Jet Propulsion Laboratory
California Institute of Technology

podaac
Physical Oceanography Distributed Active Archive Center

EARTHDATA CLOUD

Data Management in the Petabyte Era - PO.DAAC Journey to the Cloud

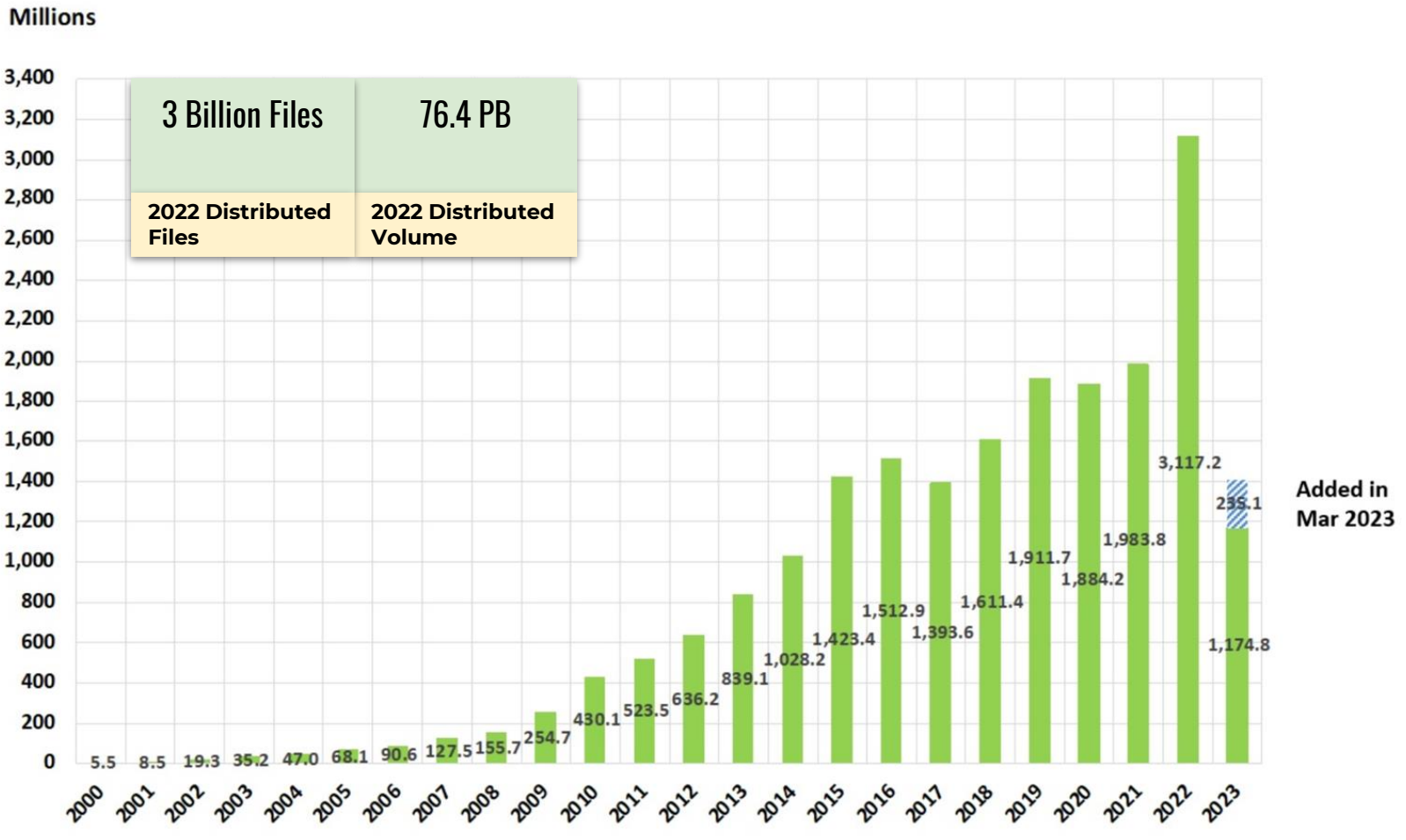
Suresh Vannan, PO.DAAC team

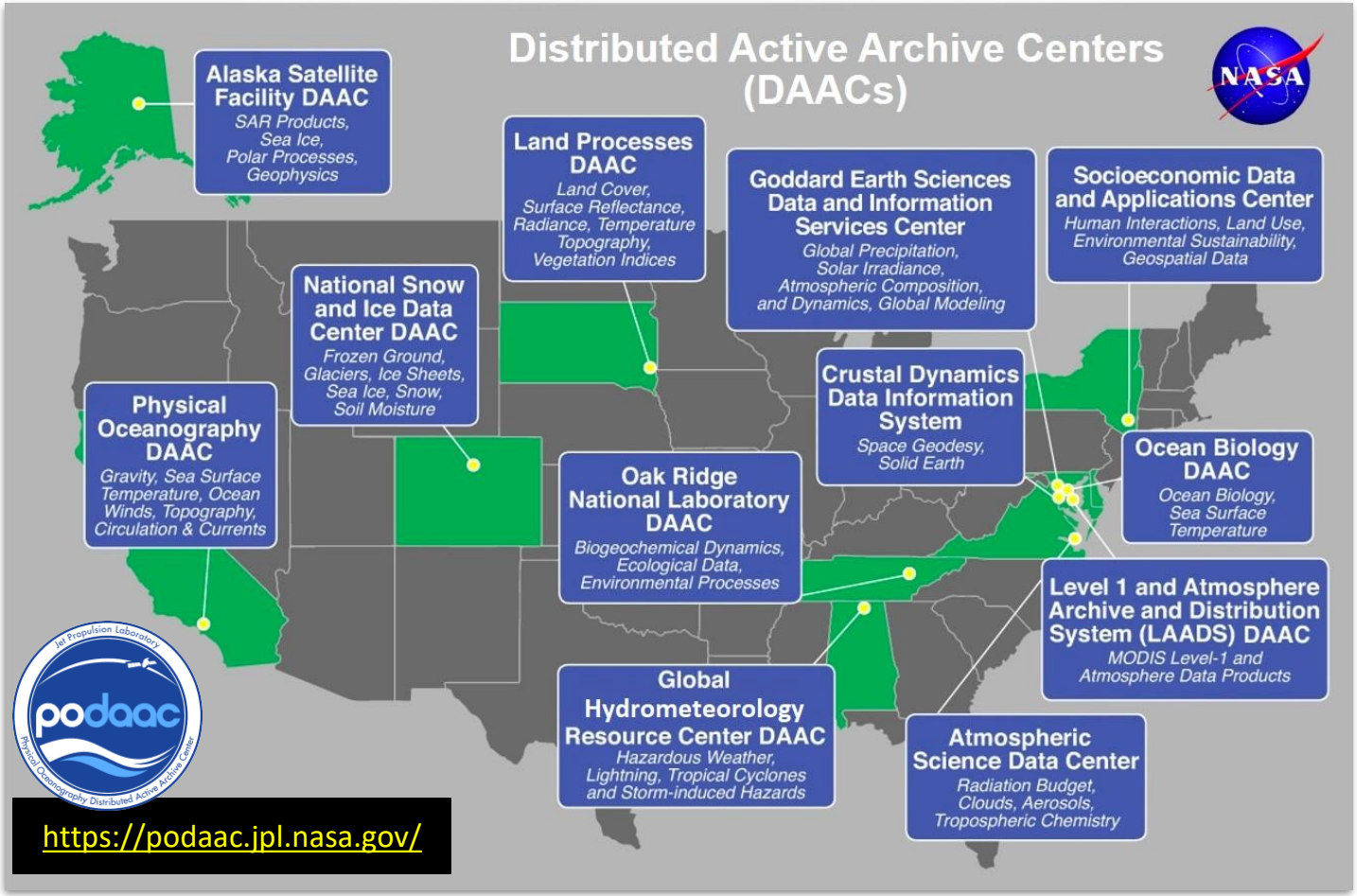
Jet Propulsion Laboratory, California Institute of Technology

May 2nd, 2023

PV2023, European Organization for Nuclear Research (CERN) Geneva (Switzerland)

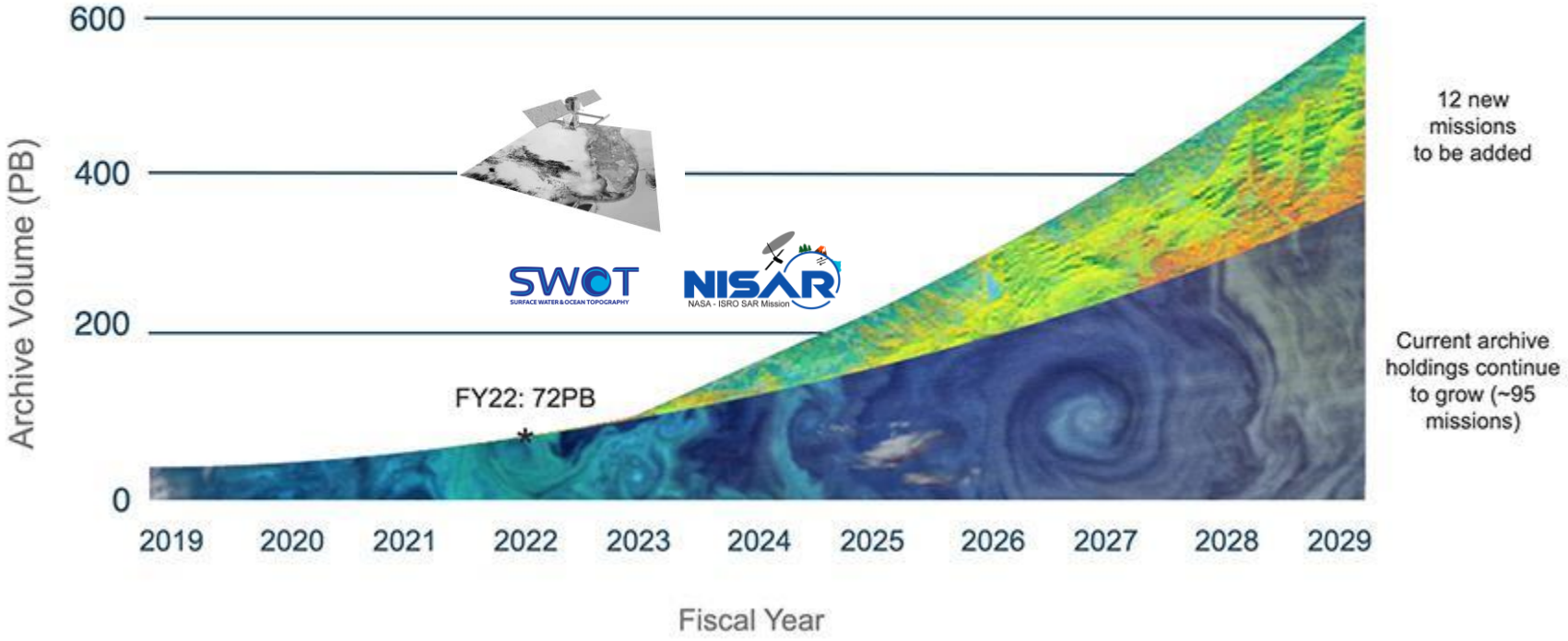

Landsat 9
ICESat-2
Aqua
ISS
Sentinel-6
Michael Freilich
CYGNSS-4
CALIPSO
Cloudsat
CYGNSS-7
GRACE-FO-1
GRACE-FO-2
CYGNSS-3
SMAP
Landsat 8
Aura
OCO-2
Suomi NPP
SWOT
Apr 20 2023 16:48





DAACs are custodians of EOS mission data and ensure that data will be easily accessible to users.

Earth Science Data Archive Growth Projection



A New Paradigm

The EOSDIS Cloud Evolution

Benefits of the Cloud

- **Easy access to data**: Data users will be able to access data directly in the cloud, making the need to download volumes of data unnecessary.
- **Rapid deployment**: Users can bring their algorithms and processing software to the cloud and work directly with the data in the cloud
- **Scalability**: The size and use of the archive can expand easily and rapidly as needed.
- **Flexibility**: Mission needs can dictate options for selecting operating systems, programming languages, databases, and other criteria to enable the best use of mission data.
- **Reduced Duplication**: The use of a common infrastructure with cloud native services will reduce redundant tools and services.

Challenges

Cost

- Storage
- Egress
- Development
- Computational
- Labor

Security

- Data protection
- Access control
- Cybersecurity

Migration

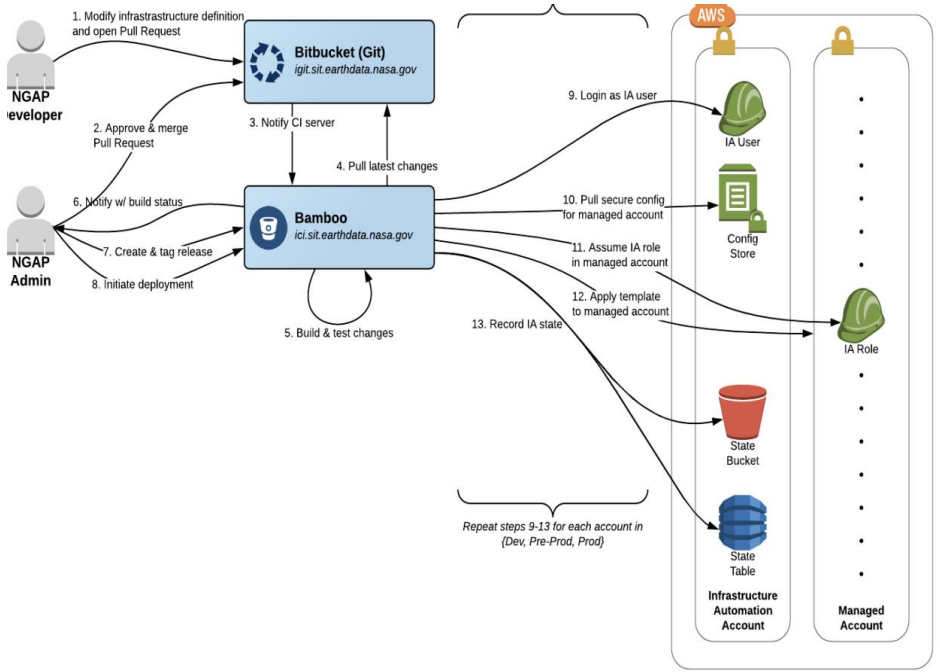
- Maintain existing system
- End-user
- Staffing
- Technical skill
- End-user migration

Earthdata Cloud Platform (NGAP - NASA General Application Platform)

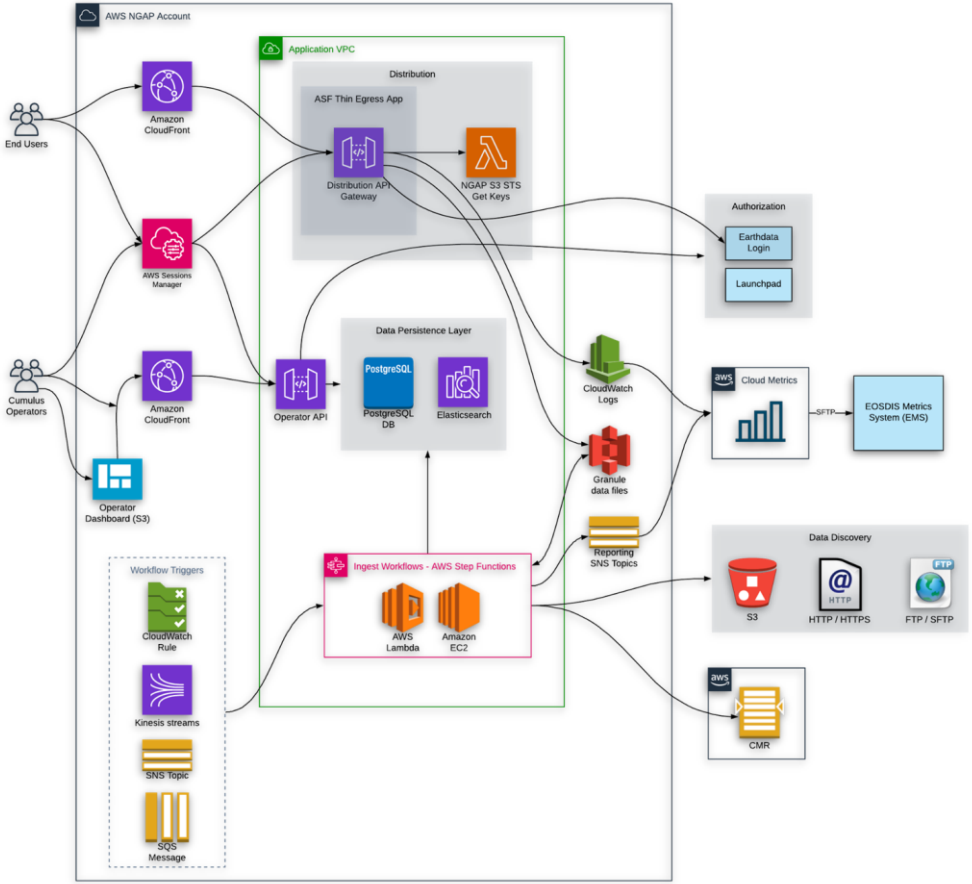
NGAP is a multi-account, Infrastructure-as-a-Service (IaaS) cloud platform operating on Amazon Web Services (AWS).

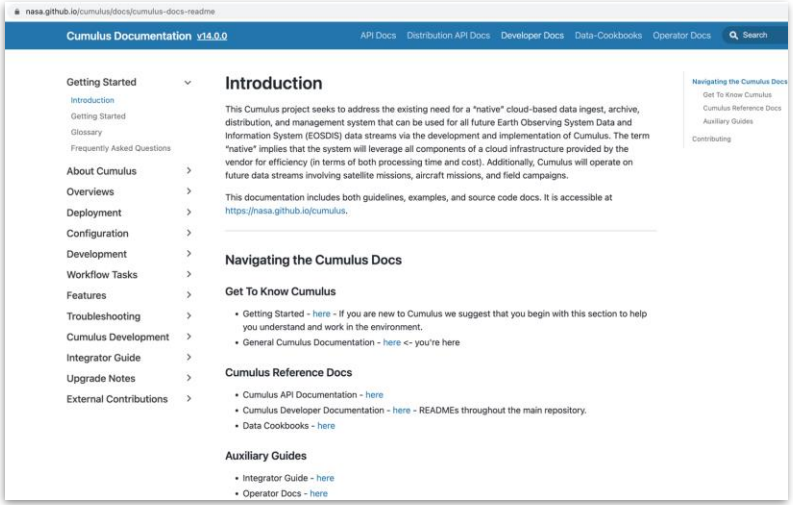
Features:

1. NASA-Approved Amazon Web Services (AWS)
2. Code Deployment Services: DevOps Pipeline
3. Use of Infrastructure as Code: including reusable template

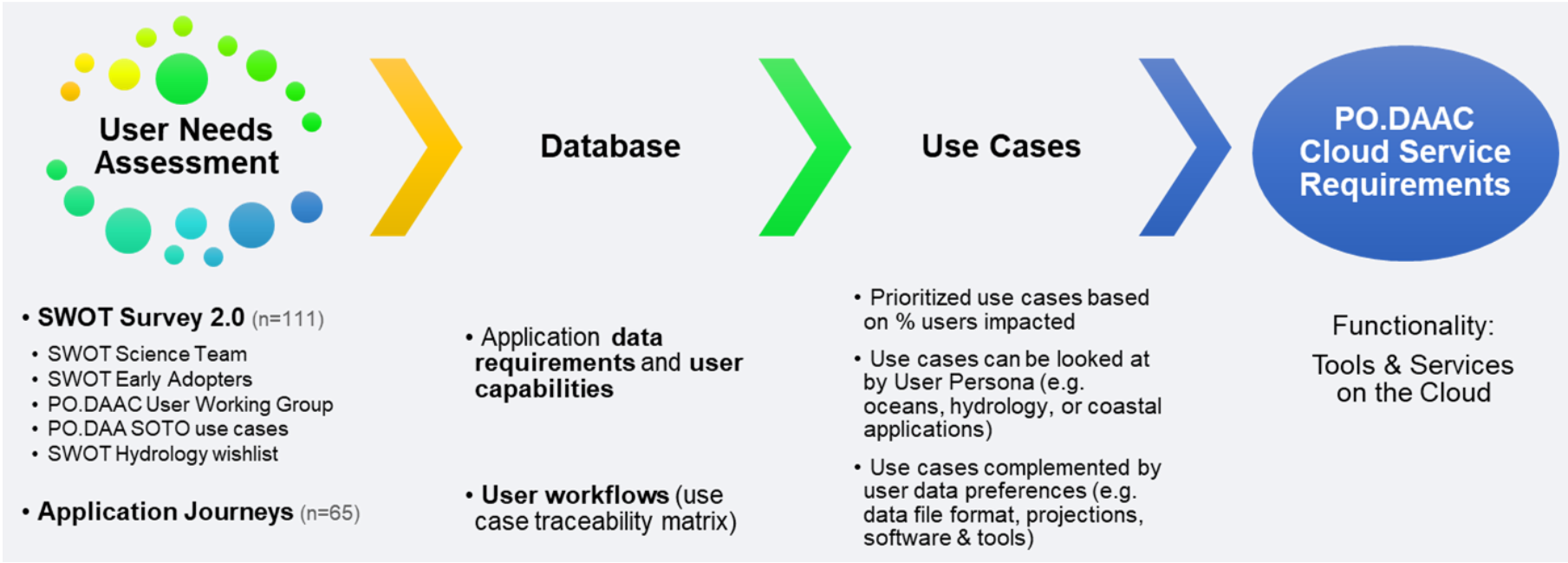


CUMULUS
EOSDIS NASA'S EARTH OBSERVING SYSTEM DATA AND INFORMATION SYSTEM



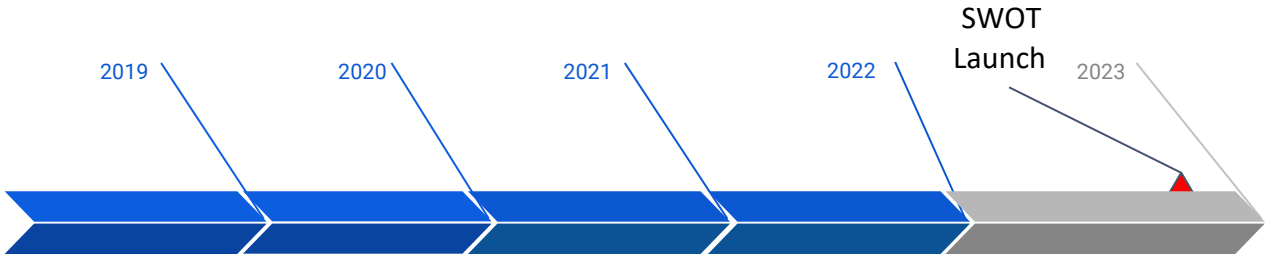


https://nasa.github.io/cumulus/docs/cumulus-docs-readme



E. N. Stavros, C. M. Oaida, J. Hausman and M. M. Gierach, "A Quantitative Framework to Inform Cloud Data System Architecture and Services Requirements Based on User Needs and Expected Demand," in *IEEE Access*, vol. 8, pp. 138088-138101, 2020, doi: 10.1109/ACCESS.2020.3012054.

Data Services Migration Timeline



DESIGN GOALS

Users will get the same level of service

Data download will continue to be freely available to users

Leverage the power of co-located data for processing large volumes of co-located datasets

Enable new frontiers in science/applications

End-User Capabilities Development

Search and Download

Dissemination & Transformation

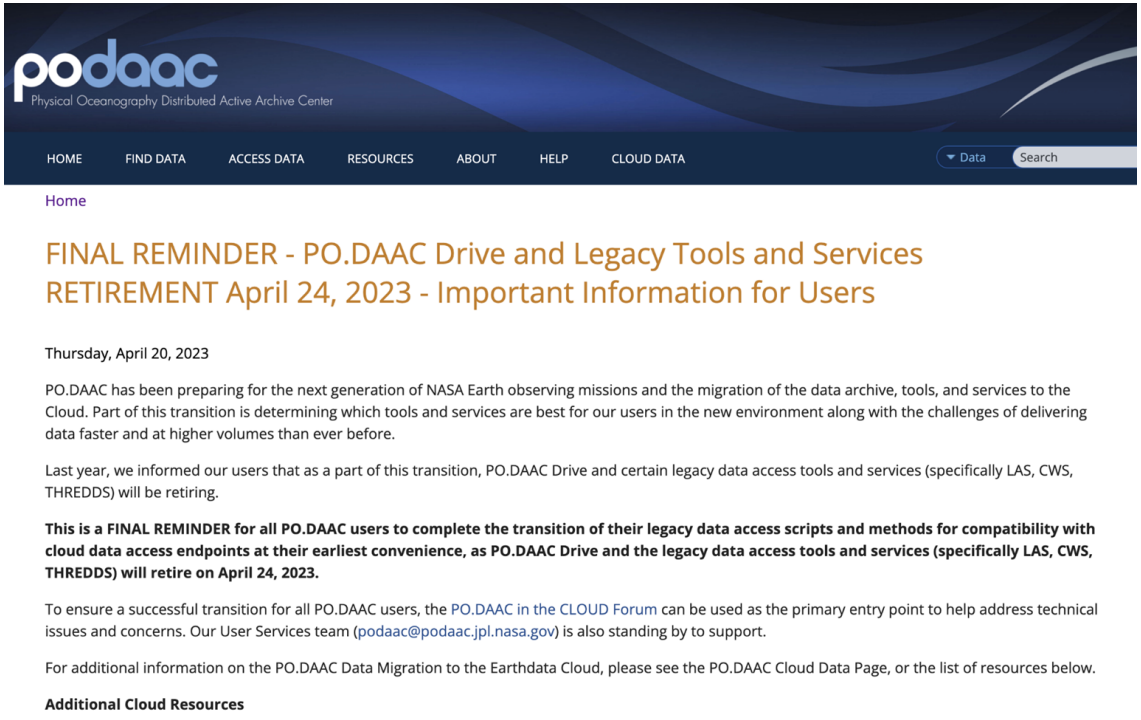
Analysis

Search and Download
- Feature based search
- Space/Time download

Dissemination & Transformation
- Subscriptions
- Subset
- Regrid , Reformat
- On-demand Raster Generation

Analysis
- Product space/time averaging
- Analysis-in-place (Cloud)
- Integration with other Datasets

Data Migration



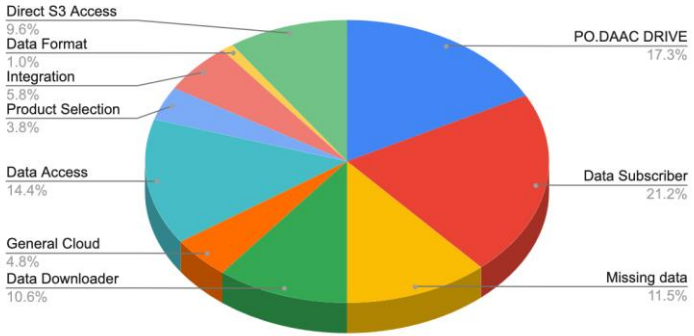
User Migration:

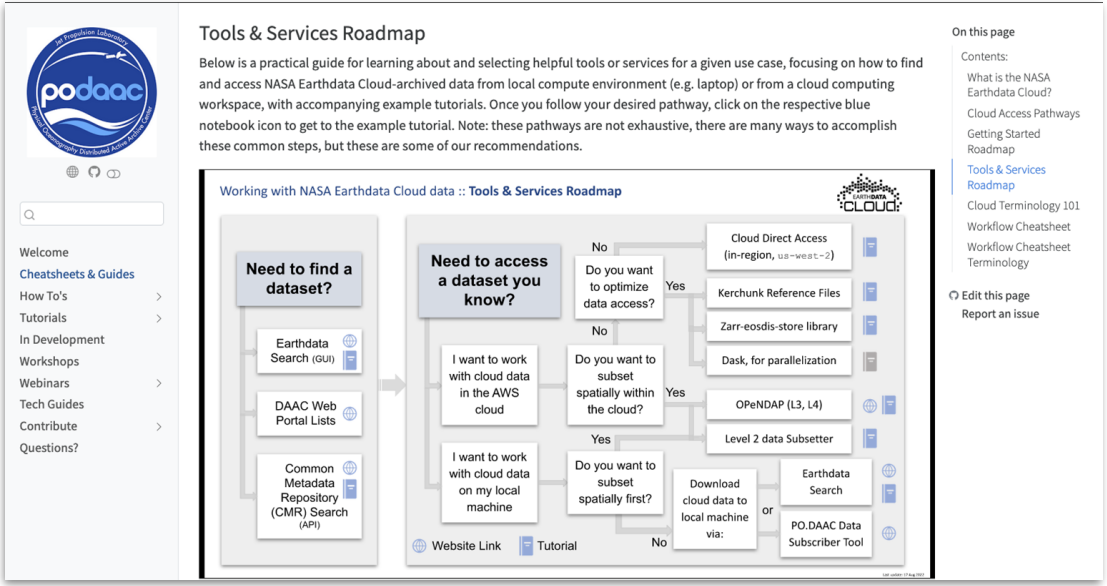
Training and Direct Outreach

- Workshops & Hackathons (Training) (**12***)
- Webinars (**3***)
- Present/participate at Science Team Meetings (**20***)
- Present/participate at Conferences (**15***)
- Tutorials/Notebooks (**40***)

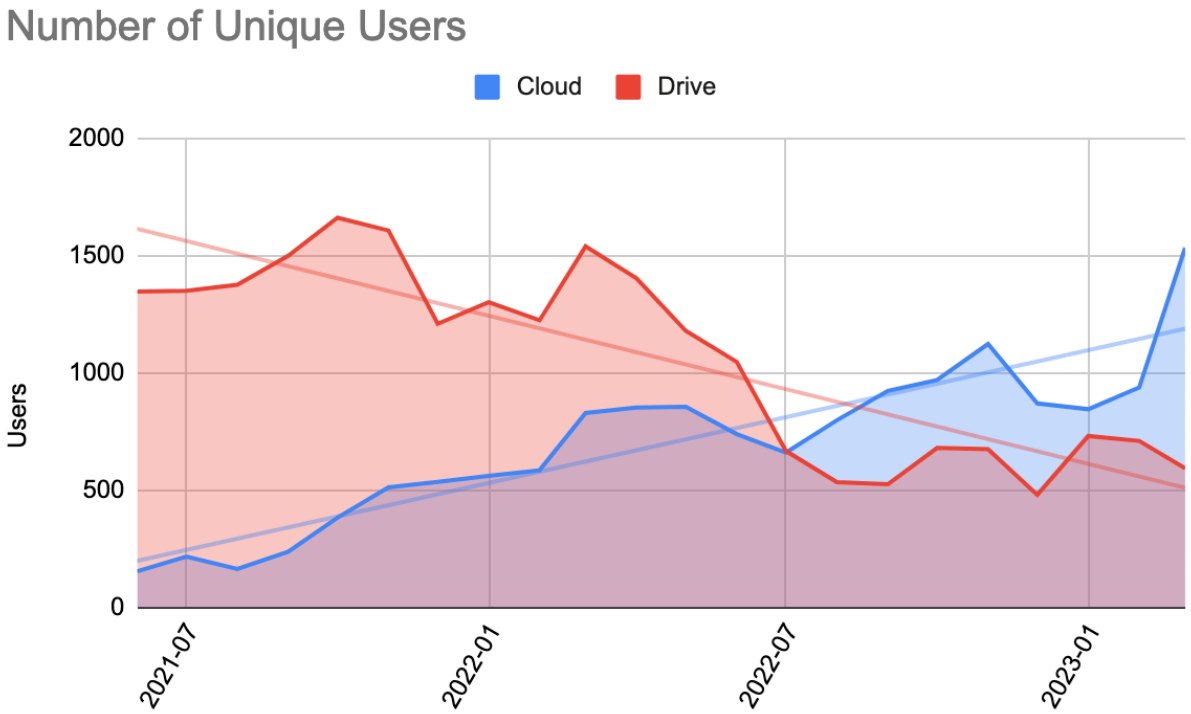
*estimate

End User Migration



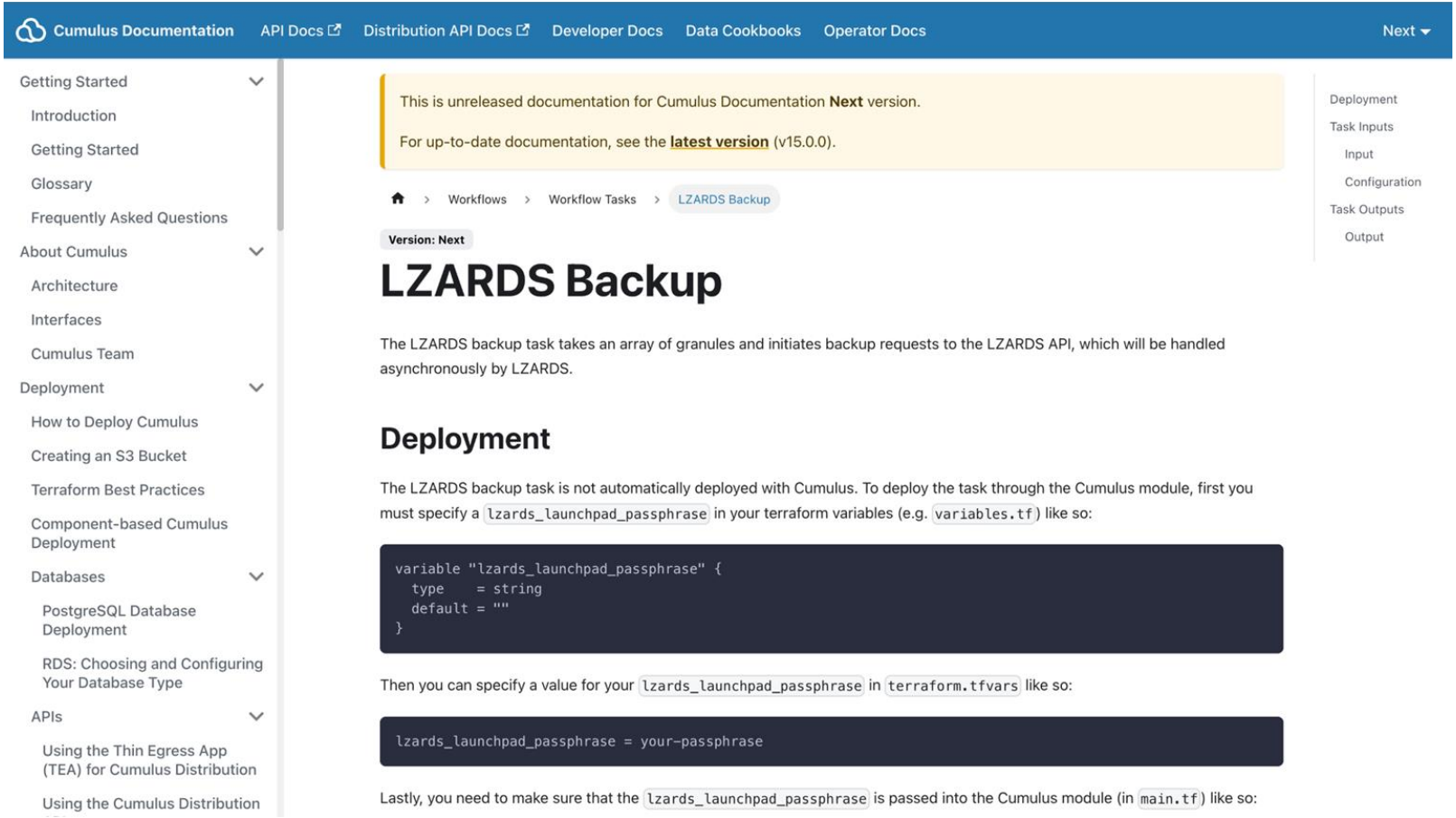


Measuring Success



- Since July of 2022, **PO.DAAC delivered data to more EarthData Cloud users than PO.DAAC Drive Users**
- Data Distribution and user adoption are trending in a positive direction

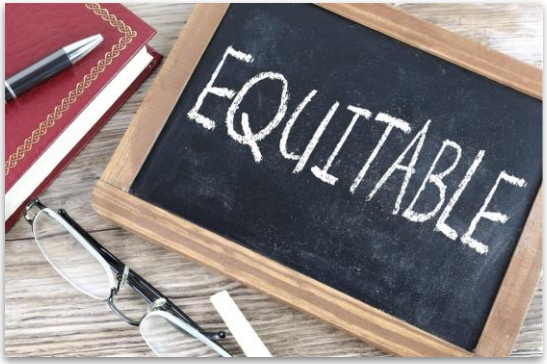
Data backup



We are not at the destination yet….

User Experience

Enable new Frontiers in Science



Open Data
Open Source Science
Equitable access to data