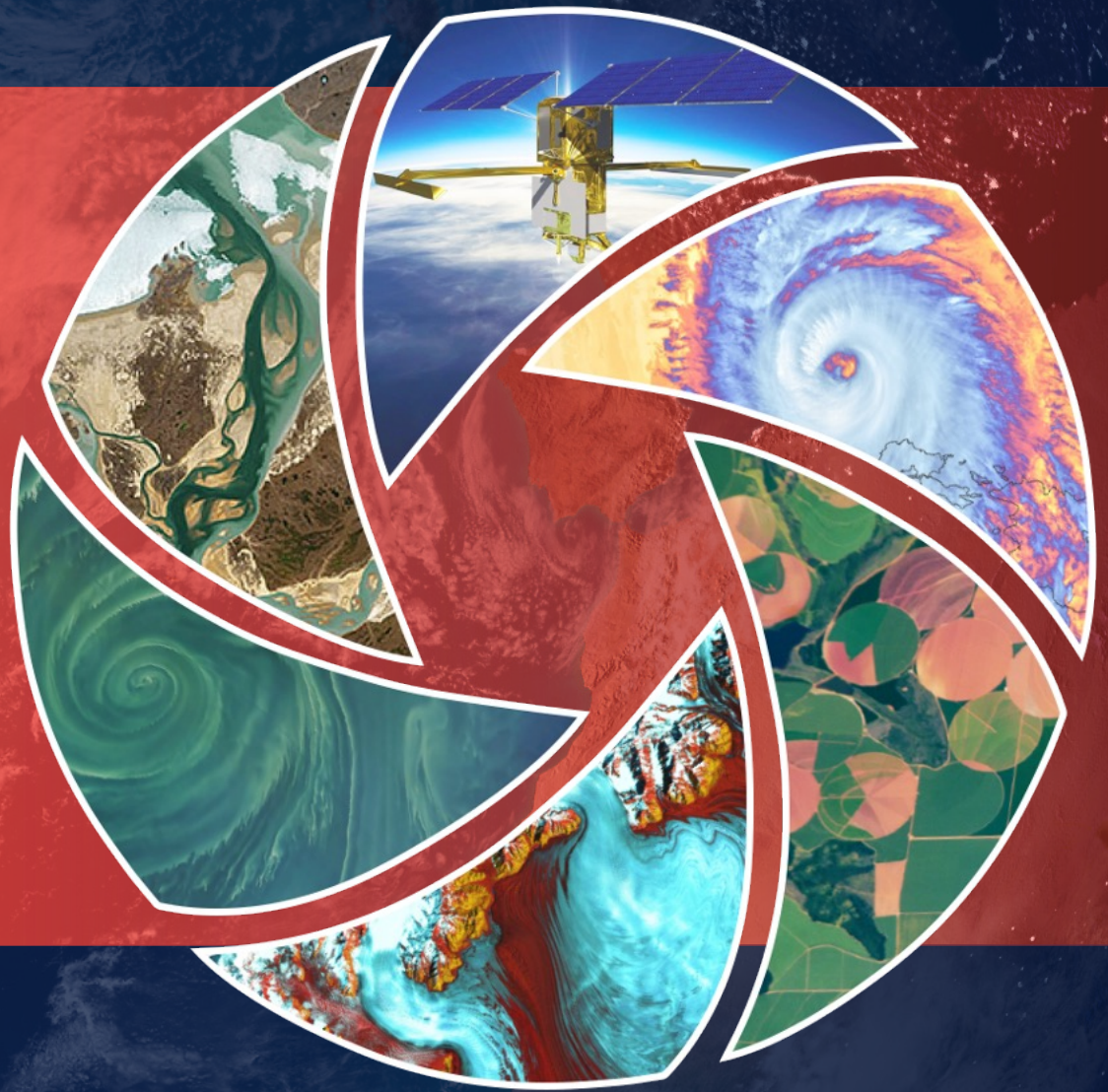
Advancing Global Aerosol Prediction through Open Science, and Cloud-Based Interoperability Strategies

November 9, 2023

Jennifer Wei, ESDIS Project Scientist, NASA Goddard Space Flight Center

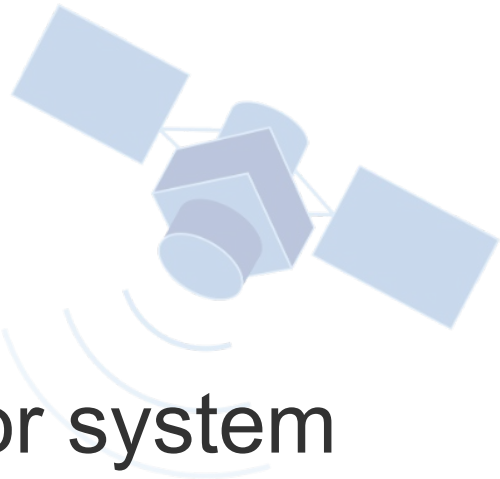


EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE

Agenda

1. What and Why is Interoperability
2. Challenges
3. NASA’s Interoperability Strategies

Interoperability



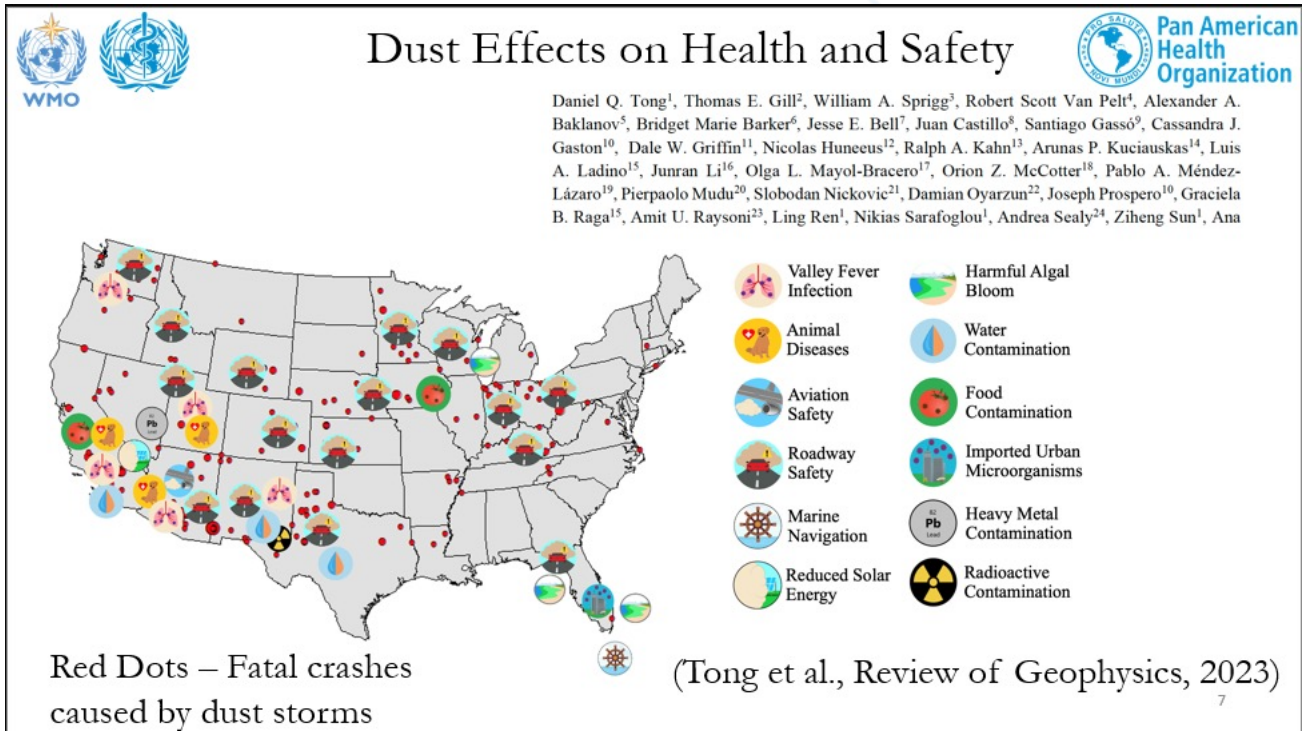
Interoperability is a characteristic of a product or system to **work with** other products or systems (Wikipedia)

Purpose – To **increase producibility or usability** of the data systems, devices, or organizations

EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE

Significance of Aerosol Prediction

- Aerosols play a pivotal role
- Environmental Monitoring
  - Environmental Quality
  - Impact on Ecosystem
- Public Health
  - Air Quality and Respiratory Health
  - Epidemiological Studies
- Climate Research
  - Climate Forcing
  - Global Climate Models



EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE

Enhancing Aerosol Prediction Through Data Integration

Approach Interoperability and Collaboration

Challenges

Data Silos, Data Incompatibility, Limited Accessibility, Limited Data Services, Redundant Efforts

Transition Research Data and Services to OPS (Open Science Requirement)

But How?

Solutions

Unified Data Standards, Unified Metadata Standards, Data Service Harmonization, Enhanced Collaboration, Minimizing Redundancy

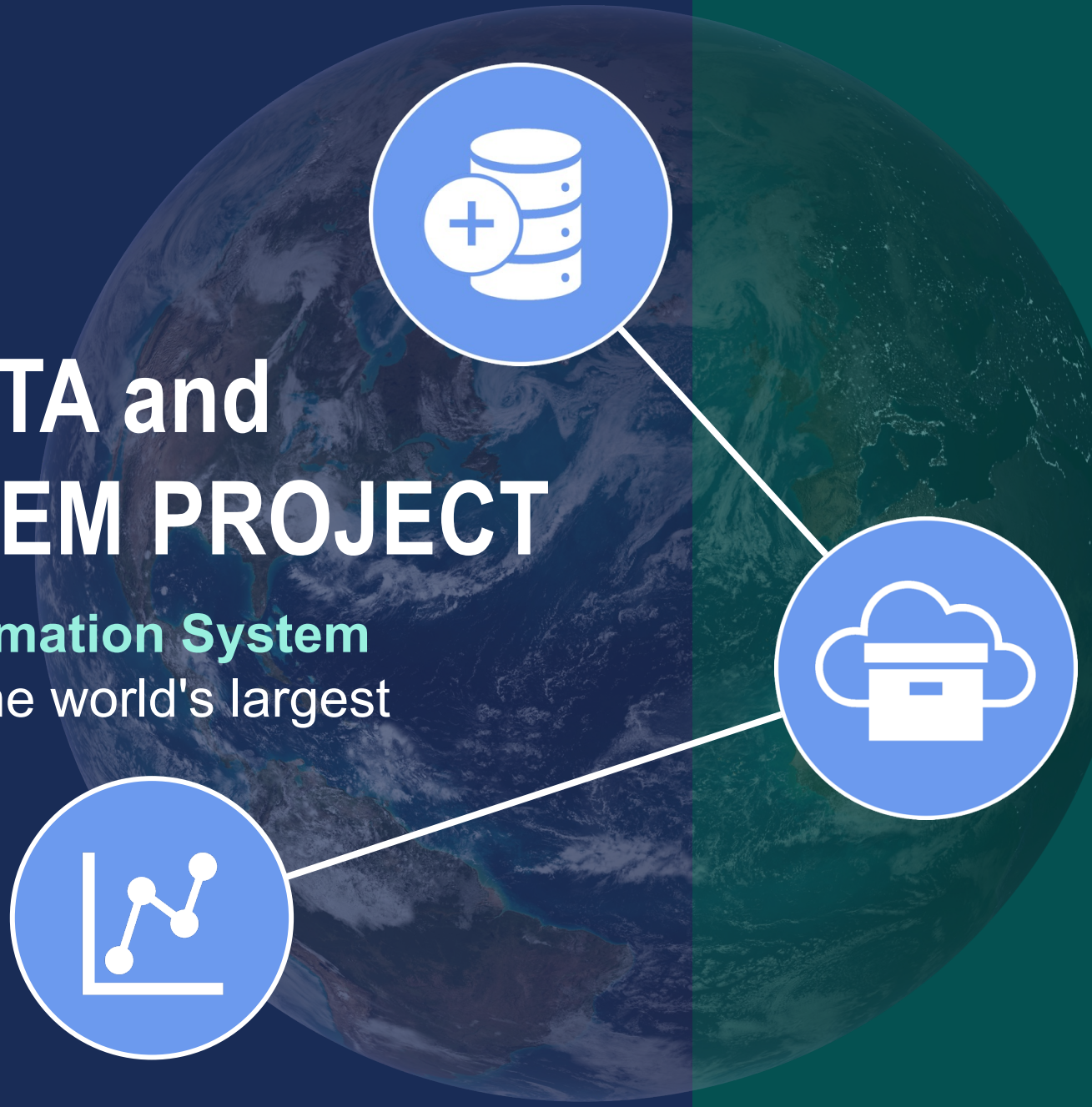
EARTH SCIENCE DATA SYSTEMS PROGRAM

NASA's Earth Science Data Systems (ESDS) Program oversees the entire Earth science data life cycle and facilitates unrestricted access to the data researchers, managers, and governments need to understand and protect our planet.

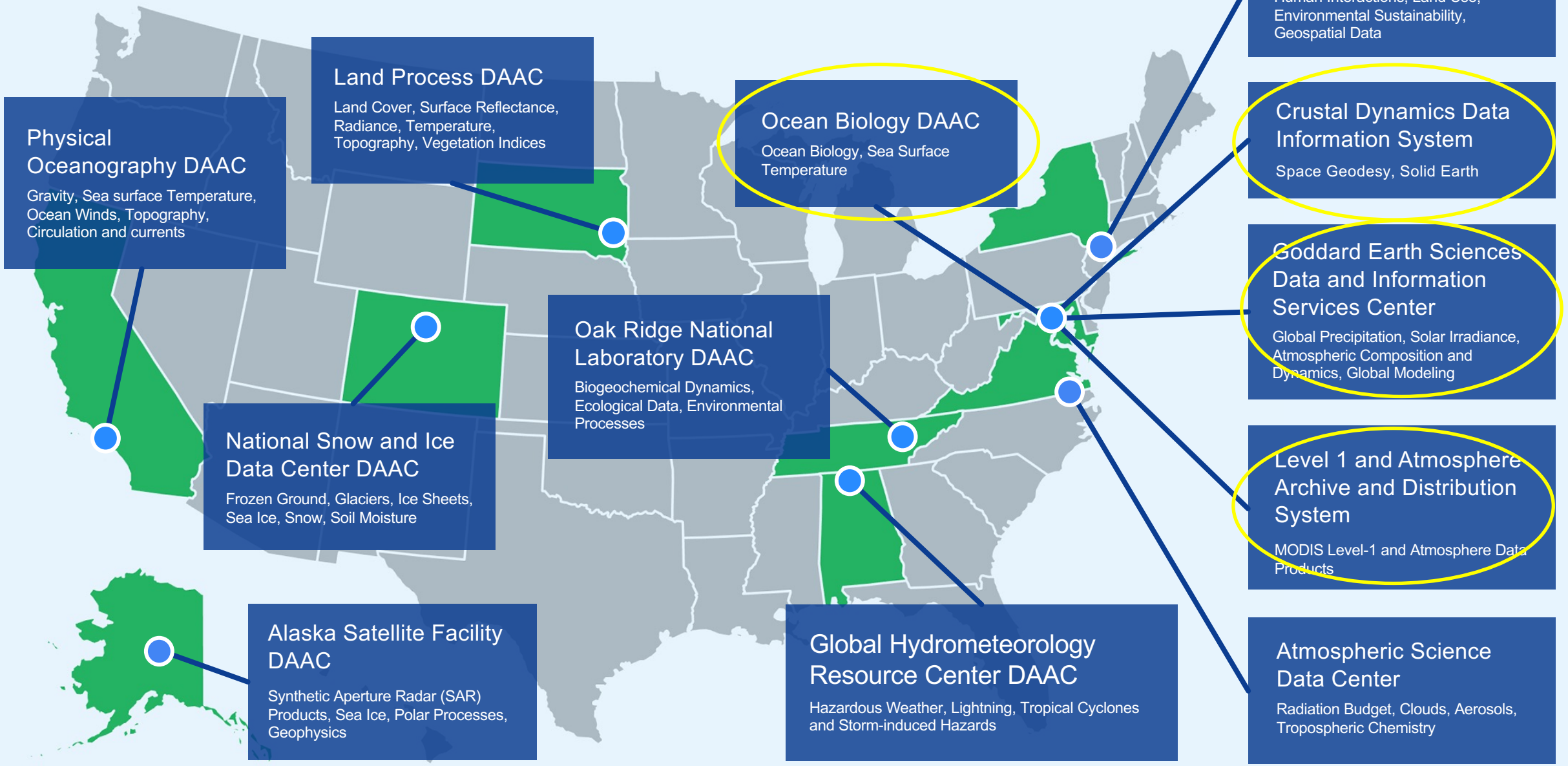
EARTH SCIENCE DATA SYSTEMS PROGRAM

EARTH SCIENCE DATA and INFORMATION SYSTEM PROJECT

The **Earth Science Data and Information System (ESDIS) Project** manages one of the world's largest archives of Earth science data.



NASA's Distributed Active Archive Centers (DAACs)



Physical Oceanography DAAC

Gravity, Sea surface Temperature, Ocean Winds, Topography, Circulation and currents

Land Process DAAC

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

Ocean Biology DAAC

Ocean Biology, Sea Surface Temperature

Socioeconomic Data and Applications Center

Human Interactions, Land Use, Environmental Sustainability, Geospatial Data

Crustal Dynamics Data Information System

Space Geodesy, Solid Earth

Goddard Earth Sciences Data and Information Services Center

Global Precipitation, Solar Irradiance, Atmospheric Composition and Dynamics, Global Modeling

Oak Ridge National Laboratory DAAC

Biogeochemical Dynamics, Ecological Data, Environmental Processes

National Snow and Ice Data Center DAAC

Frozen Ground, Glaciers, Ice Sheets, Sea Ice, Snow, Soil Moisture

Level 1 and Atmosphere Archive and Distribution System

MODIS Level-1 and Atmosphere Data Products

Alaska Satellite Facility DAAC

Synthetic Aperture Radar (SAR) Products, Sea Ice, Polar Processes, Geophysics

Global Hydrometeorology Resource Center DAAC

Hazardous Weather, Lightning, Tropical Cyclones and Storm-induced Hazards

Atmospheric Science Data Center

Radiation Budget, Clouds, Aerosols, Tropospheric Chemistry

Interoperability Strategy
What was then?
(~10 years ago)

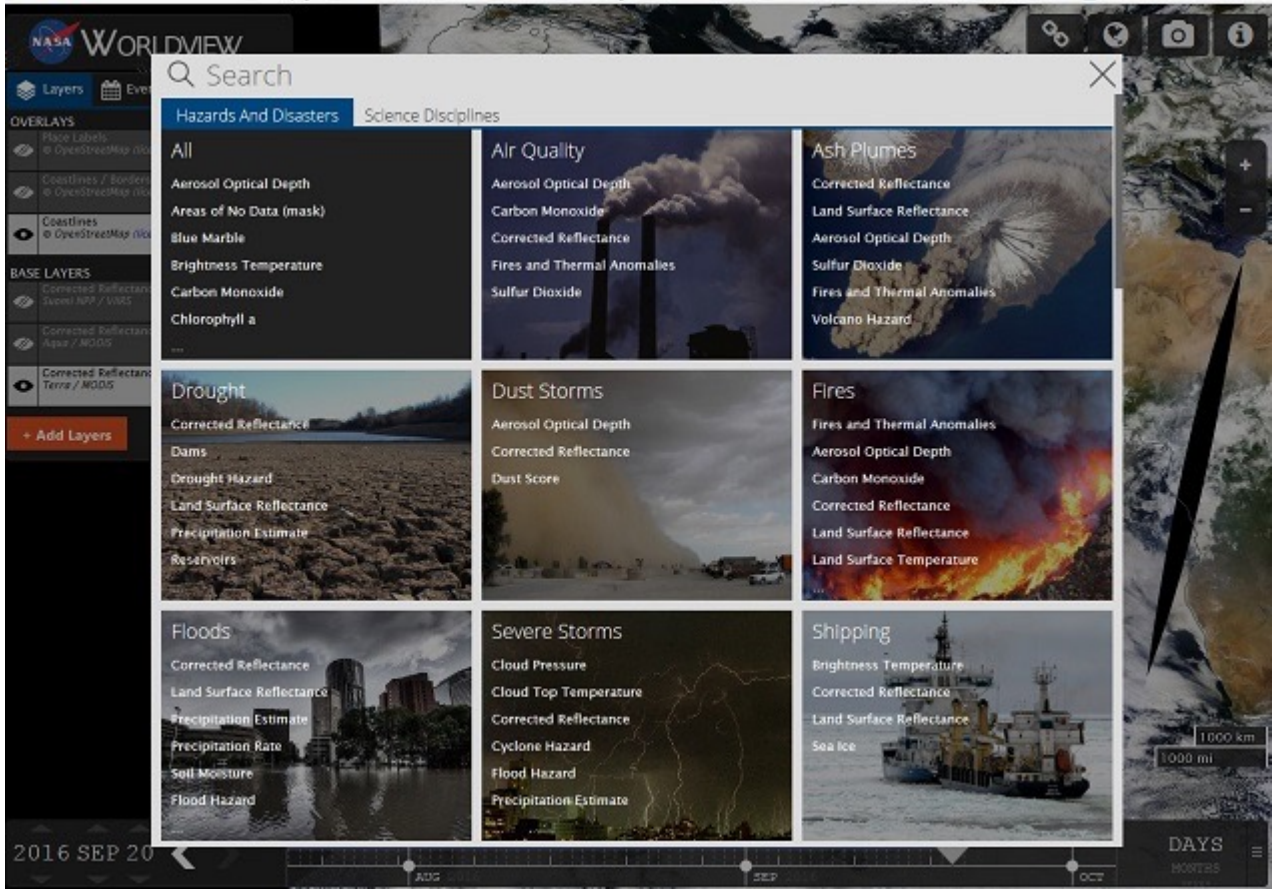
Big Earth Data Initiative (BEDI)

- In **2012, President Obama's "Big Data Research and Development Initiative"** seeks to improve our ability to acquire knowledge and discover insights into large and complex collections of digital data.
- In **January 2013**, the White House Office of Science and Technology Policy (OSTP) hosted an Interagency Panel on Interoperability at the ESIP Federation Winter Meeting
- Then **NASA , NOAA , and USGS** proposed the **Big Earth Data Initiative (BEDI)** focused on **interoperability** of data between agencies, and in particular, increasing the **Discoverability**, **Accessibility** and **Usability** of earth observation data.
- The **President's 2014 budget** request included funding for NASA, NOAA, USGS, and USDA to implement the BEDI objectives.

Interoperability - Target User Model

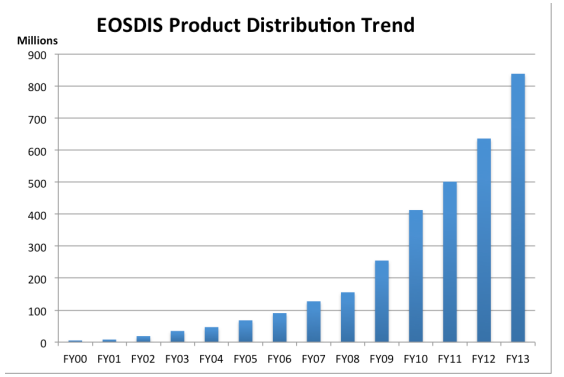
Priority	User Persona
High	Interdisciplinary Scientists
	Applications Practitioners Government Private Sector
	Decision Support tools (machine-level)
Medium	Discipline Research Scientists
	Applications Researchers (esp. if funded by NASA)
	Citizen Scientists
Low	Data Analytics Scientists (i.e. working on techniques)

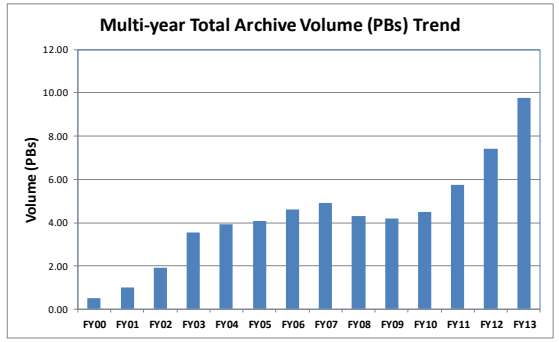
Enhancing EOSDIS Data Usability - Societal Benefit Areas (SBAs)

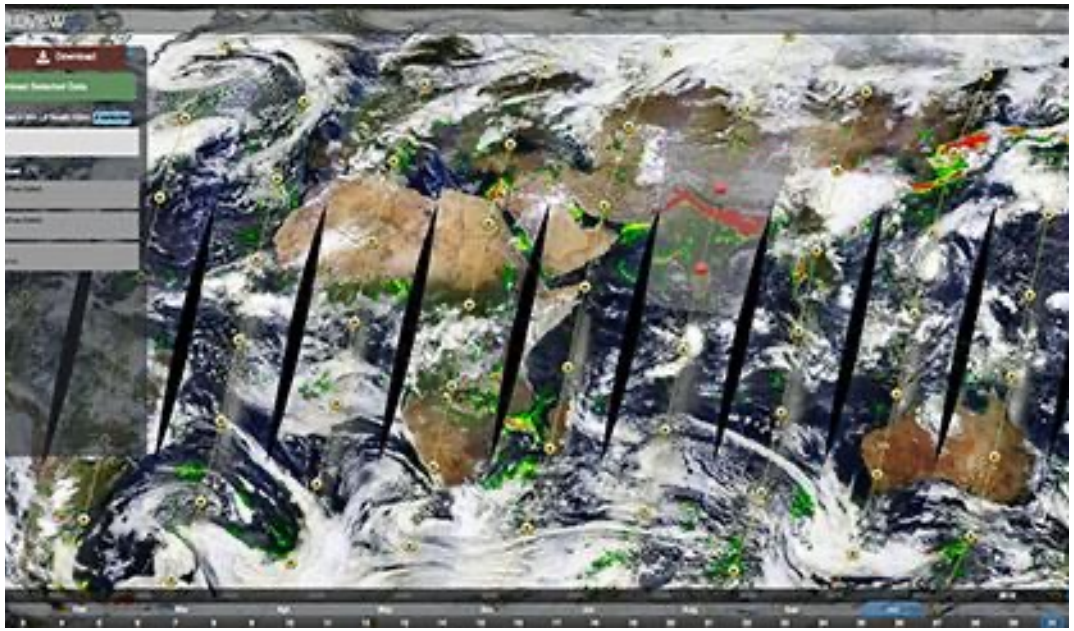


How - EOSDIS Approach

- Enhancing the ability to access EOSDIS data via web accessible APIs (e.g., OPeNDAP, WEBIFICATION [W10N]
- Increasing the ability of commercial search engines to discover EOSDIS data sets
- Enhancing our Global Imagery Browse Services (GIBS) capability to provide pre-generated full resolution browse imagery (with links to the underlying data).
- Enhancing and formalizing standards like GeoTIFF
- Improved support for Open Standards (via ISO and W3C, as well as OGC)

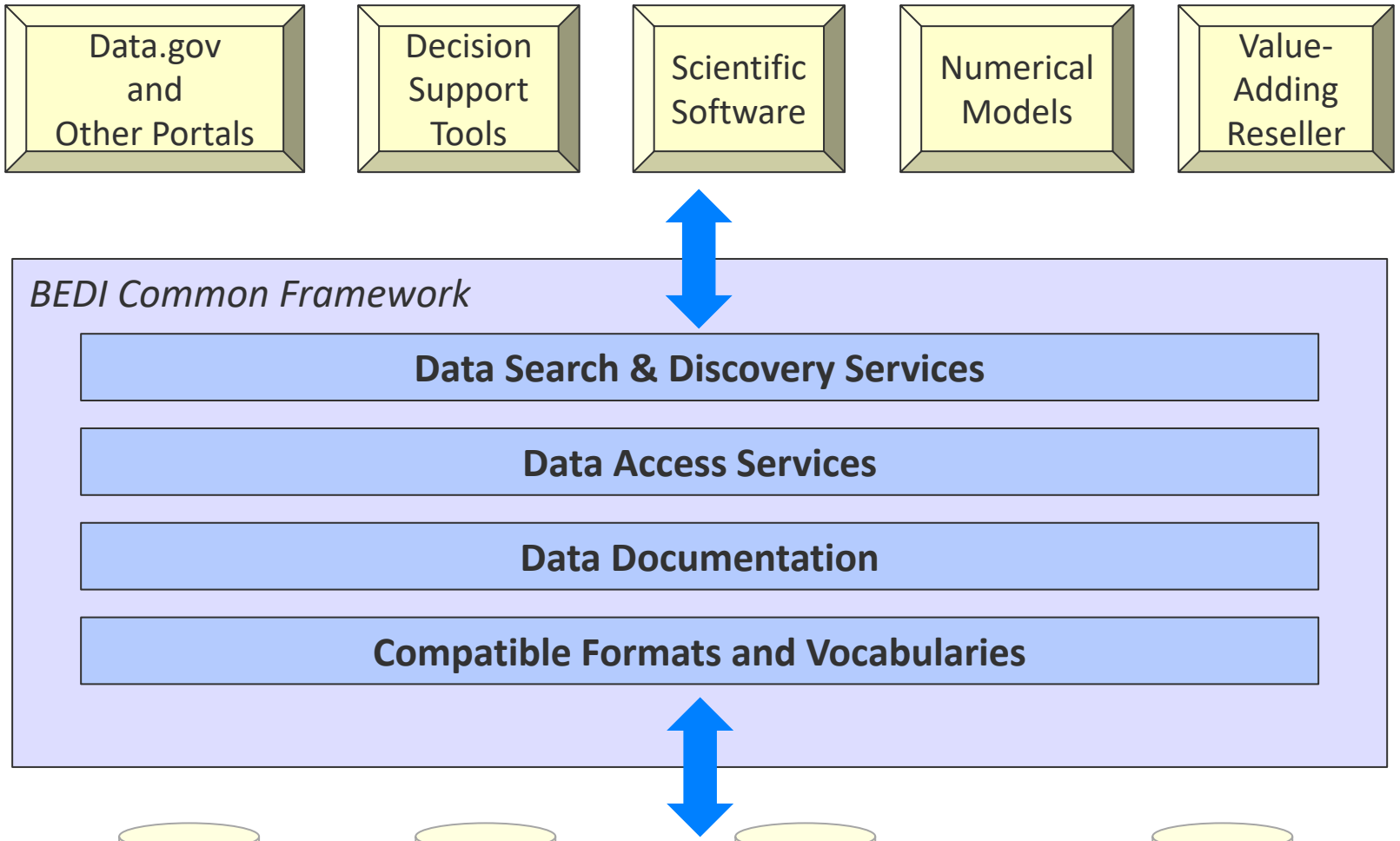






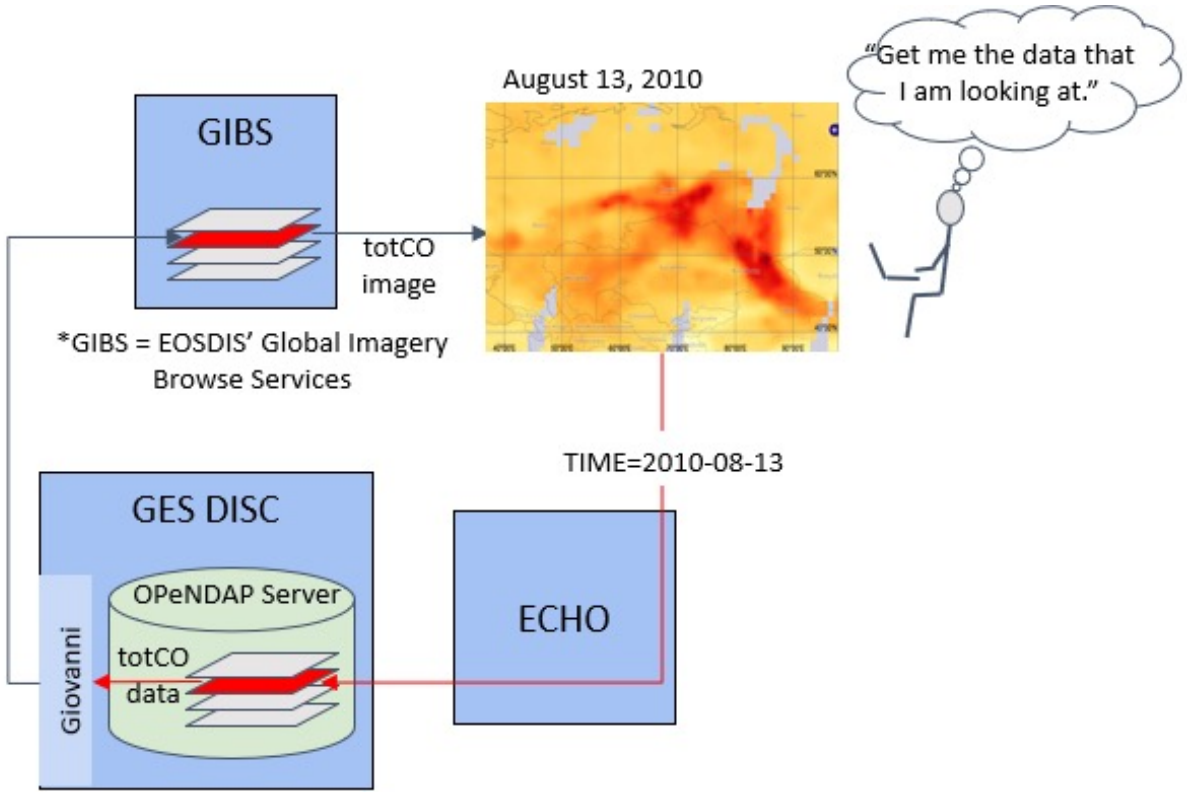
EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE

BEDI Common Framework



How a DAAC follows?
Use Case: GES DISC Giovanni

Goal	Approach
Fine-grained machine-level access (from applications community)	➔ More OPeNDAP
Better discoverability	➔ CMR publication ➔ Dataset landing pages + DOIs ➔ More Browse for GIBS*
Easier to Use	➔ Time aggregated OPeNDAP ➔ "Right-sized" Variable-level User Guide ➔ Create more data recipe (HowTo) ➔ Improve Giovanni workflow



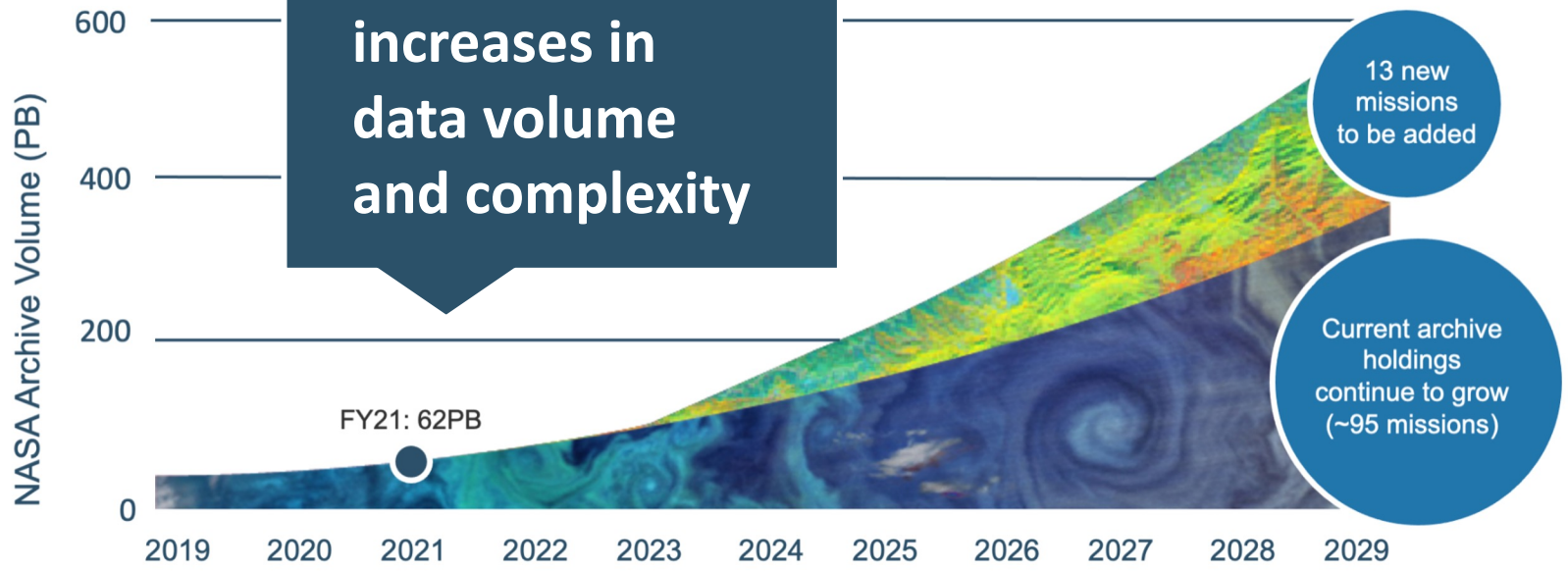
EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE

What is Current and Future Landscape of Interoperability (till 2030)

Challenges facing NASA's Earth Data ecosystem

Push for transdisciplinary, multi-mission science

Dramatic increases in data volume and complexity



Increased emphasis on open, accessible, and reproducible science

Addressing Users' Needs Together

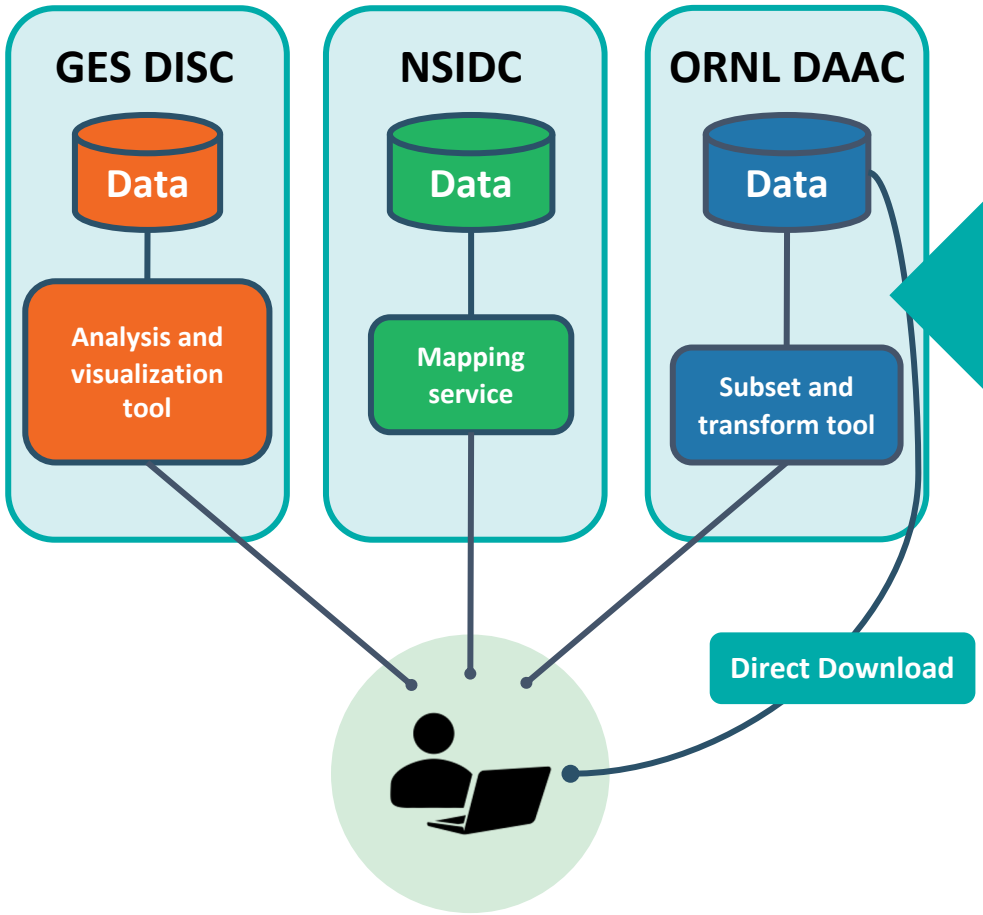
Empower Via Inclusivity

Our goal is to develop opportunities for all data users—from the novice learner to the experienced program scientist.

Our users include:

- GIS Users
- Data Scientists
- Practical Data Users
- Non-Traditional Remote Sensing Data Users

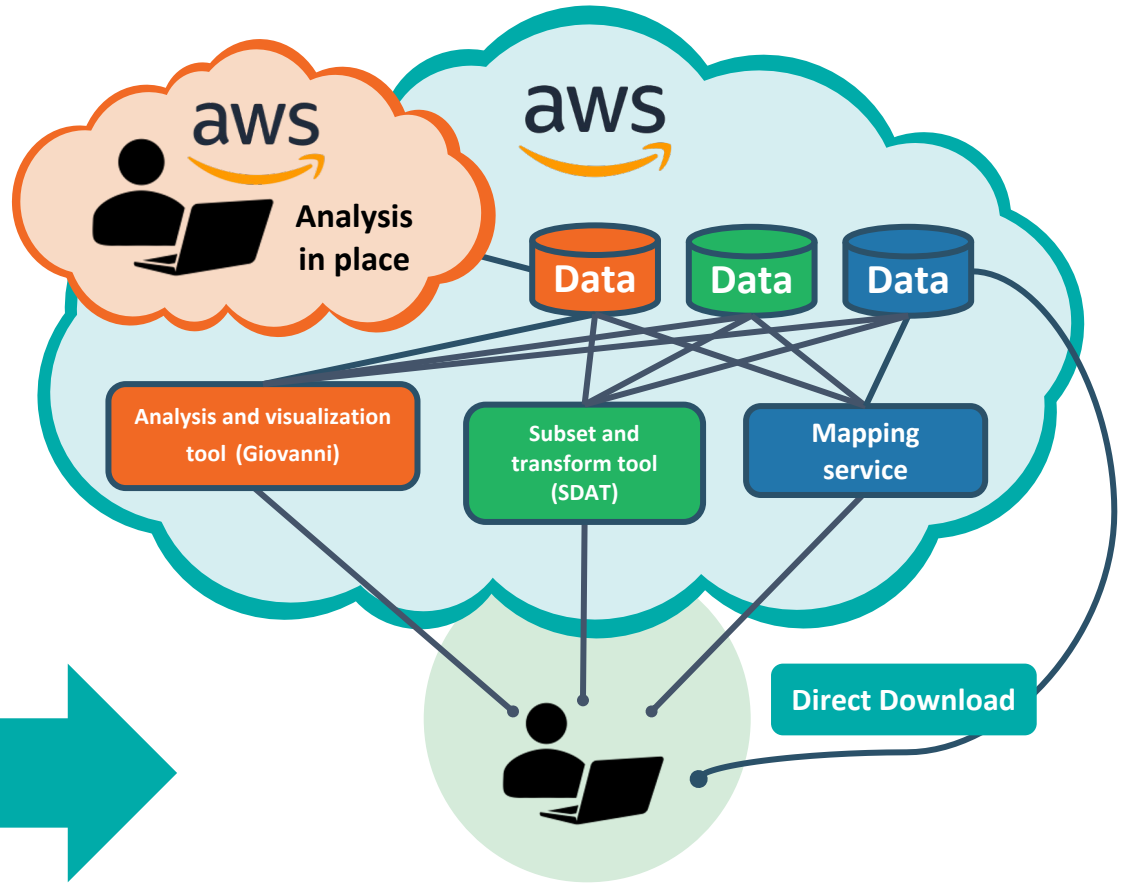
EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE



Where we are

VS

Where we're going



What will stay the same?

- All NASA Earth Science data will continue to be 100% free and open to public.
- Existing data services (including direct download) will continue to work without disruption
- On-premise HPC will continue to play an important role in the NASA computing ecosystem

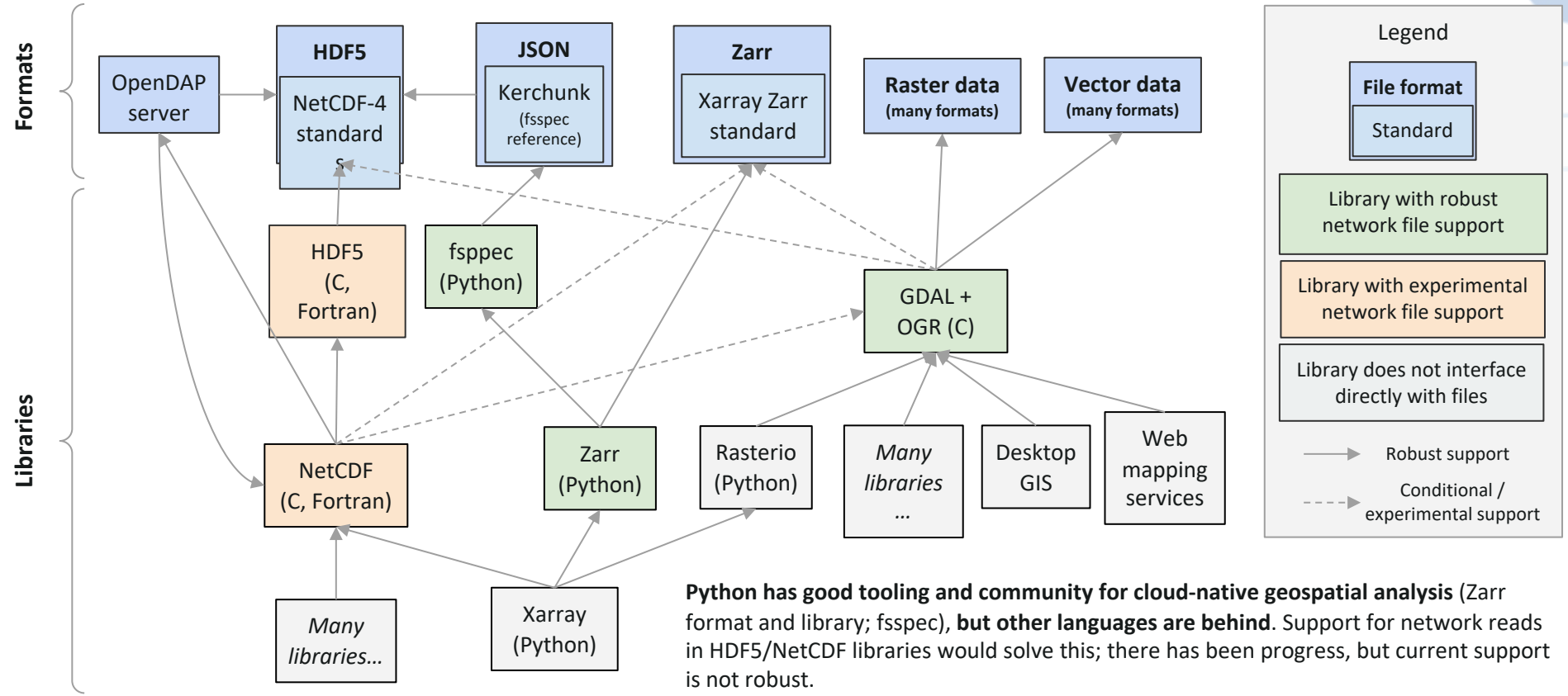
What will change?

- It will be easier for DAACs to collaborate and develop tools that work with more datasets, now that they always have direct access to each other's data.
- *New* options for analyzing data and developing tools "in place" in the cloud, without needing to download data.

Cloud Migration Progress/Timeline

LAADS DAAC Phased Dataset Migration				
Phase	**Collection**	**Year**	**Total Volume (TB)**	**Migration Status**
1.1	MODIS C61 L1, Geolocation and Cloud Mask (Cohort-2)	2022	1,260	Complete
1.2	LAADS DAAC product "Top 50 + 25" List	2022	355	Complete
1.3	MODIS C61 L2 and L3 Atmosphere Products	2023	35	Complete
2.1	C2 SNPP VIIRS L1B and Geolocation	2023	562	Complete
2.2	C2.1 J1 VIIRS L1B and Geolocation	2023	306	Complete
2.3	SNPP/J1 VIIRS C1/C1.1/C2 Atmosphere Products	2023	572	Complete
3.1	MODIS C61 L2 Land Surface Reflectance	2023	781	Complete
4.1	C6 Long Term Data Records from NOAA POES and ESA MetOps	2023	17	Complete
4.2	MERIS, Sentinel 3A, and Sentinel 3B L1 Products	2024	1,133	Complete

Common Data Formats and Open-Sourced Libraries



Python has good tooling and community for cloud-native geospatial analysis (Zarr format and library; fsspec), **but other languages are behind**. Support for network reads in HDF5/NetCDF libraries would solve this; there has been progress, but current support is not robust.

Desktop GIS and web mapping are the most popular methods for data analysis, but **fragmentation between multi-dimensional (NetCDF/Zarr) and geospatial (GDAL/OGR)** standards, tools, and communities inhibits usage of model data in these tools.

EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE

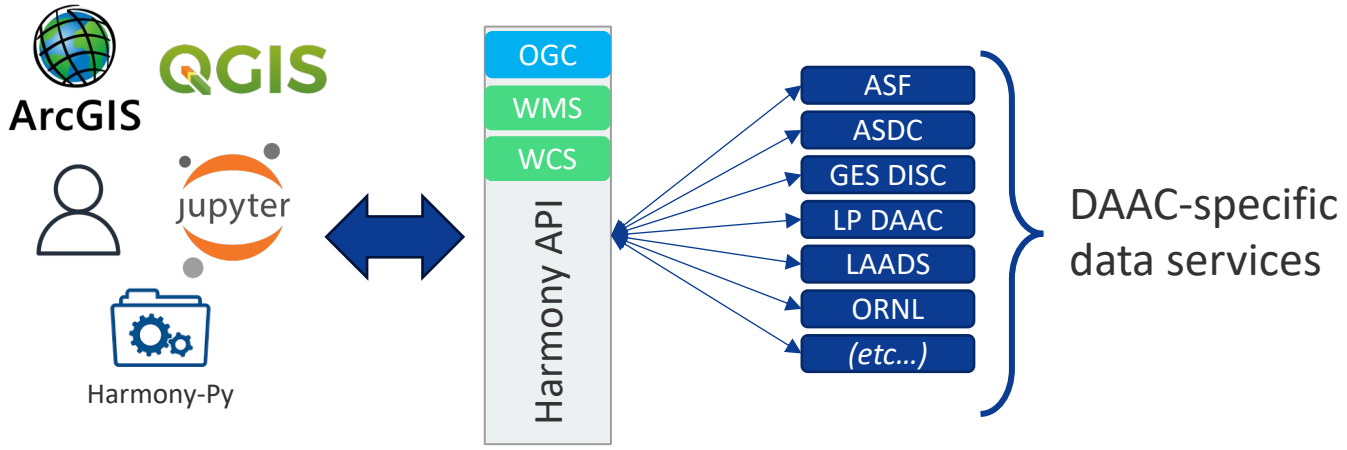
Courtesy of Alexey Shiklomanov, ESDIS Project Scientist

Harmony

https://harmony.earthdata.nasa.gov/

Unified API for data transformation and subsetting services across the DAACs.

Example: Requesting a re-gridded version of a dataset.




```
In [ ]: l3_collection = 'C1234088182-EEDTEST'
        l2_collection = 'C1233860183-EEDTEST'

In [ ]: coverages_root = 'https://harmony.uat.earthdata.nasa.gov/{collection}/ogc-api-coverages/1.0.0/collections/{variable}/coverag
```

Single central API endpoint; not DAAC-specific!

Basic Regridding

```
In [ ]: response = get(
            coverages_root.format(
                collection=l2_collection,
                variable='all'),
            params={
                'outputCrs': 'EPSG:4326',
                'subset': 'time("2020-01-15T16:00:00Z":"2020-01-15T17:00:00Z")'})
        show(response)
```

Advanced Regridding

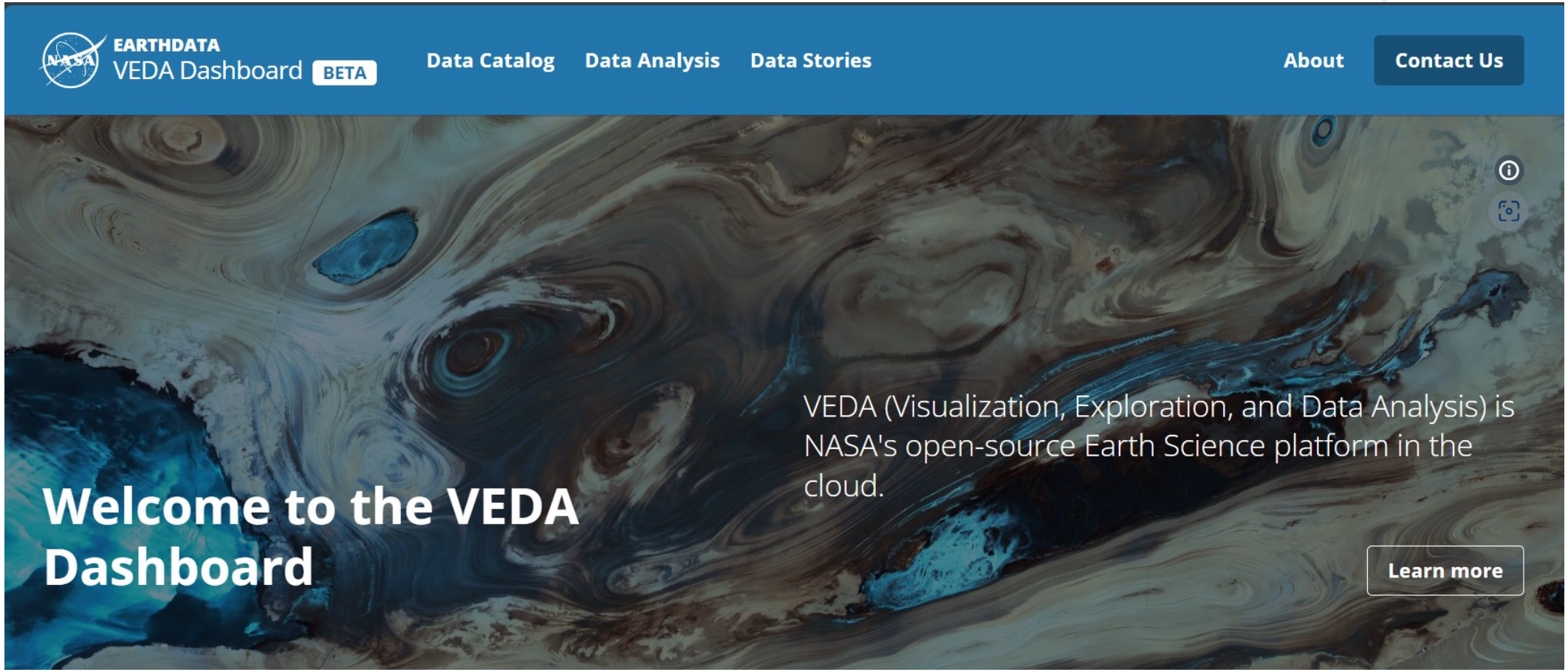
```
In [ ]: response = get(
            coverages_root.format(
                collection=l2_collection,
                variable='all'),
            params={
                'outputCrs': '+proj=lcc +lat_1=43 +lat_2=62 +lat_0=30 +lon_0=10 +x_0=0 +y_0=0 +ellps=intl +units=m +no_defs',
                'interpolation': 'near',
                'scaleExtent': '-4000000,-1000000,5000000,7000000',
                'subset': 'time("2020-01-15T16:00:00Z":"2020-01-15T17:00:00Z")'})

        show(response)
```

Generic regridding request, specific to regridding service but *not* specific to DAAC-specific implementation or dataset characteristics

EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE

VEDA – Visualization, Exploration, and Data Analysis

https://www.earthdata.nasa.gov/dashboard/



EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE

A Simple Question

Do you want to increase your data usage and enhance aerosol and dust prediction?

Yes

Target User Types

- Environmental Monitoring
  - Environmental Quality
  - Impact on Ecosystem
- Public Health
  - Air Quality and Respiratory Health
  - Epidemiological Studies
- Climate Research
  - Climate Forcing
  - Global Climate Models

Data visualization of an April 5-8, 2022 dust event using DustTracker-AI, a machine learning model. It tracks dust into the night and is compared with NASA's CALIPSO satellite data.

Thank You!

Email Address:
jennifer.c.wei@nasa.gov