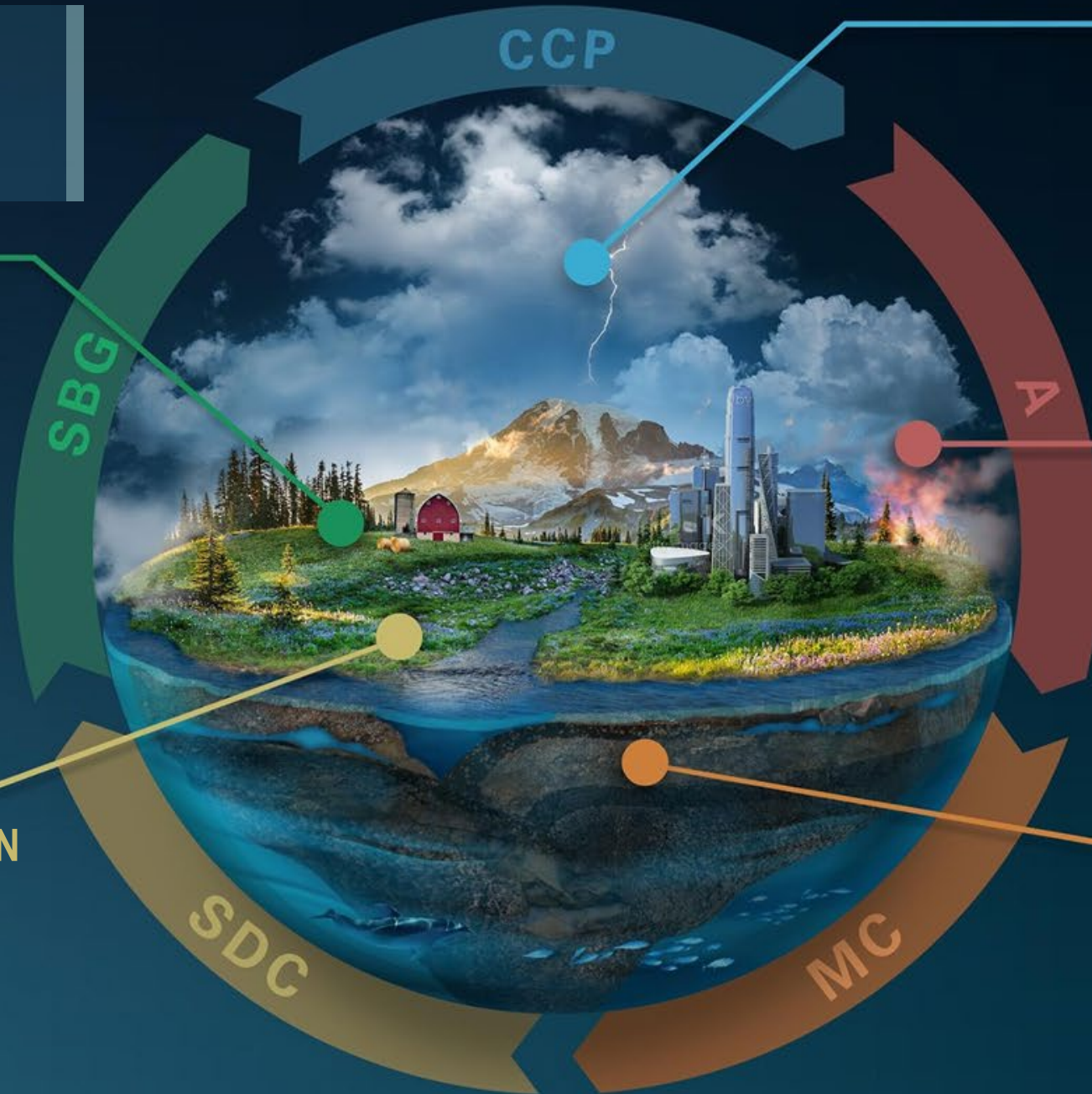🟡 PREFORMULATION
🟠 FORMULATION
🟢 OPERATING
🔵 EXTENDED

TSIS-2

🌐 SENTINEL-6B

MISSIONS

2025

05.25.21

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

NASA

Land Processes DAAC
Land Cover, Surface Reflectance, Radiance, Temperature Topography, Vegetation Indices

Goddard Earth Sciences Data and Information Services Center
Global Precipitation, Solar Irradiance, Atmospheric Composition, and Dynamics, Global Modeling
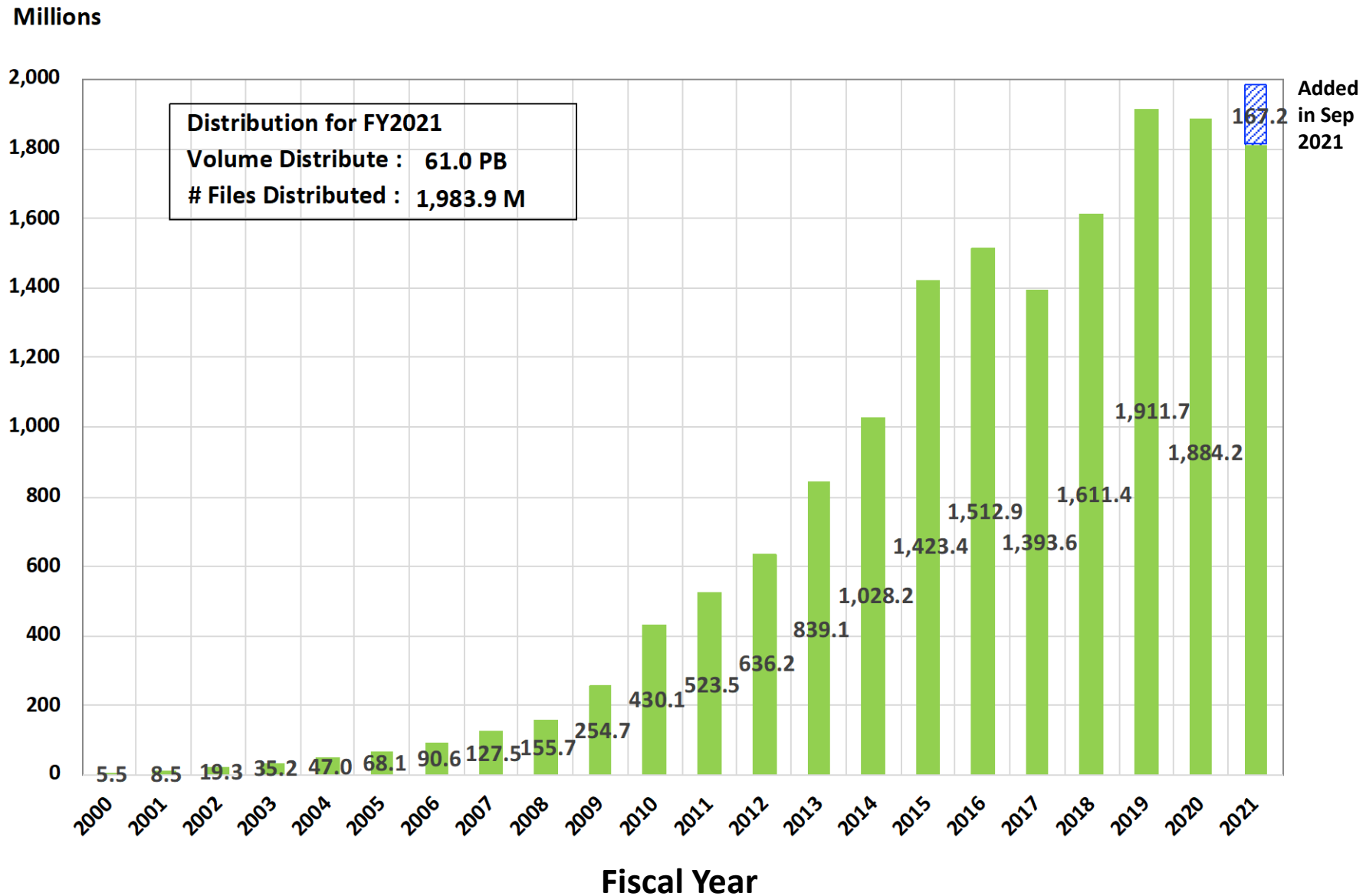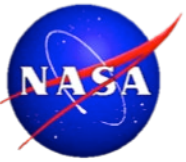
Socioeconomic Data and Applications Center
Human Interactions, Land Use, Environmental Sustainability, Geospatial Data

National Snow and Ice Data Center DAAC
Frozen Ground, Glaciers, Ice Sheets, Sea Ice, Snow, Soil Moisture

Physical Oceanography DAAC
Gravity, Sea Surface Temperature, Ocean Winds, Topography, Circulation & Currents

Suomi National Polar-orbiting Partnership (Suomi-NPP) Atmosphere

Crustal Dynamics Data Information System
Space Geodesy, Solid Earth

Ocean Biology DAAC
Ocean Biology, Sea Surface Temperature

Measurements of Pollution in the Troposphere (MOPITT)

Oak Ridge National Laboratory DAAC
Biogeochemical Dynamics, Ecological Data, Environmental Processes

Level 1 and Atmosphere Archive and Distribution System (LAADS)
MODIS Level-1 and Atmosphere Data Products

Suomi National Polar-orbiting Partnership (Suomi-NPP) (ATMS) and (CrIS) Sounder

Microwave Limb Sounder (MLS)

Global Hydrometeorology Resource Center DAAC
Hazardous Weather, Lightning, Tropical Cyclones and Storm-induced Hazards

Ozone Mapping Profiler Suite (OMPS) Ozone

•

Ozone Monitoring Instrument (OMI)

•

Ozone Mapping Profiler Suite (OMPS) Ozone

•

Ocean Data Processing System (OCDPS)

•

MODIS Adaptive Processing System (MODAPS)

•

Visible Infrared Imaging Radiometer Suite (VIIRS) Land

Advanced Microwave Scanning Radiometer for EOS 2 (AMSR-E/2)

LaRC Atmospheric Science Data Center
Radiation Budget, Clouds, Aerosols, Tropospheric Chemistry

Alaska Satellite Facility DAAC
SAR Products, Sea Ice, Polar Processes, Geophysics

5

Millions

Distribution for FY2021
Volume Distribute : 61.0 PB
 # Files Distributed : 1,983.9 M

Added in Sep 2021

Fiscal Year	Value
2000	5.5
2001	8.5
2002	19.3
2003	35.2
2004	47.0
2005	68.1
2006	90.6
2007	127.5
2008	155.7
2009	254.7
2010	430.1
2011	523.5
2012	636.2
2013	839.1
2014	1,028.2
2015	1,423.4
2016	1,512.9
2017	1,393.6
2018	1,611.4
2019	1,911.7
2020	1,884.2
2021	167.2 (added in Sep 2021)

Fiscal Year

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

Yearly EOSDIS Archive Growth Projections for Current Holdings and New Missions (FY 2019 to FY 2029)

Legend: Current Holdings, New Missions

Y-axis: Archive Voulme (Tbytes), from 0 to 300,000

X-axis: Fiscal Year, 2019 to 2029
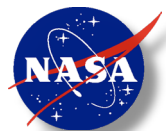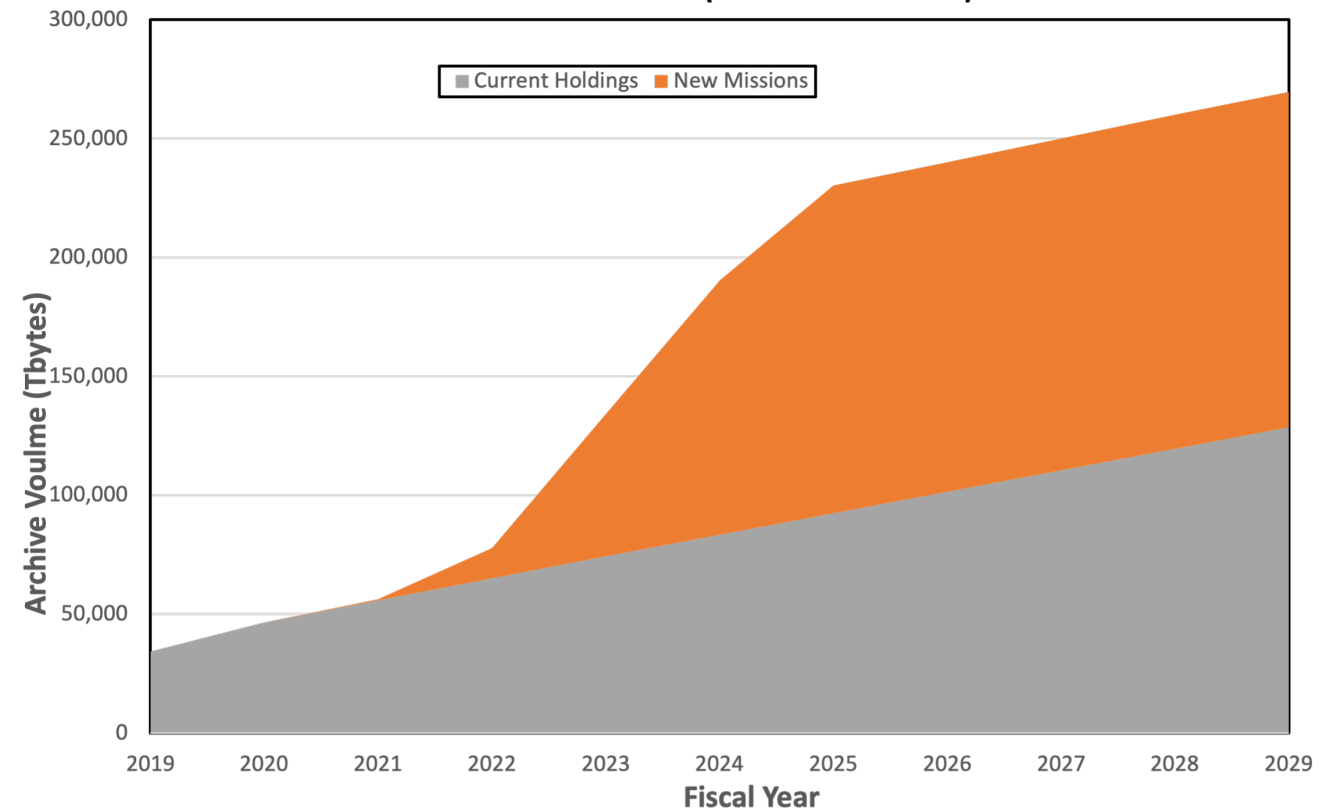
Current User Interaction

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
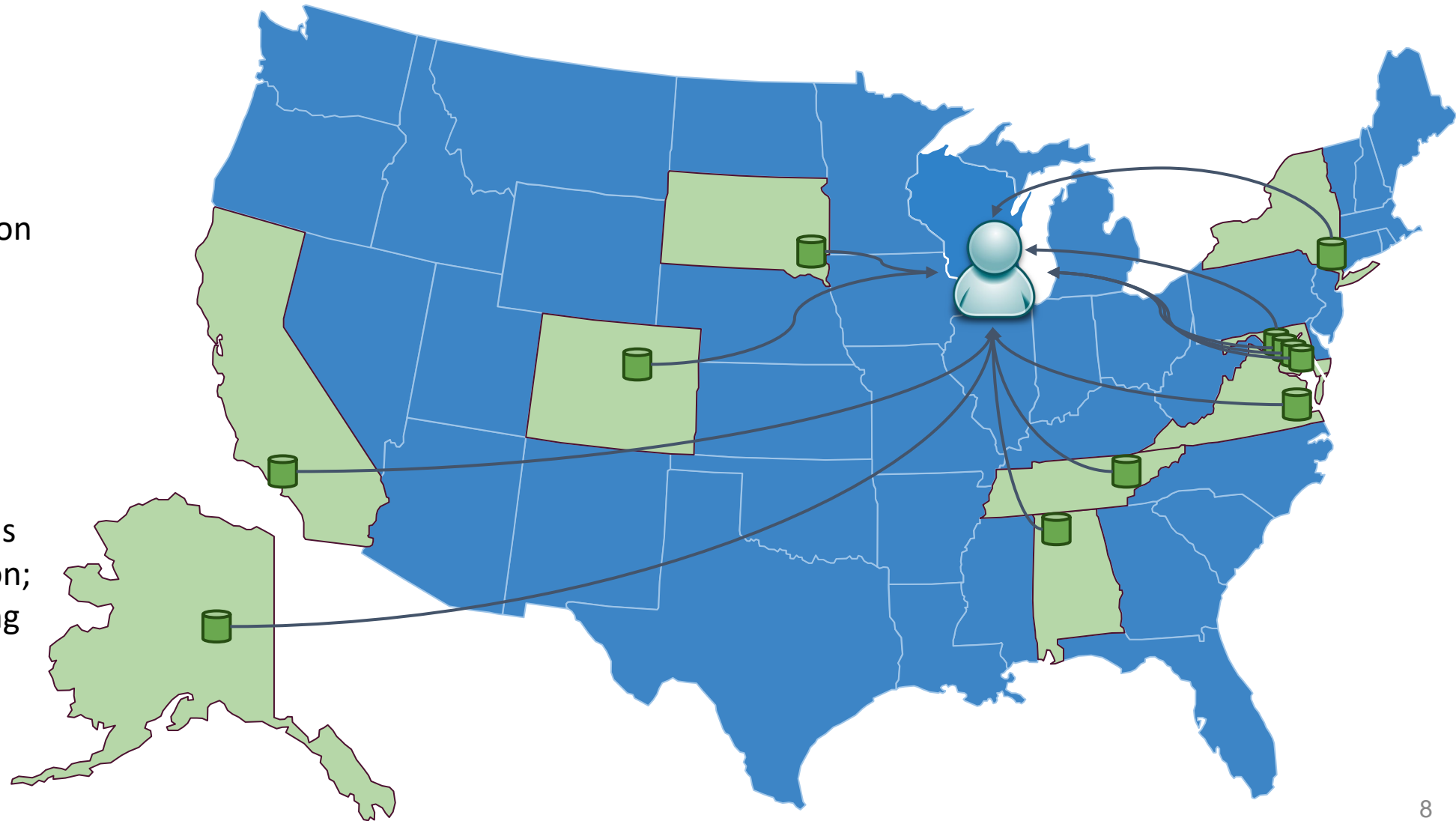
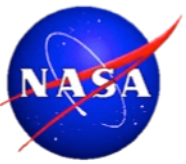# Envisioning an EOSDIS "Data Lake"
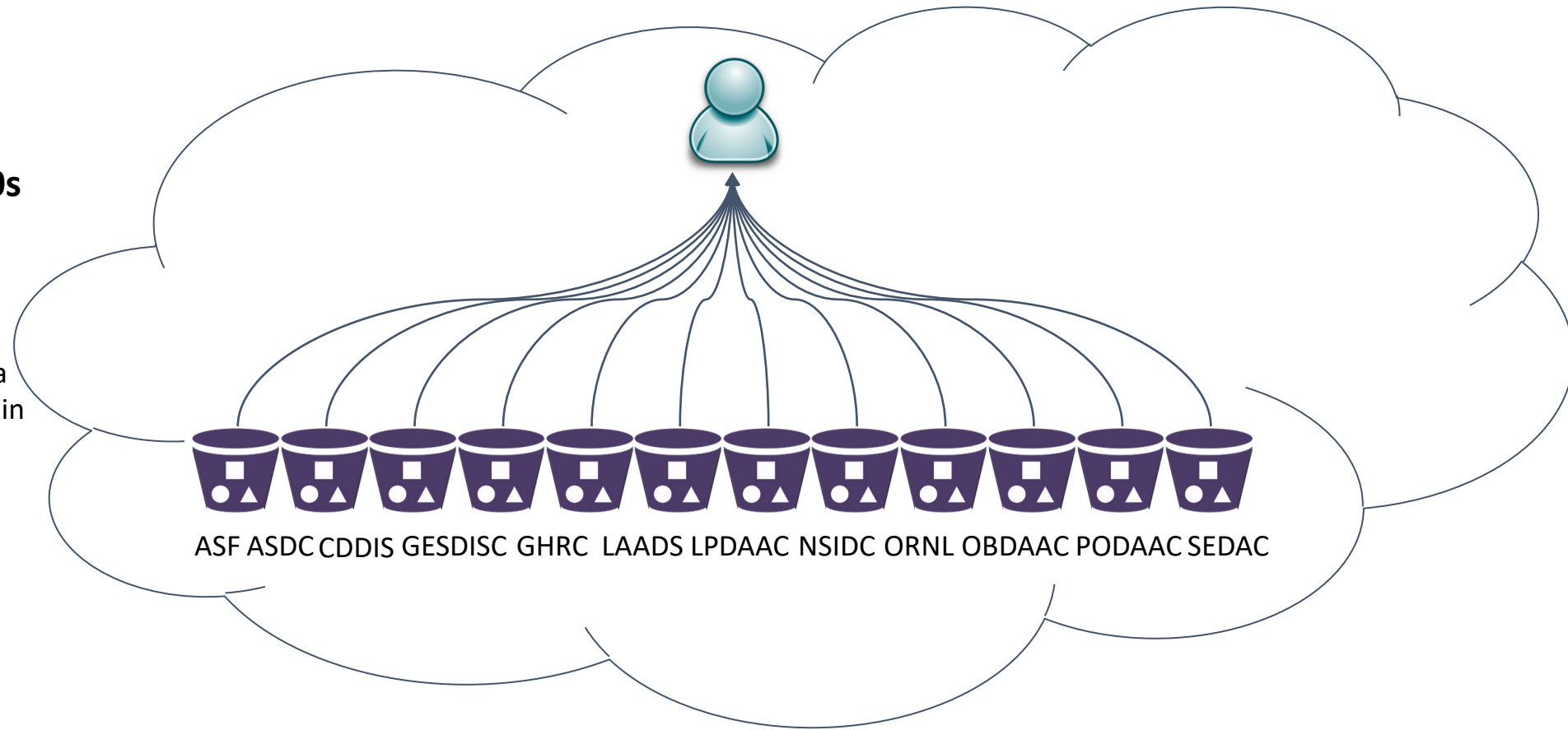
Evolution of EOSDIS in the 2020s

- 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

ASF ASDC CDDIS GESDISC GHRC LAADS LPDAAC NSIDC ORNL OBDAAC PODAAC SEDAC

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.

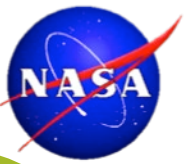- **Choice**: Users can still download data if they prefer.

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

Earthdata Cloud -EDC-

- Improve the efficiency of NASA's data systems operations – <u>continues free and open access to data</u>

- 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

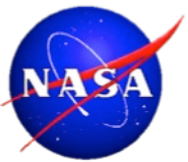- Transparent/extendable open source processing framework

NASA Earthdata Cloud (EDC) – http://earthdata.nasa.gov
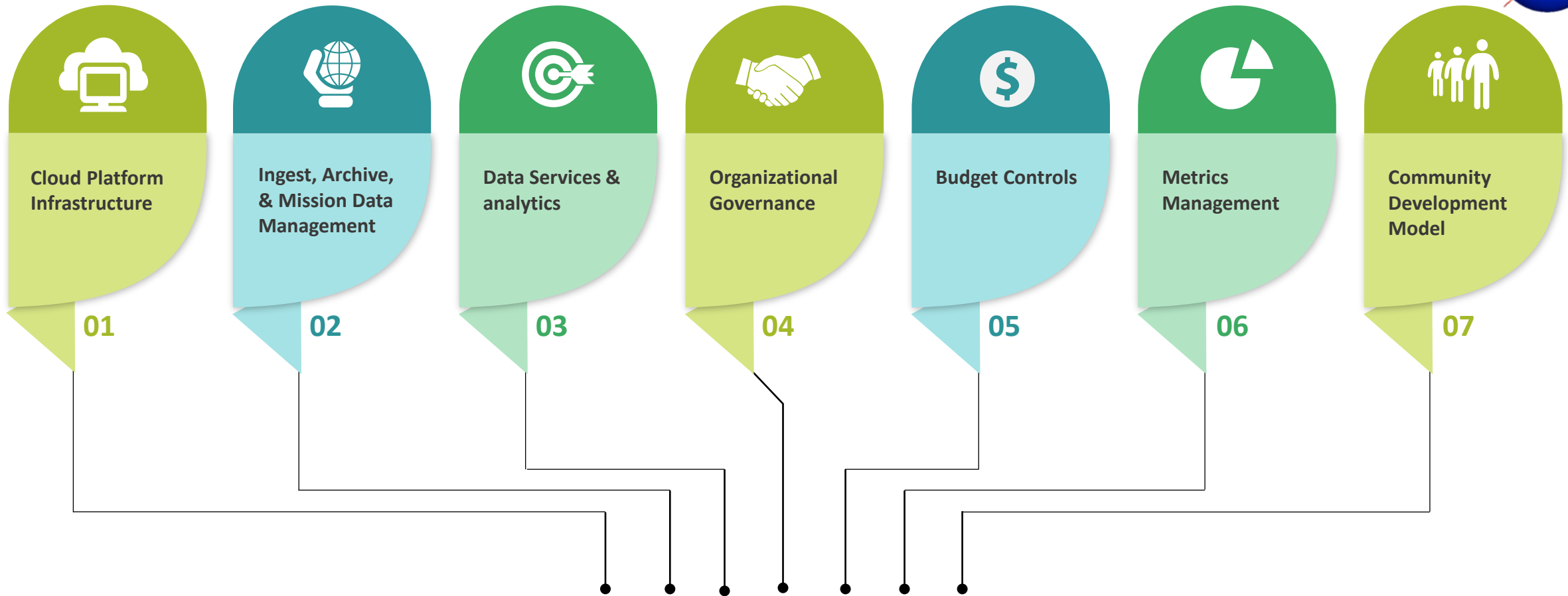
Components of the Earthdata Cloud

01 Cloud Platform Infrastructure
02 Ingest, Archive, & Mission Data Management
03 Data Services & analytics
04 Organizational Governance
05 Budget Controls
06 Metrics Management
07 Community Development Model

# What is the Earthdata Cloud Platform?

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

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

NASA Earthdata Cloud (EDC) – http://earthdata.nasa.gov
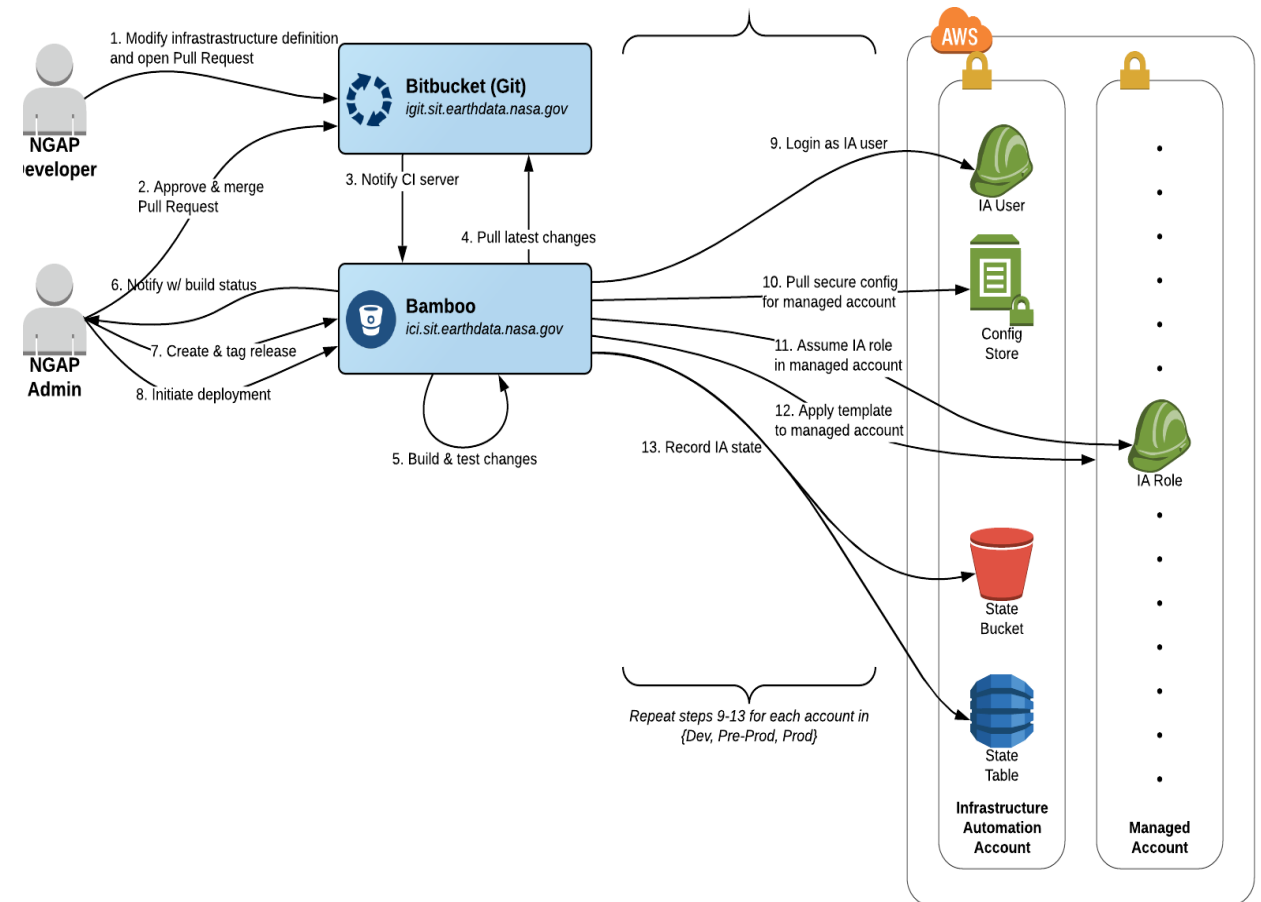
Unifying Ingest & Archive in the Cloud: Cumulus

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

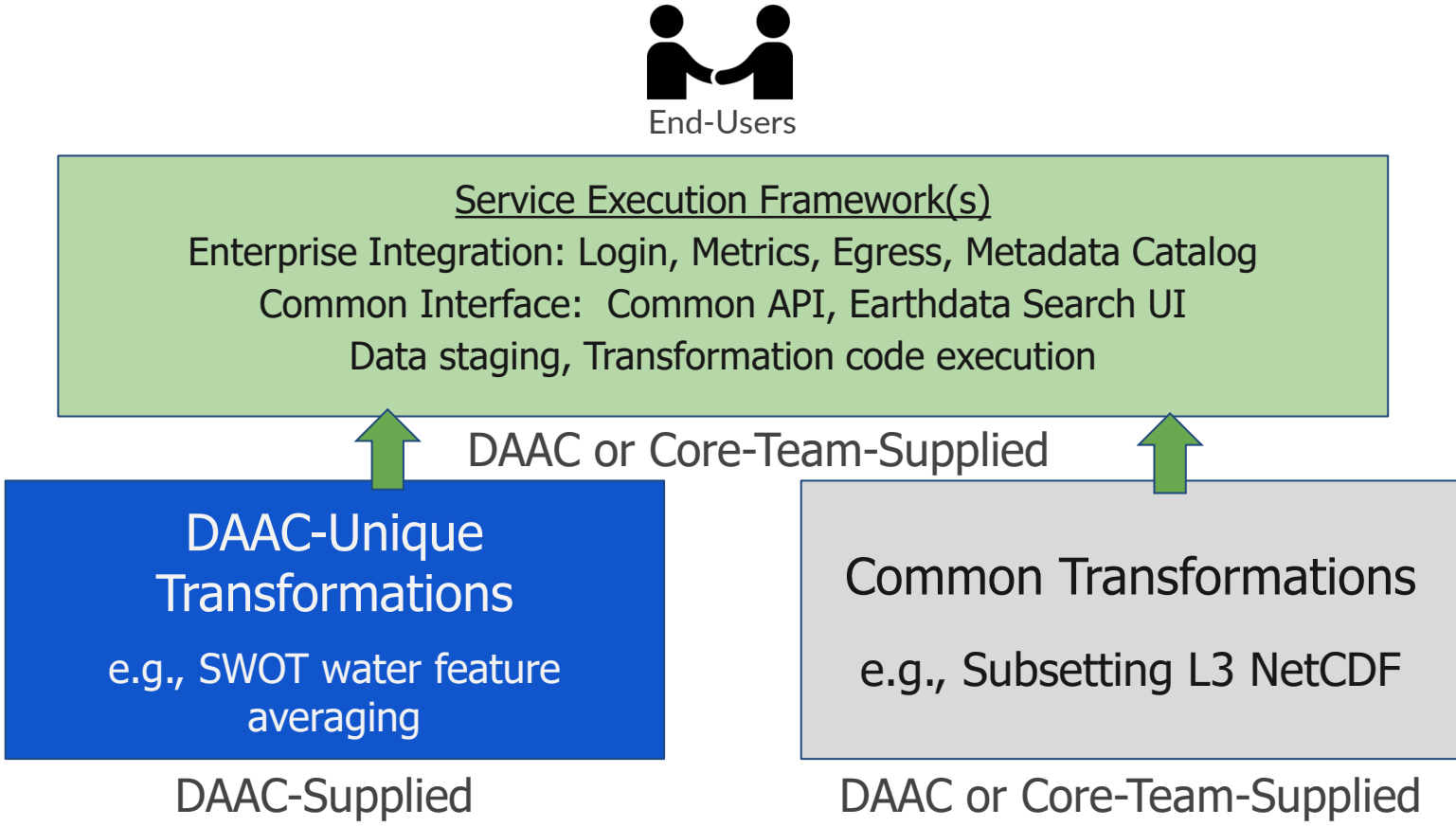**Database** store status, logs, and other system state information

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

Dashboard create and execute workflows, monitor system

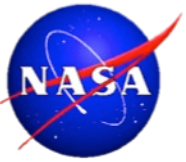NASA Earthdata Cloud (EDC) – http://earthdata.nasa.gov
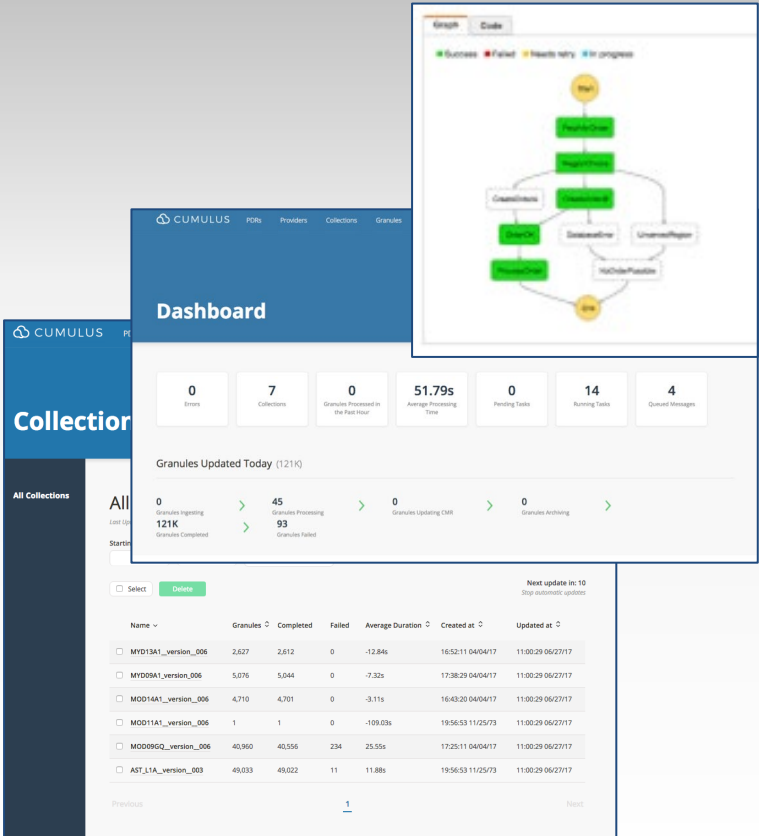
Unifying Data Services in the Cloud: Harmony

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.
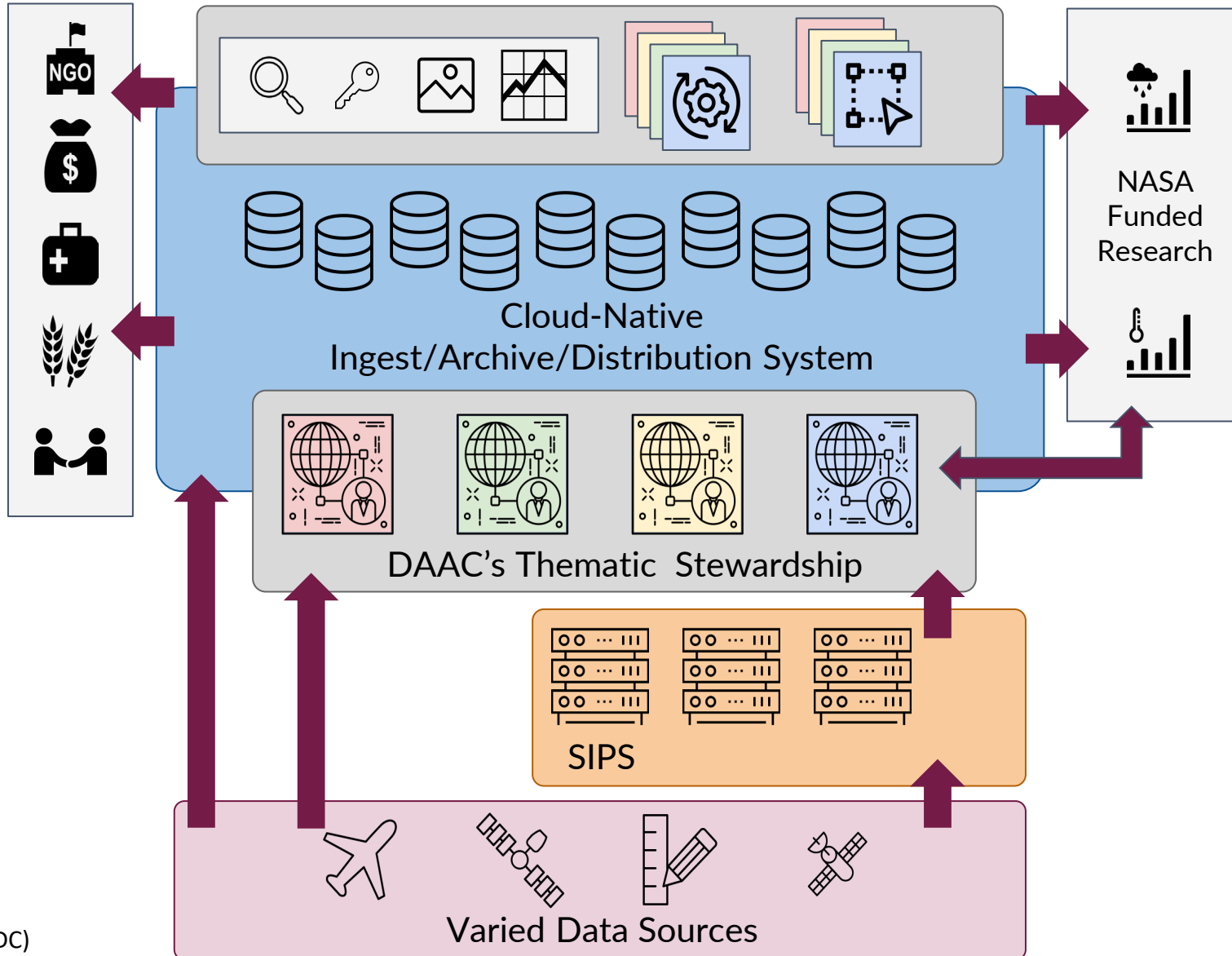
End-Users

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

Common Transformations

e.g., Subsetting L3 NetCDF

DAAC-Supplied

DAAC or Core-Team-Supplied

NASA Earthdata Cloud (EDC) – http://earthdata.nasa.gov

15
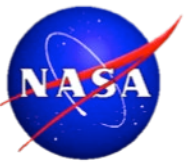
EOSDIS Conceptual Cloud Based Architecture

Benefits

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

Expert user support

Streamlined product addition

Reduced duplication of tools and services

Cloud-Native
Ingest/Archive/Distribution System

DAAC's Thematic Stewardship

SIPS

Varied Data Sources

NASA Funded Research

Challenges

Development coordination

Cost Management

Shifting Labor Needs

Security/Export Compliance

Vendor Lock In

NASA Earthdata Cloud (EDC)

16

Additional information is available at:

http://earthdata.nasa.gov/

https://earthdata.nasa.gov/learn/articles/cloud-articles-announcement

Thanks for the support from all of the scientists, engineers, accountants, developers and operators of the EOSDIS system!