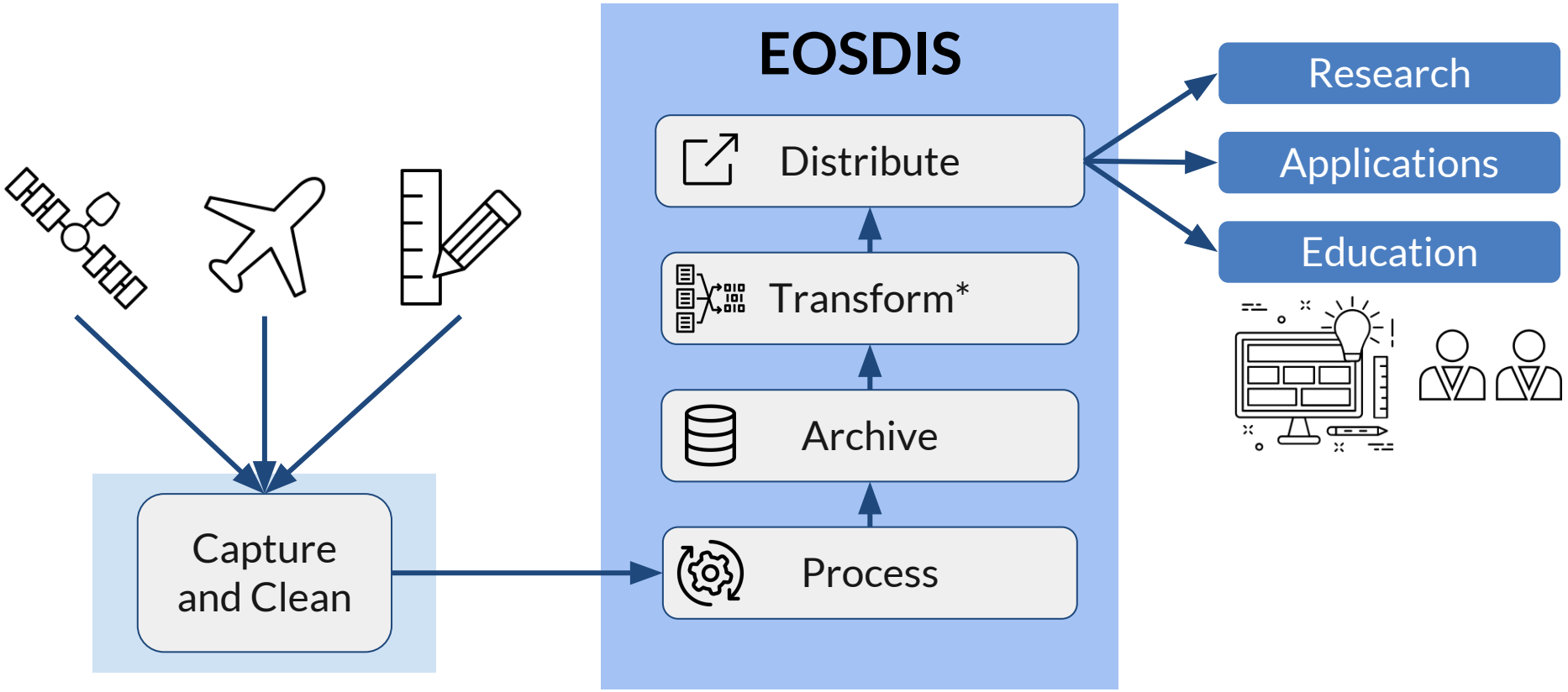
EOSDIS
NASA'S EARTH OBSERVING SYSTEM
DATA AND INFORMATION SYSTEM

Preparing for the Onslaught: NASA Earth Sciences Big Data Future

Katie Baynes, EOSDIS System Architect, GSFC

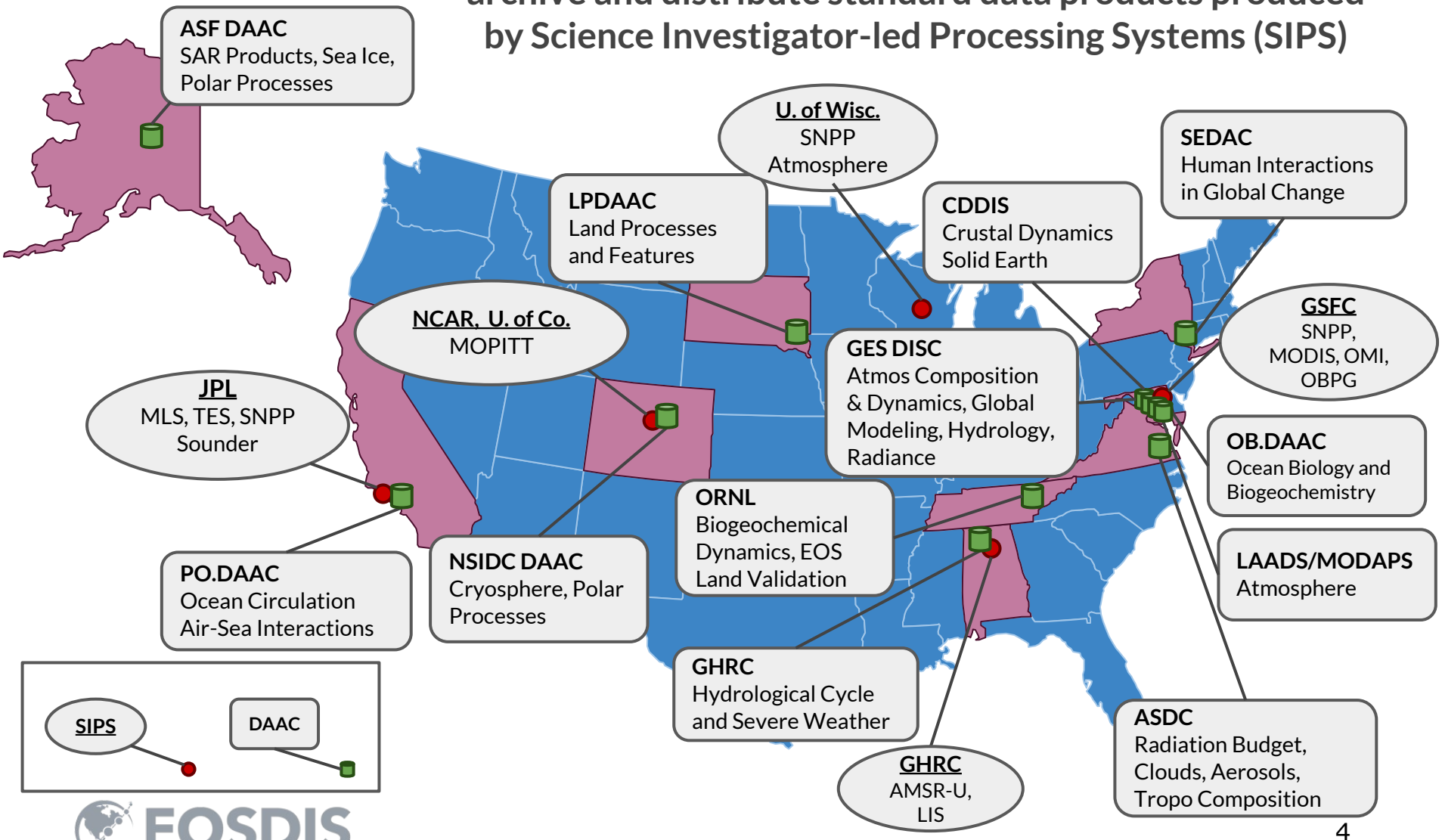
EOSDIS Overview

Putting EOSDIS in Context

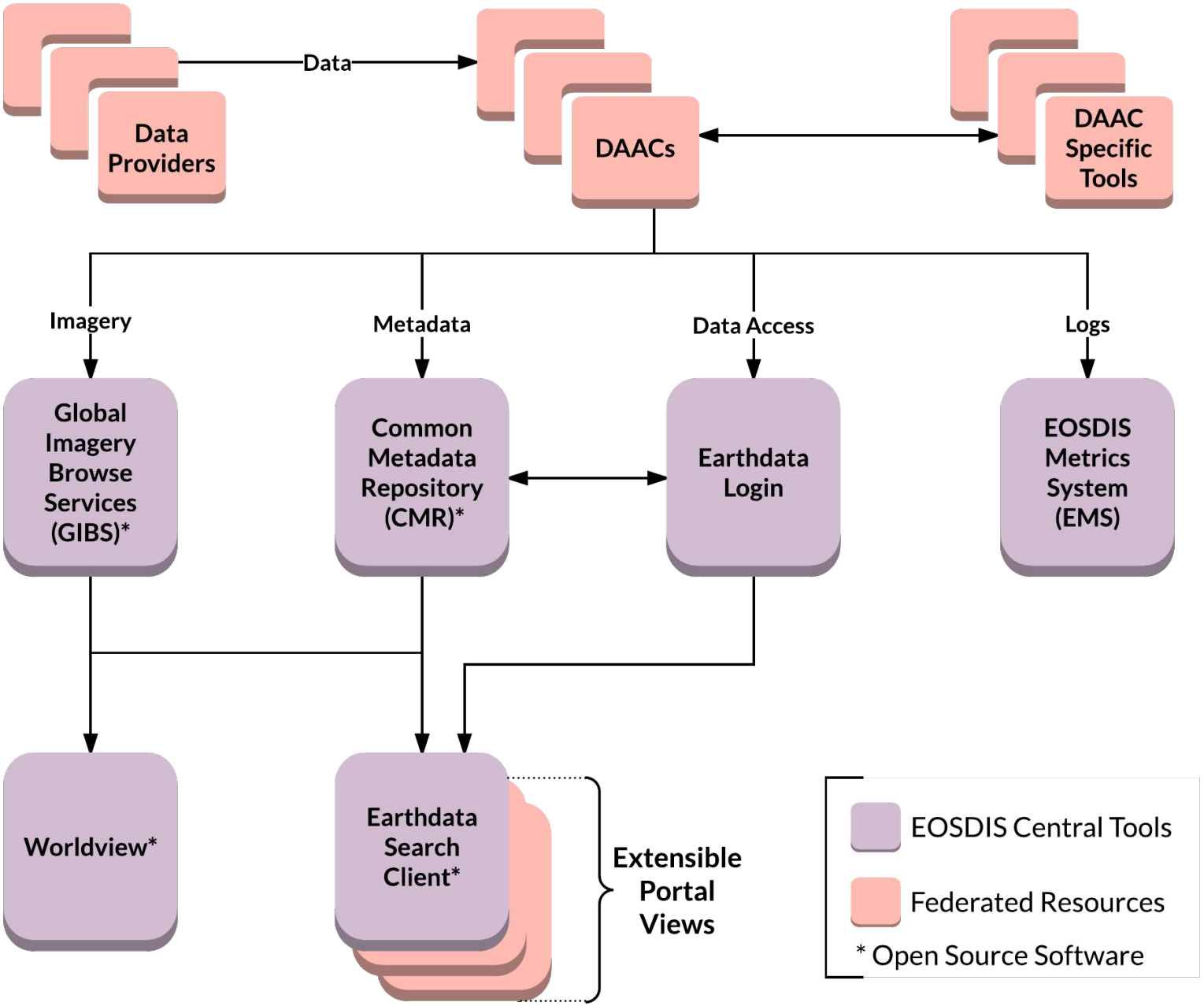


*Subset, reformat, reproject

Distributed Active Archive Centers (DAACs), collocated with centers of science discipline expertise, archive and distribute standard data products produced by Science Investigator-led Processing Systems (SIPS)

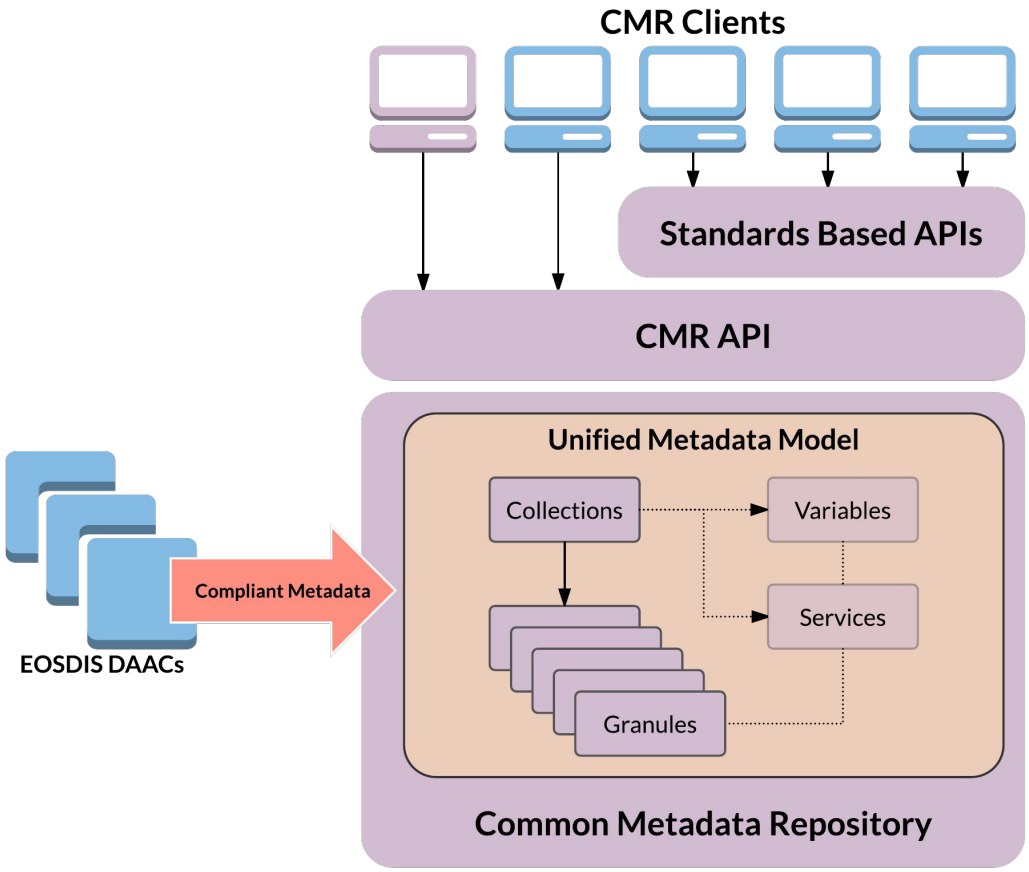


EOSDIS


Data Providers
Data
DAACs
DAAC Specific Tools
Imagery
Metadata
Data Access
Logs
Global Imagery Browse Services (GIBS)*
Common Metadata Repository (CMR)*
Earthdata Login
EOSDIS Metrics System (EMS)
Worldview*
Earthdata Search Client*
Extensible Portal Views
EOSDIS Central Tools
Federated Resources
* Open Source Software

EOSDIS

Common Metadata Repository



Lightning fast, always available

- *95% queries complete in <1s*
- *99.98% uptime (last 365d)*

Big Data Ready

- *34K collections*
- *367 million files indexed*
- *Prepared to scale 1B+ records*

Standards-focused

- *ISO-19115 metadata*
- *OpenSearch/OGC CSW*
- *REST based APIs*

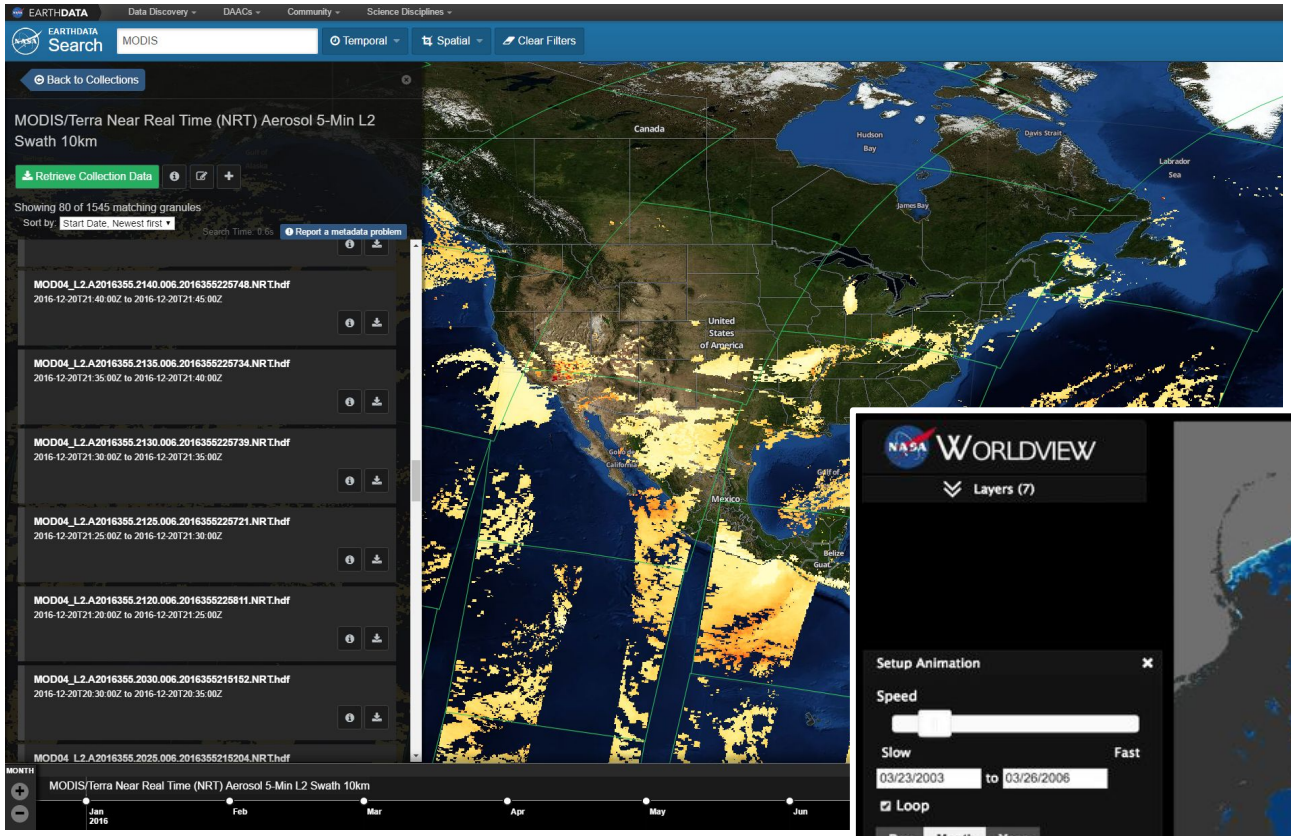
Community-focused

- *Developer's portal*
- *Active Developer's forum*
- *Ecosystem of supported tools*
- *Open Source codebase in github*

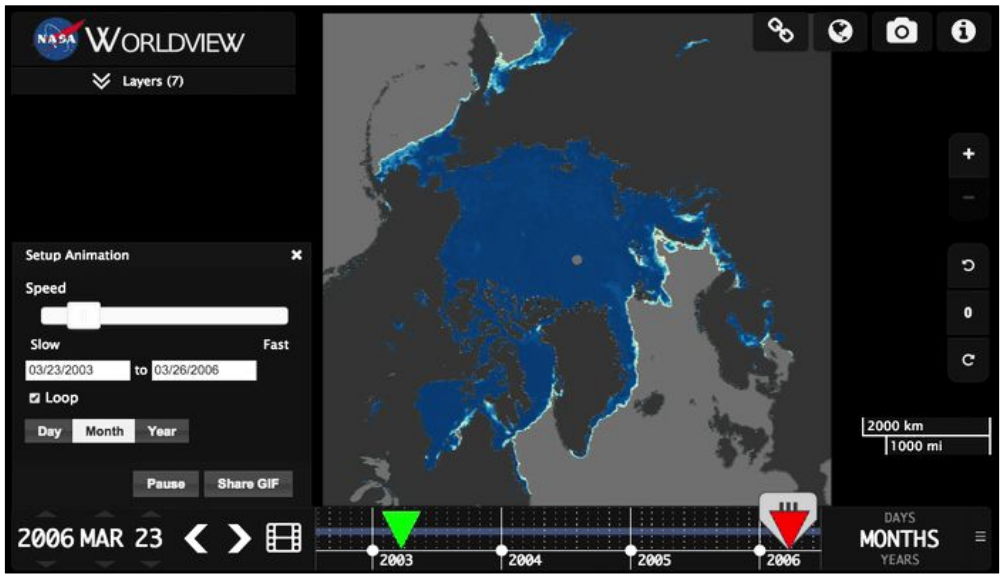
Internationally Recognized

- *Provides the backbone of the Community of Earth Observing Satellites International Directory Network (CEOS IDN)*

EOSDIS



Data Centric End Users
https://search.earthdata.nasa.gov



Imagery Centric End Users
https://worldview.earthdata.nasa.gov

EOSDIS

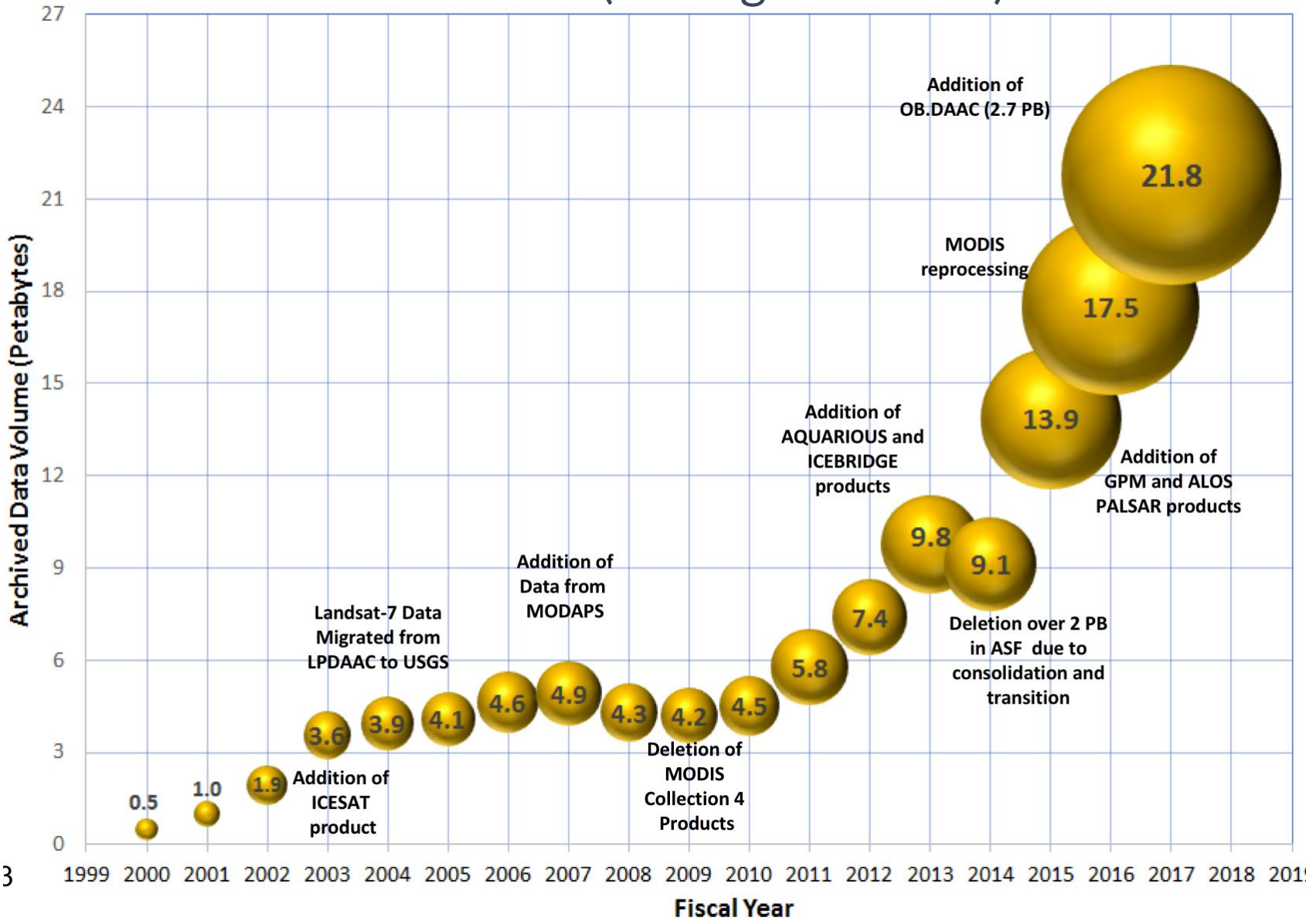
Mission Phase Lifecycle Events

NASA Life-Cycle Phases	FORMULATION			IMPLEMENTATION			
Project Life-Cycle Phases	Pre-Phase A Concept Studies	Phase A Concept & Technology Development	Phase B Preliminary Design & technology Completion	Phase C Final Design & Fabrication	Phase D System Assembly, Integration & Test, Launch & Checkout	Phase E Operations & Sustainment	Phase F Closeout
Key Decision Points (KDP) & Major Reviews	MCR (A)	SRR (B)	PDR (C)	CDR (D)	ORR (E)	DR (F)	
Major Flight Project Events		Baseline Requirements	Baseline Project Plan	Baseline Interface Control Documents	Launch	End of Mission	Final Archival of Data
Major Science/ Data Management Events	DAAC Assignment → DAAC and Flight Project responsibilities → Estimated data volumes, latency requirements → Key Milestones	L1/PLRA Requirements DAAC Assignment ESDIS-Flight Project Inter-Project Agreement	Algorithm Theoretical Basis Documents → Delivery mechanism → Data formats → Product lists/volumes → Network Bandwidth/Latency → Network Diagrams	Interface Control Documents (DAAC & Network)	Baseline Science Data Management Plan Operations Agreements Support End-to-end tests	On-going Processing, Archiving and Distribution	Implement Preservation Content Specification

MCR – Mission Concept Review; SRR – System Requirements Review; PDR – Preliminary Design Review; CDR – Critical Design Review; ORR – Operational Readiness Review; DR – Decommissioning review; PLRA - Program Level Requirements Appendix

The Data Onslaught

Total EOSDIS Data Archive Volume (Petabytes) 2000-2017 (through 1/31/17)



Mission Volumes - Previous 5 years

Missions	Launch Date	Daily Data Volume	DAAC(s)
S-NPP	Oct 28, 2011	1,700 GB/day	GES DISC, OB, LP, NSIDC, LAADS, ASDC
Shizuku (GCOM-W1)	May 18, 2012	8 GB/day	GHRC / NSIDC
GPM	Feb 27, 2014	410 MB/day	GES DISC
Sentinel-1A	Apr 3, 2014	1,800 GB/day	ASF
OCO-2	Jul 2, 2014	82 GB/day	GES DISC
RapidScat on ISS	Sept 20, 2014	3.3 GB/day	PO
CATS on ISS	Jan 10, 2015	14 GB/day	ASDC
SMAP	Jan 31, 2015	350 GB/day	ASF / NSIDC
DSCOVR	Feb 11, 2015	859 MB/day	ASDC
Sentinel-3A	Feb 16, 2016	939 GB/day	OB, LAADS
Sentinel-1B	April 22, 2016	1,800 GB/day	ASF
CYGNSS (EV-M)	Oct 28, 2016	9,872 MB/day	PO
SAGE-III on ISS	Feb 19, 2017	174 MB/day	ASDC
LIS on ISS	Feb 19, 2017	100 MB/day	GHRC
TOTAL		**6.7 TB/day**	

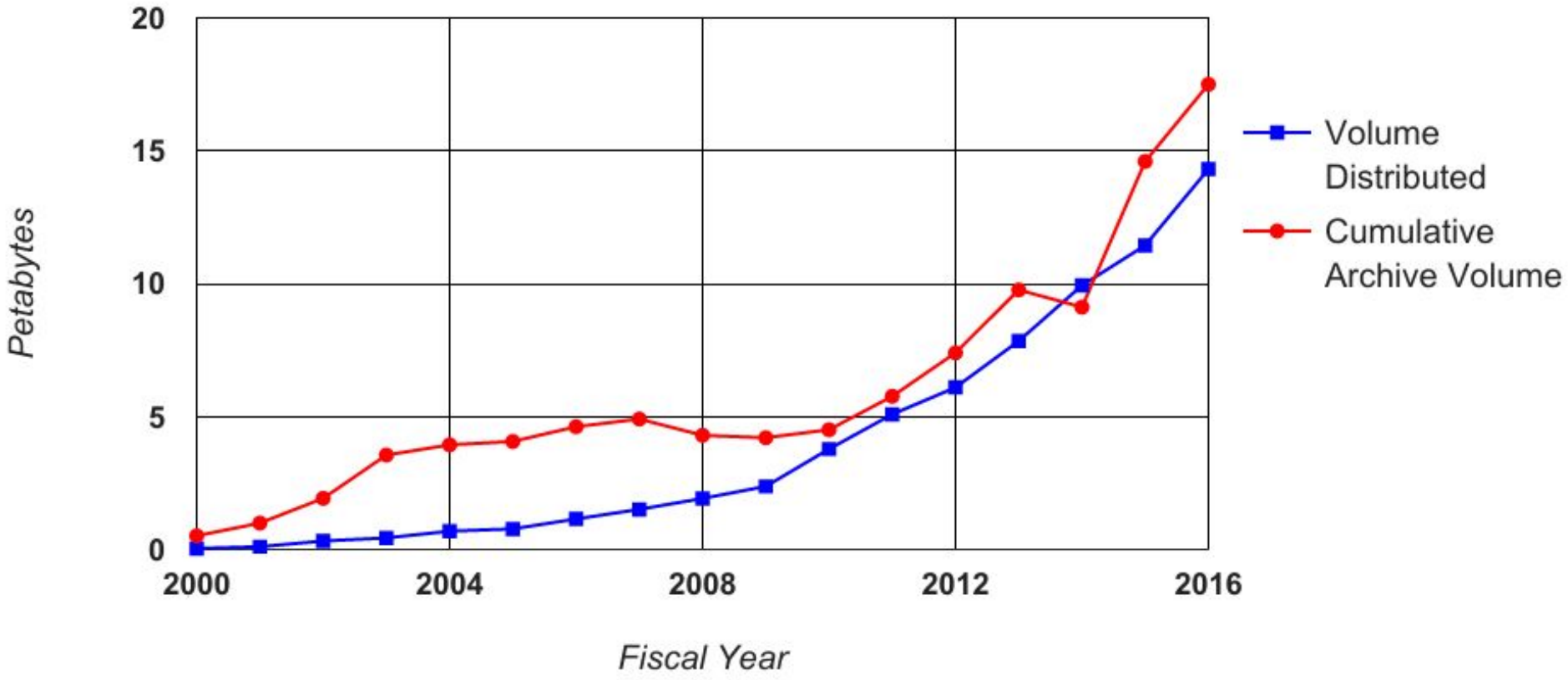
EOSDIS

Mission Volumes - Next 5 years

Missions	Launch Date	Daily Data Volume	DAAC(s)
Sentinel-5P	June 2017	626 GB/day	ASF
JPSS-1	July-Sept 2017	1,700 GB/day	GES DISC, OB, LP, NSIDC, LAADS, ASDC
Sentinel-3B	Nov 2017	128 GB/day	OB, LAADS
GRACE FO	Feb 2018	19 MB/day	PO
TSIS on ISS	Apr 2018	541 MB/day	GES DISC
ICESat-2	Sept 2018	891 GB/day	NSIDC
ECOSTRESS	2019	585 GB/day	LP
GEDI on ISS	Mar 2019	3.5 GB/day	LP
SWOT	April 2021	15.5 TB/day	PO
TEMPO	2021	1.7 TB/day	ASDC
NISAR	Dec 2021	86 TB/day	ASF
JPSS-2	Oct 2021-Feb 2022	1,700 GB/day	GES DISC, OB, LP, NSIDC, LAADS, ASDC
PACE	Jun 2022	3.5 TB/day	OB
TOTAL		**114 TB/day**	

EOSDIS

Data Coming In vs. Data Going Out

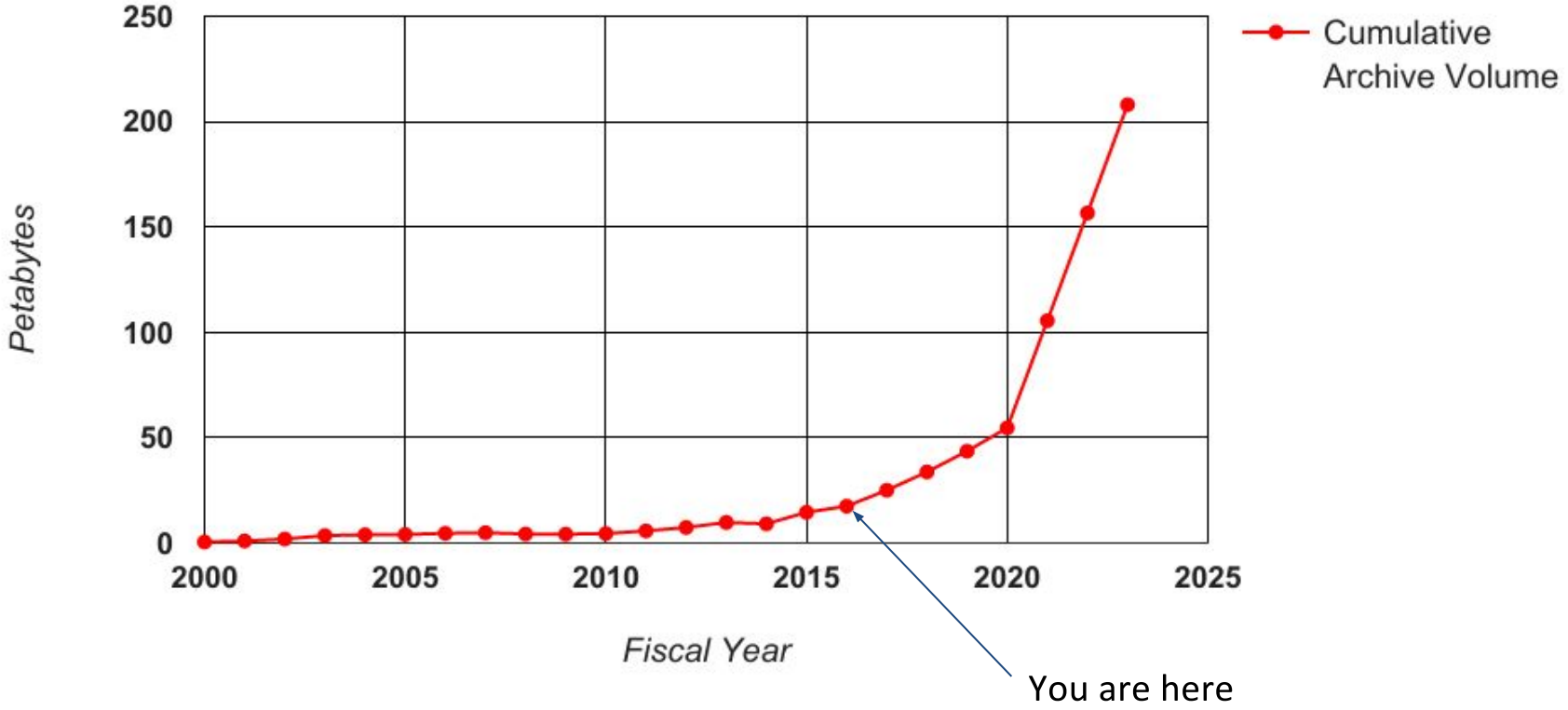


EOSDIS

Asking Big Questions

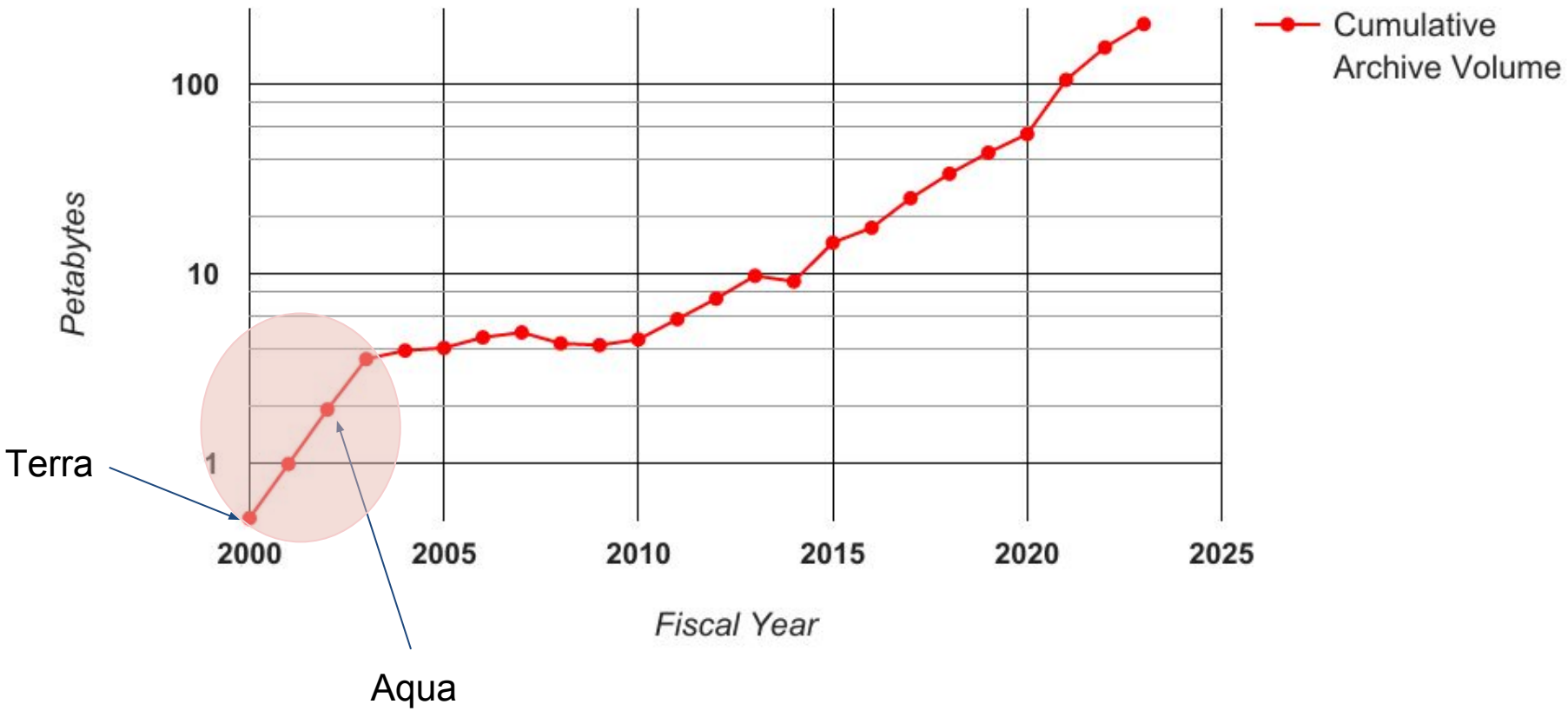
- Where is EOSDIS today?
- Where does EOSDIS need to be in 5 years?
- How will missions with dramatically greater data volumes be handled?

So Our Archive Slated to Grow Substantially…



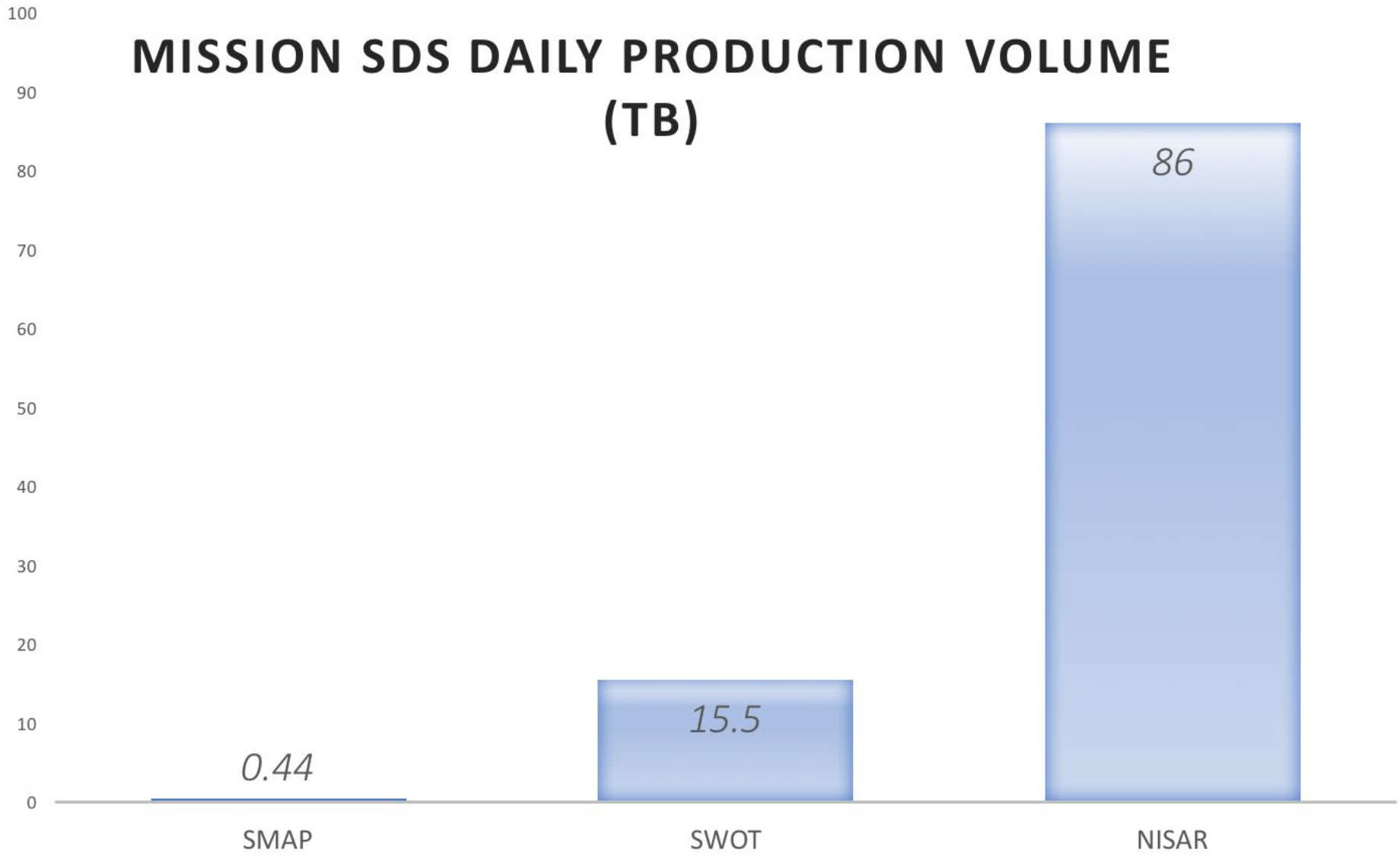
EOSDIS

...But Not Our First Rodeo



EOSDIS

Changing Data Landscape



Looking towards Cloud

Conceptual "Data Close to Compute"

Large Volume Data Storage

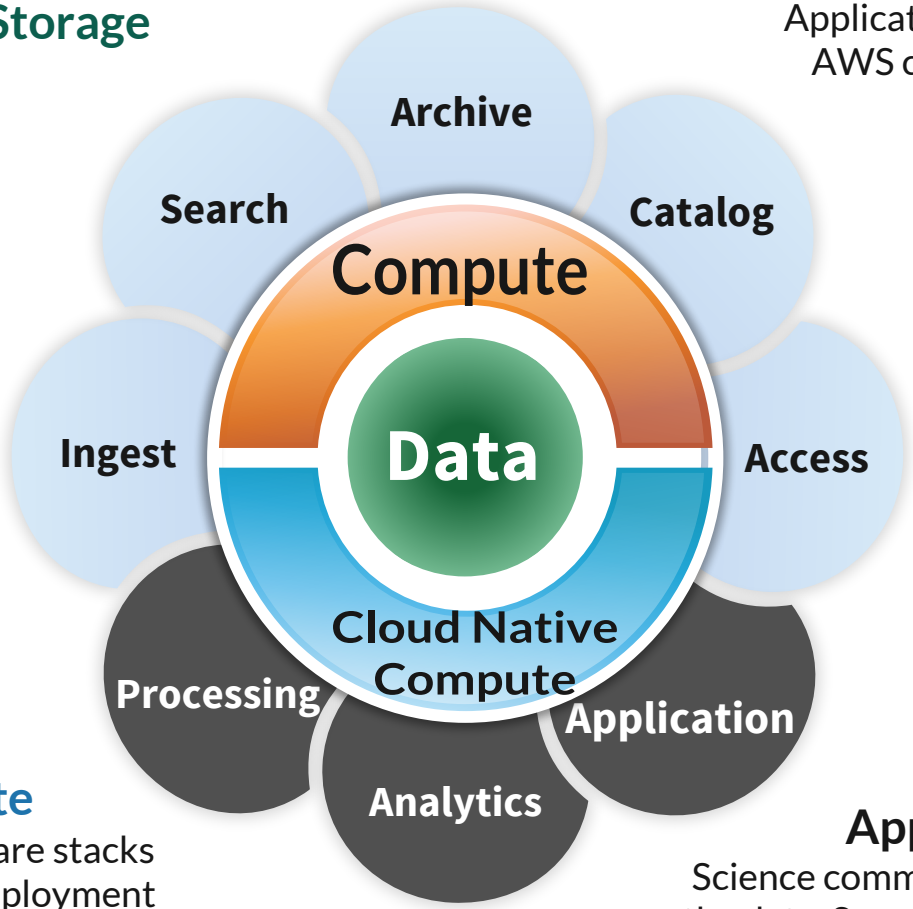
Centralized mission observation & model datasets stored in auto graduated AWS object storage (S3, S3-IA, Glacier)

Scalable Compute

Provision, Access, and terminate dynamically based on need. Cost by use

Cloud Native Compute

Cloud vendor service software stacks and microservices easing deployment of user based applications



EOSDIS Applications & Services

Application and service layer using AWS compute, storage (S3, S3IA, Glacier), and cloud native technologies

Non-ESDIS / Public Applications & Services

Science community brings algorithms to the data. Support for NASA & non-NASA

EOSDIS

Cloud Benefits for Data Systems (EOSDIS)

Cost-Effective

Cost-Effective: Only pay for the compute and storage needed. Easy to separate and fund by project/function.

Scalable Performance

Scalable Performance: Data Close to compute, auto-scaling, elastic load balancing, scale up or down based on need.

Flexibility

Flexibility: Maximum flexibility selecting operating systems, compute (CPU), programming languages, databases, and more based on mission need.

Deployment Speed

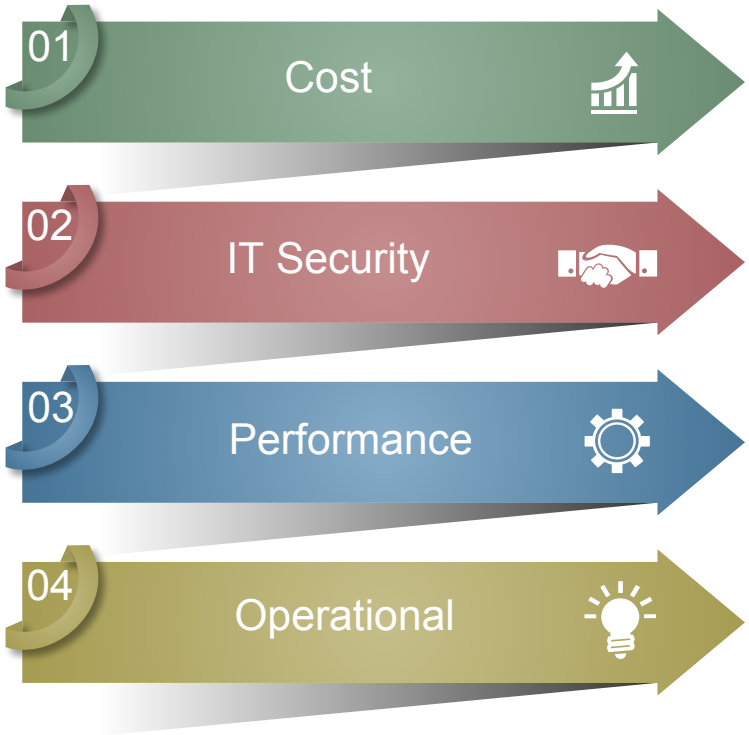
Deployment Speed: With an established cloud platform development is faster, access; to compute, storage, and IT Services requires only funding. Significantly simplifying procurement and provisioning.

Accessibility

Accessibility: Centralized, redundant, enterprise level holdings in the cloud allows for "additional" more effective ways of accessing extremely large datasets of new missions.

Decision Considerations

High level decision considerations for individual project prototypes and capabilities to operationalize into AWS (commercial cloud)



Is AWS (commercial cloud) affordable?.

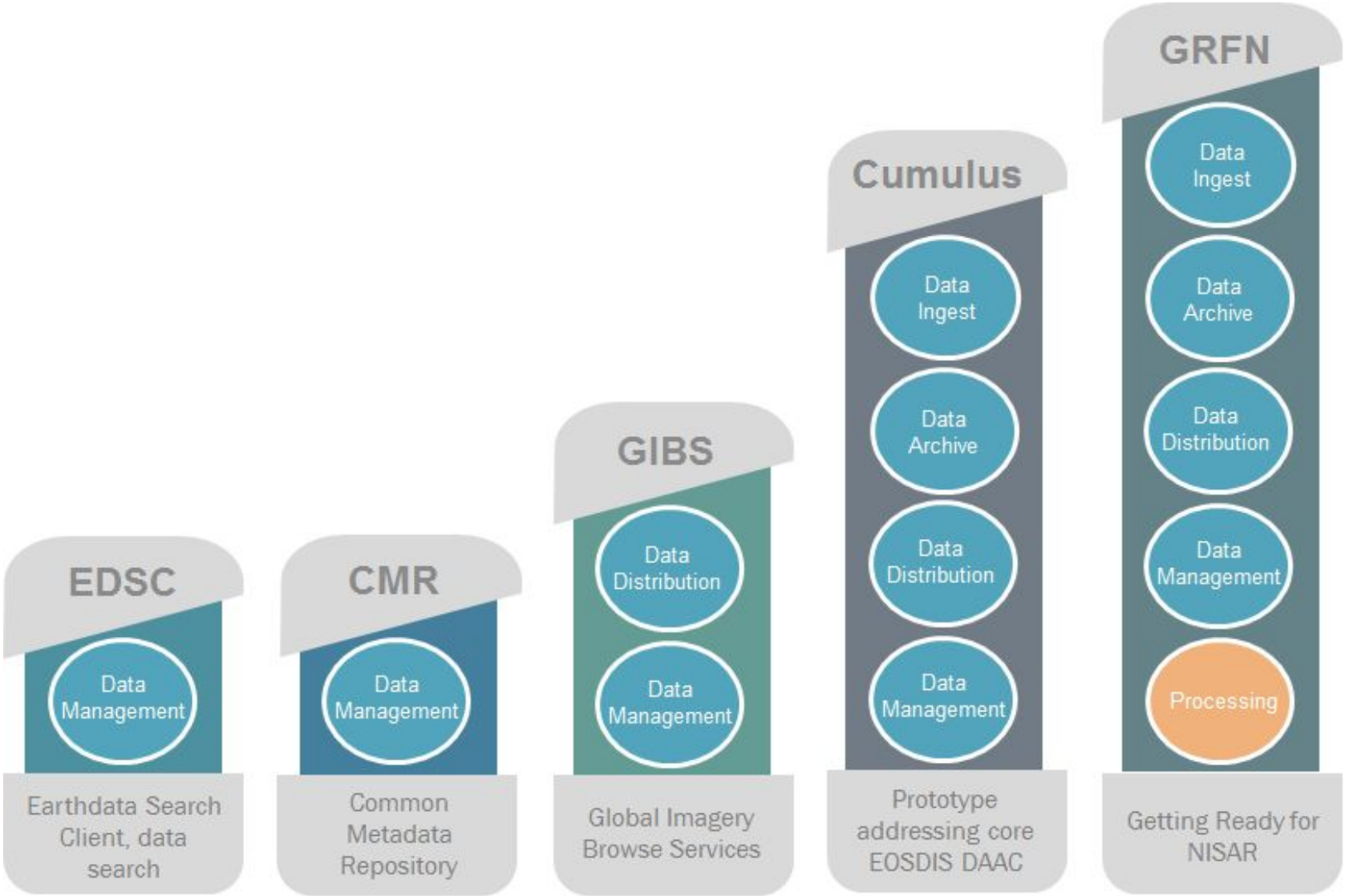
Is NASA IT Security compliance and tactical operations achievable in AWS (commercial cloud)?

Is performance equal to or better than current on-premises solutions?

Can we operate “Operationally” in AWS (commercial cloud), technical and business?

EOSDIS

Technical Foundational Capability Mapping

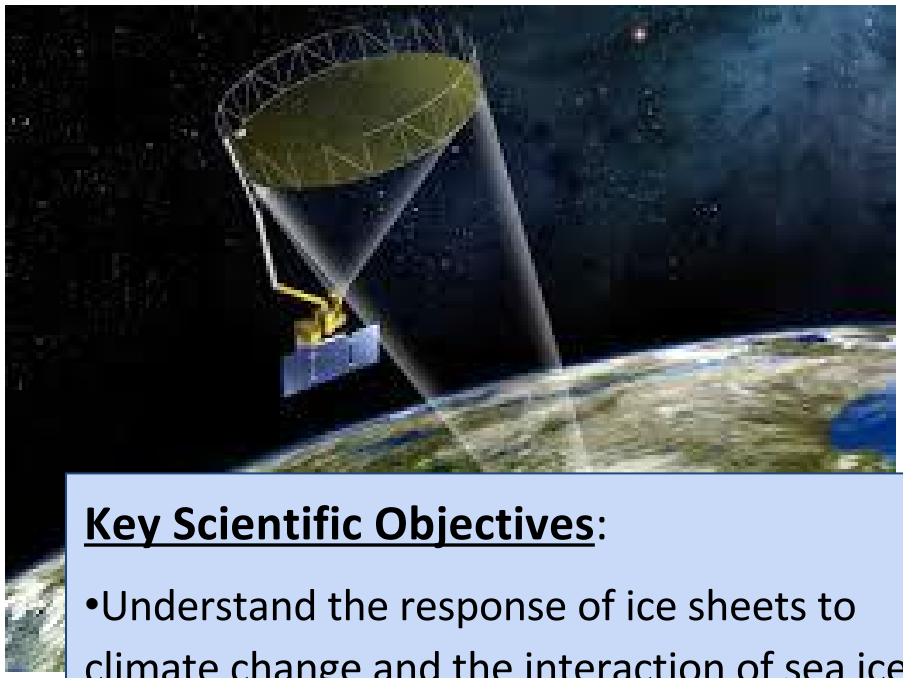


EOSDIS

GRFN: Getting Ready for NISAR

EOSDIS

NISAR Quick Facts



Key Scientific Objectives:

- Understand the response of ice sheets to climate change and the interaction of sea ice and climate
- Understand the dynamics of carbon storage and uptake in wooded, agricultural, wetland, and permafrost systems
- Determine the likelihood of earthquakes, volcanic eruptions, and landslides

*"The **NASA-ISRO Synthetic Aperture Radar** (**NISAR**) mission is a joint project between NASA and ISRO to co-develop and launch a dual frequency synthetic aperture radar satellite. The satellite will be the first radar imaging satellite to use dual frequency and it is planned to be used for remote sensing to observe and understand natural processes of the Earth."*

https://en.wikipedia.org/wiki/NISAR_(satellite)

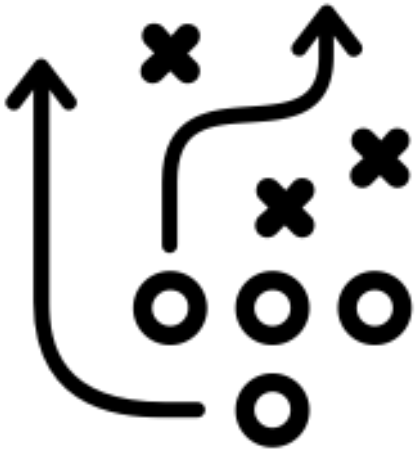
Payload:
L-band (24-centimeter wavelength)
polarimetric SAR (NASA)
S-band (12-centimeter wavelength)
polarimetric SAR (ISRO)

Launch: 2021-ish from India

GRFN Project Overview

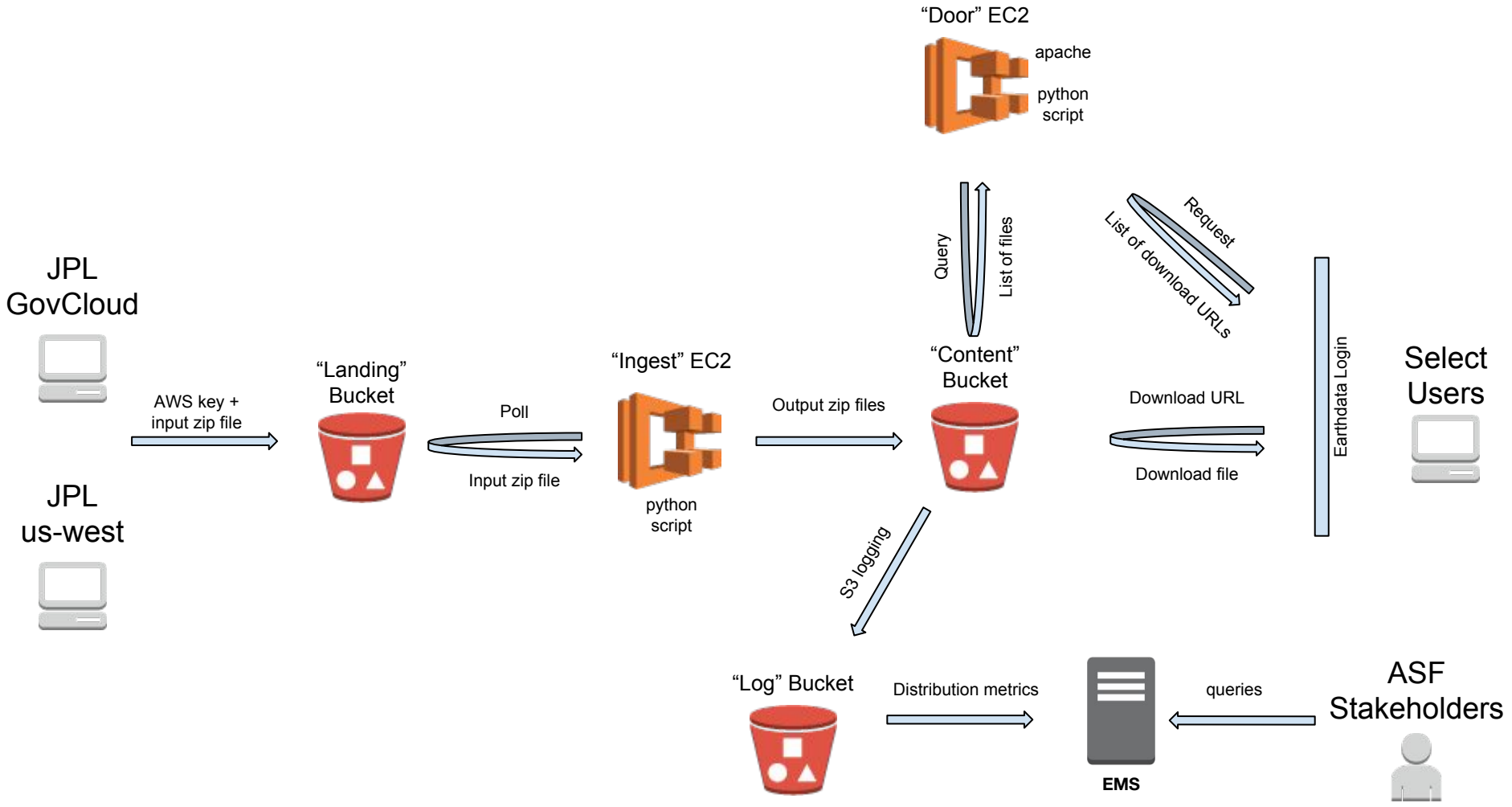
The GRFN project will aid in preparing for the large data volumes expected from the NISAR mission using ESA's Sentinel-1 data as a NISAR surrogate.

GRFN Technical Approach

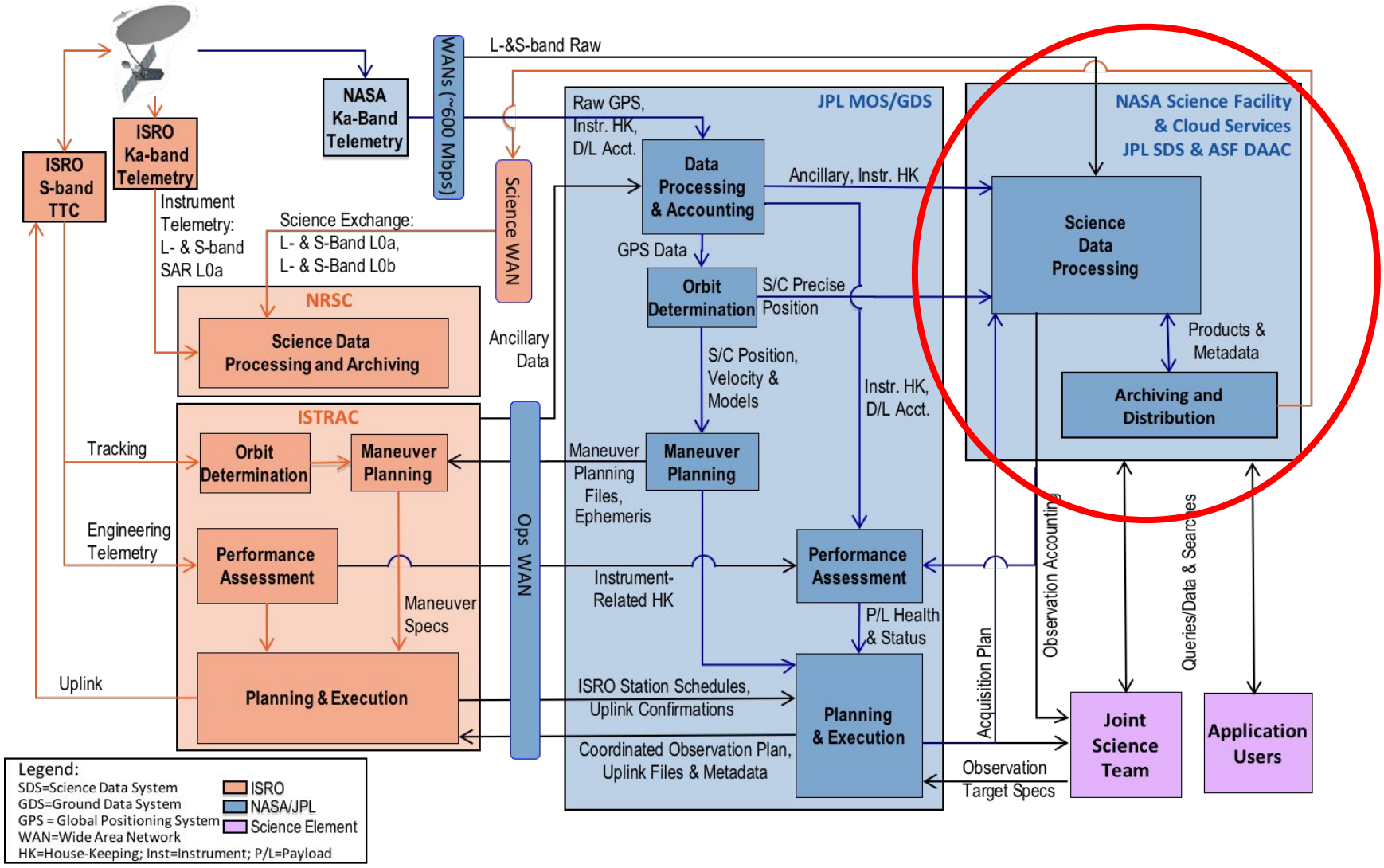


- Leveraging AWS Cloud
- JPL SDS and ASF DAAC working collaboratively to understand and react to NISAR impacts
- **Investigating seamless data delivery, bulk reprocessing scenarios, and on-demand processing**
- Engaging and **encouraging science community** to take advantage of cloud-based compute capabilities
- Engaging SWOT DAAC as appropriate to provide lessons-learned and guidance
- Proving out **use cases and cost models** and discovering **pain points** prior to launch

GRFN Short Term Architecture



NISAR Mission System Functional Elements and Interfaces



EOSDIS

CUMULUS

Prototyping DAACs in the Cloud

EOSDIS

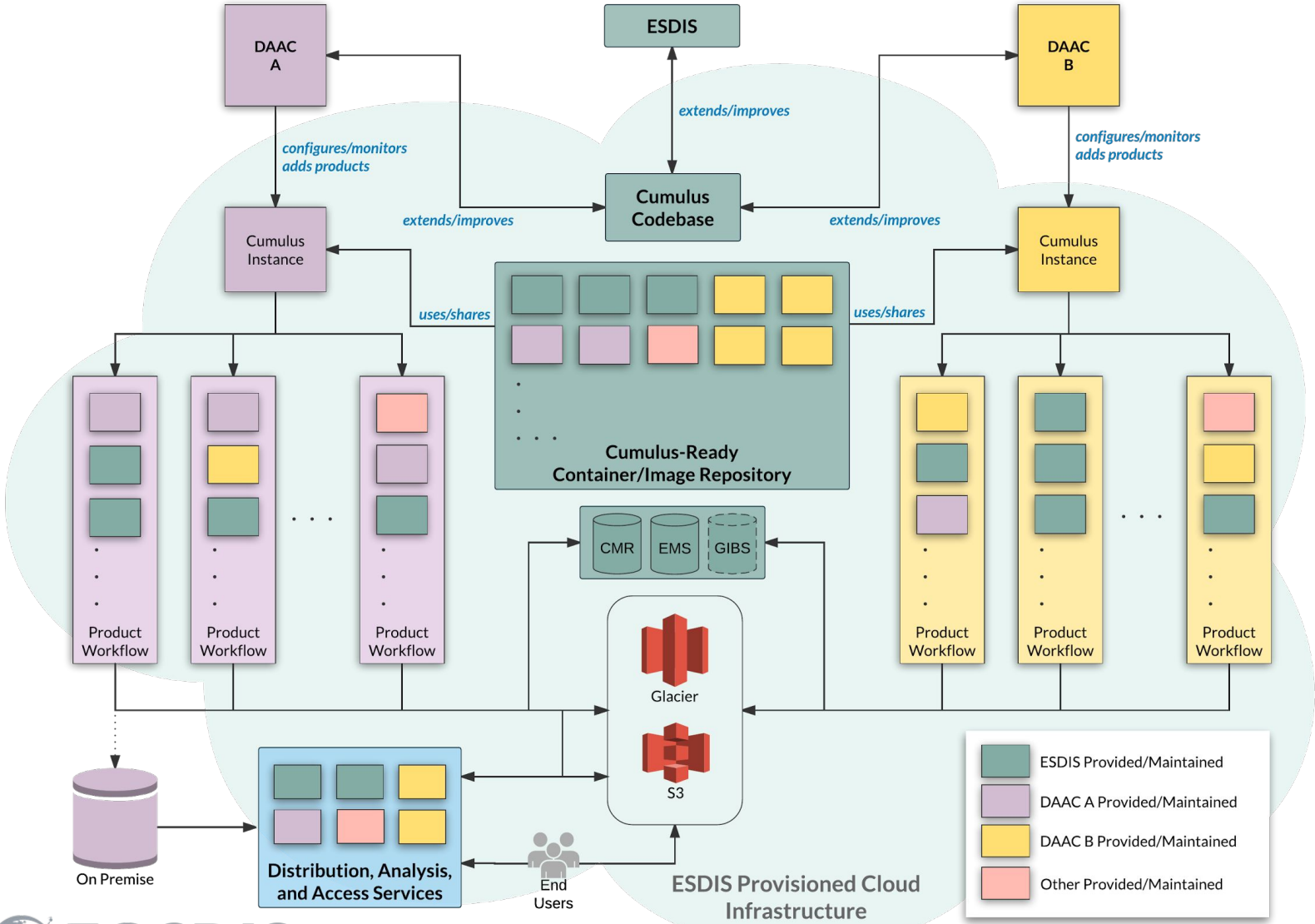
What is Cumulus?

Lightweight cloud-native framework for data ingest, archive, distribution and management

Goals

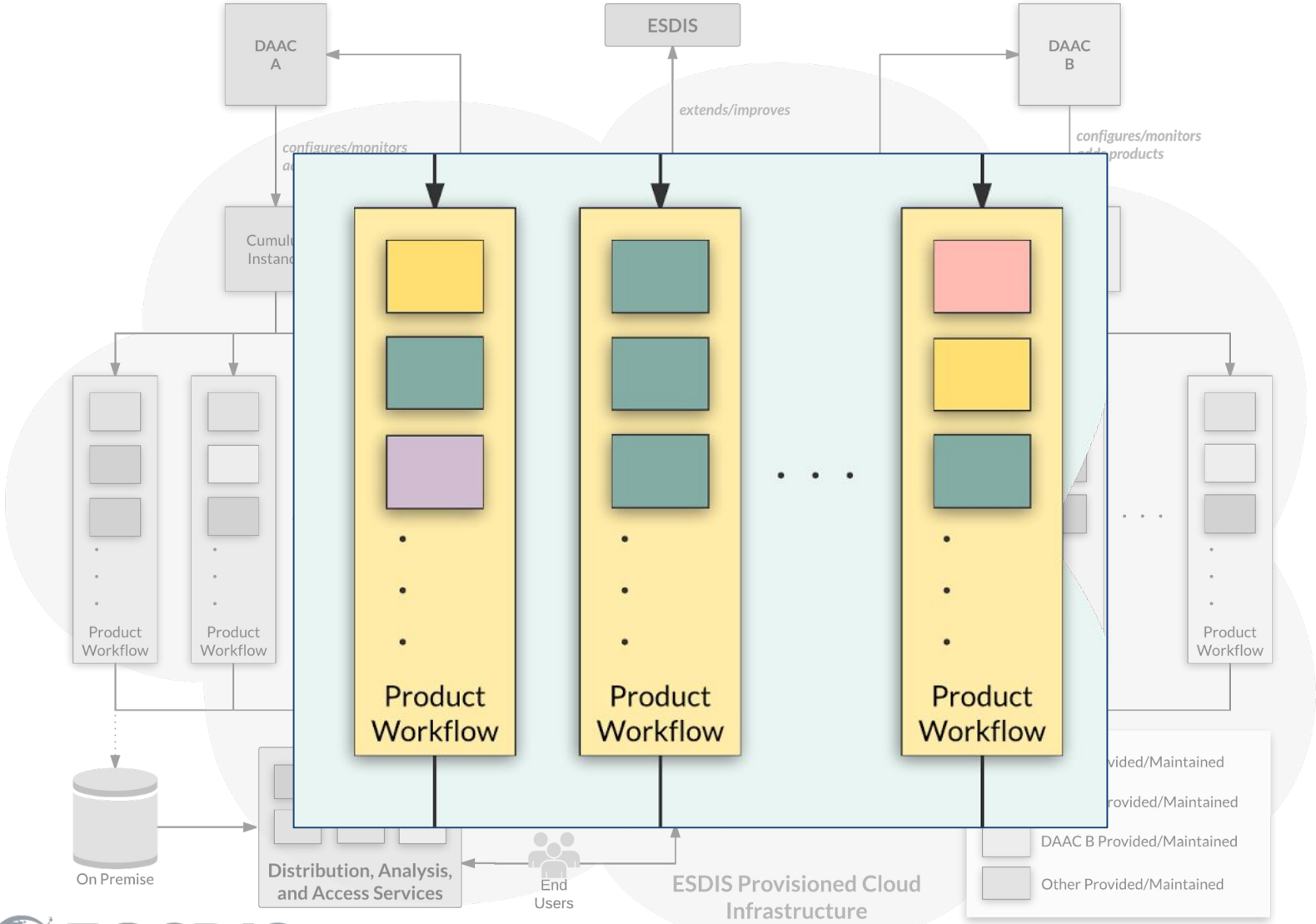
- Provide core DAAC functionality in a configurable manner
  - Data acquisition
  - Data ingest (Validation, Preprocessing)
  - Metadata harvesting, creation, publication into the catalog
  - Data archiving and distribution
  - Metrics publication
- Enable DAACs to help each other with re-usable, compatible containers (e.g. widely applicable GIS components or sub-setters)
- Enable DAAC-specific customizations

Cumulus Vision

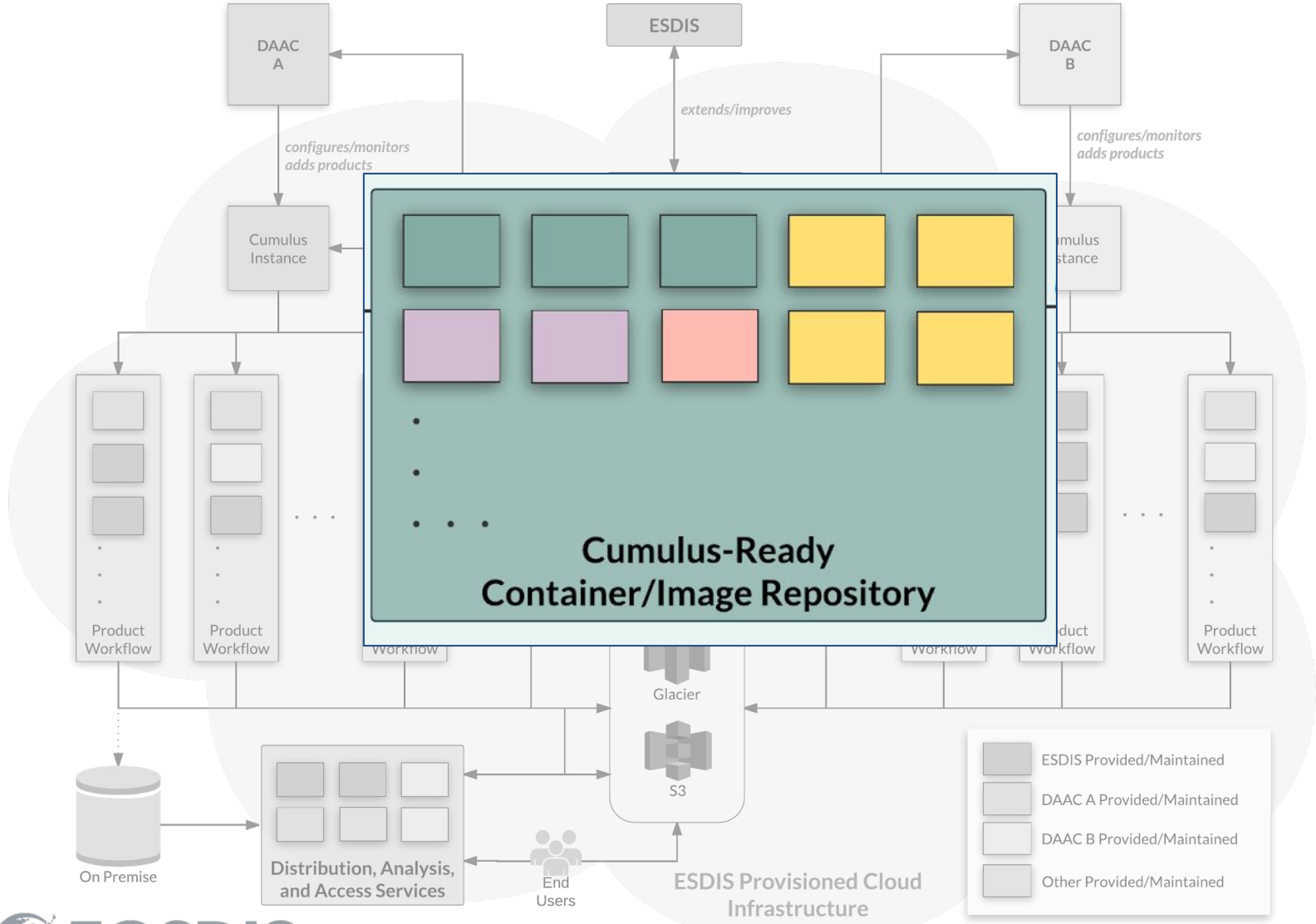


EOSDIS

Configurable Workflows

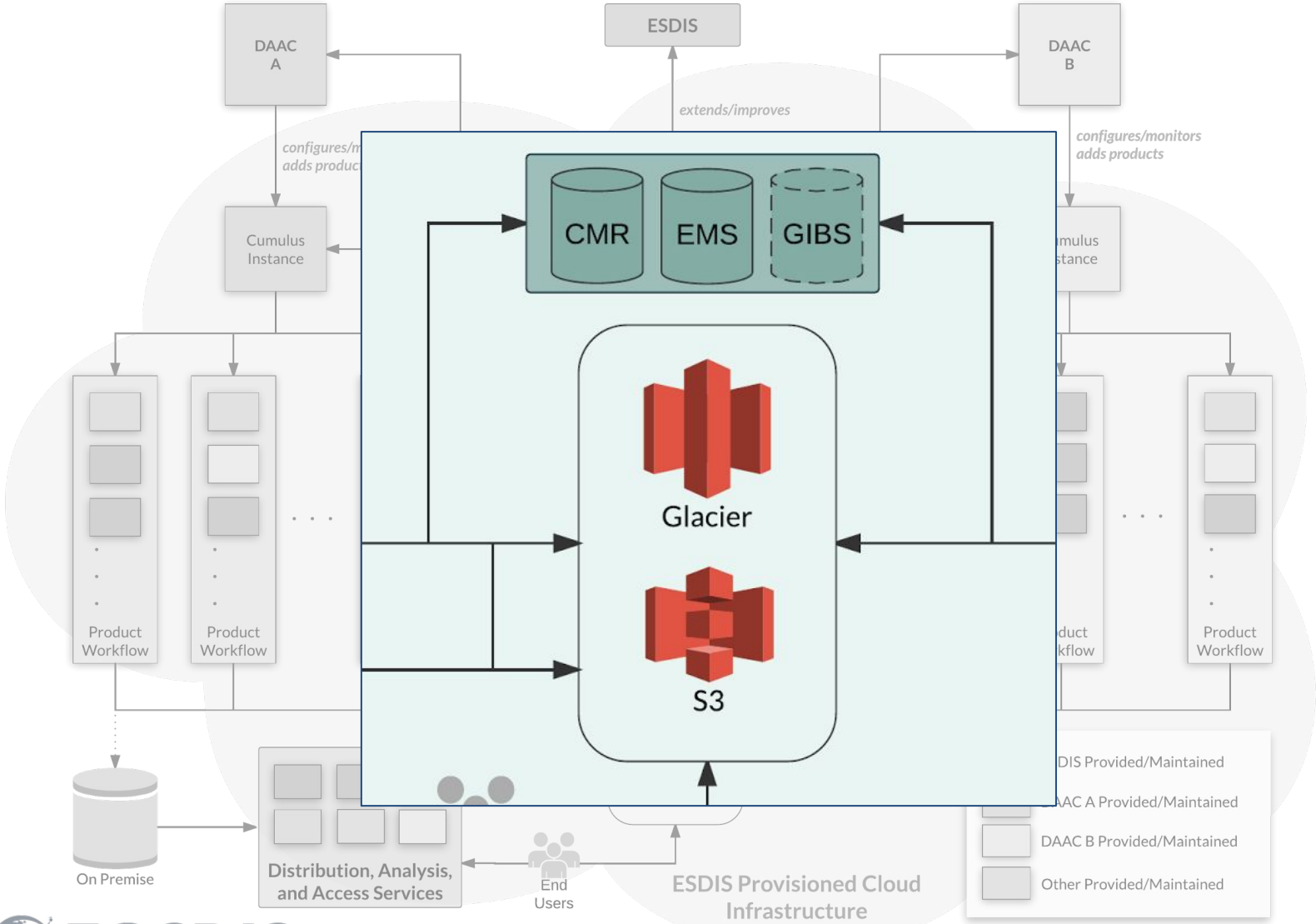


Reusable Components



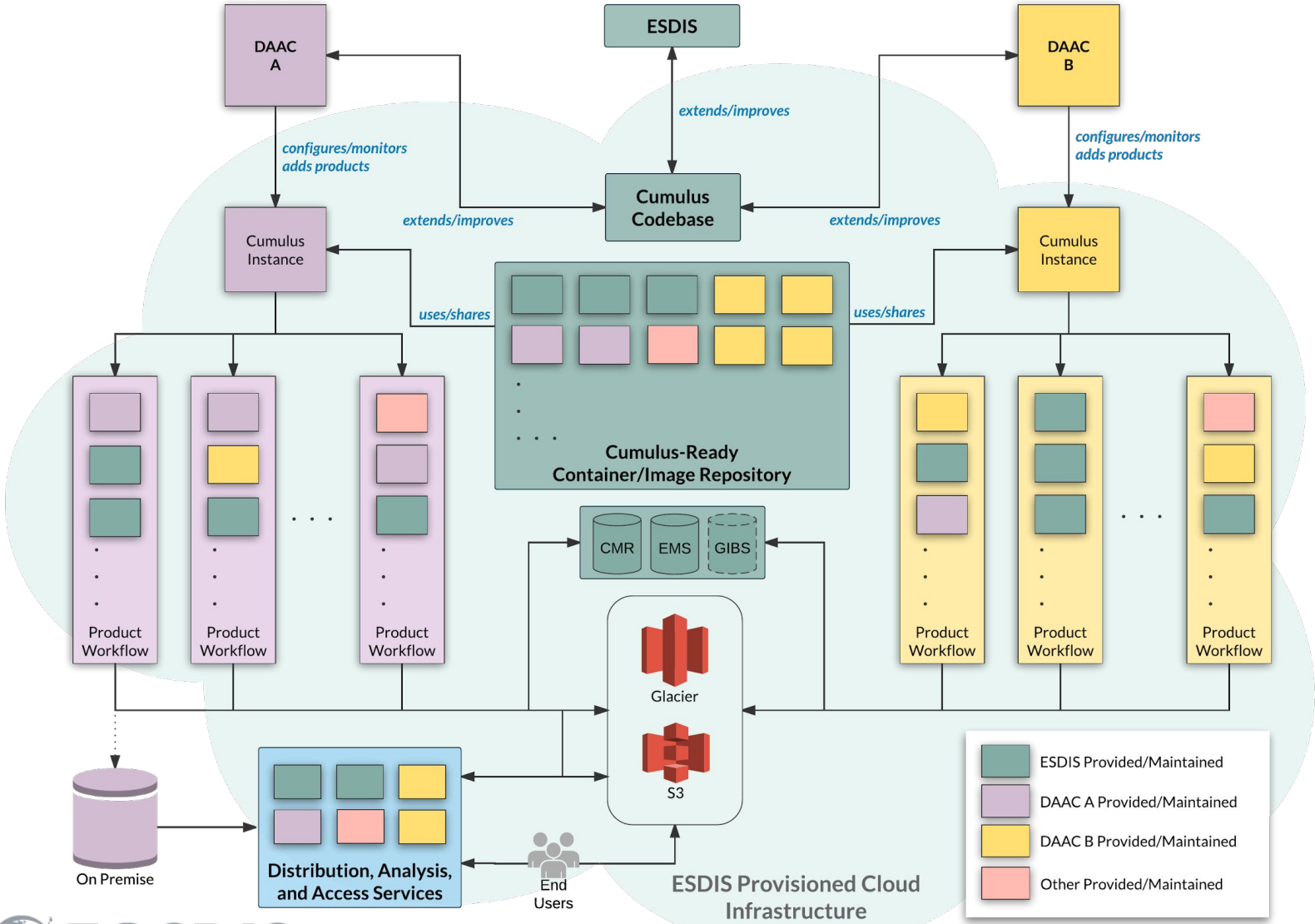
EOSDIS

Common Landing Place

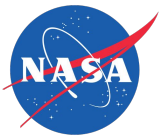


EOSDIS

Cumulus Vision



EOSDIS



Working Towards Streamlining Data Access in Web Object Storage

EOSDIS

OPeNDAP Data Access in the Cloud

- Three Architectures Studied
  - Baseline: HDF S3 backing cache on EBS full reads
  - Full HDF file S3 objects with range index file and range reads
  - "Shredded" HDF files in S3 with index file and full reads
- 21 use cases exercised, some based on real world logs
- Results varied WIDELY based on usage

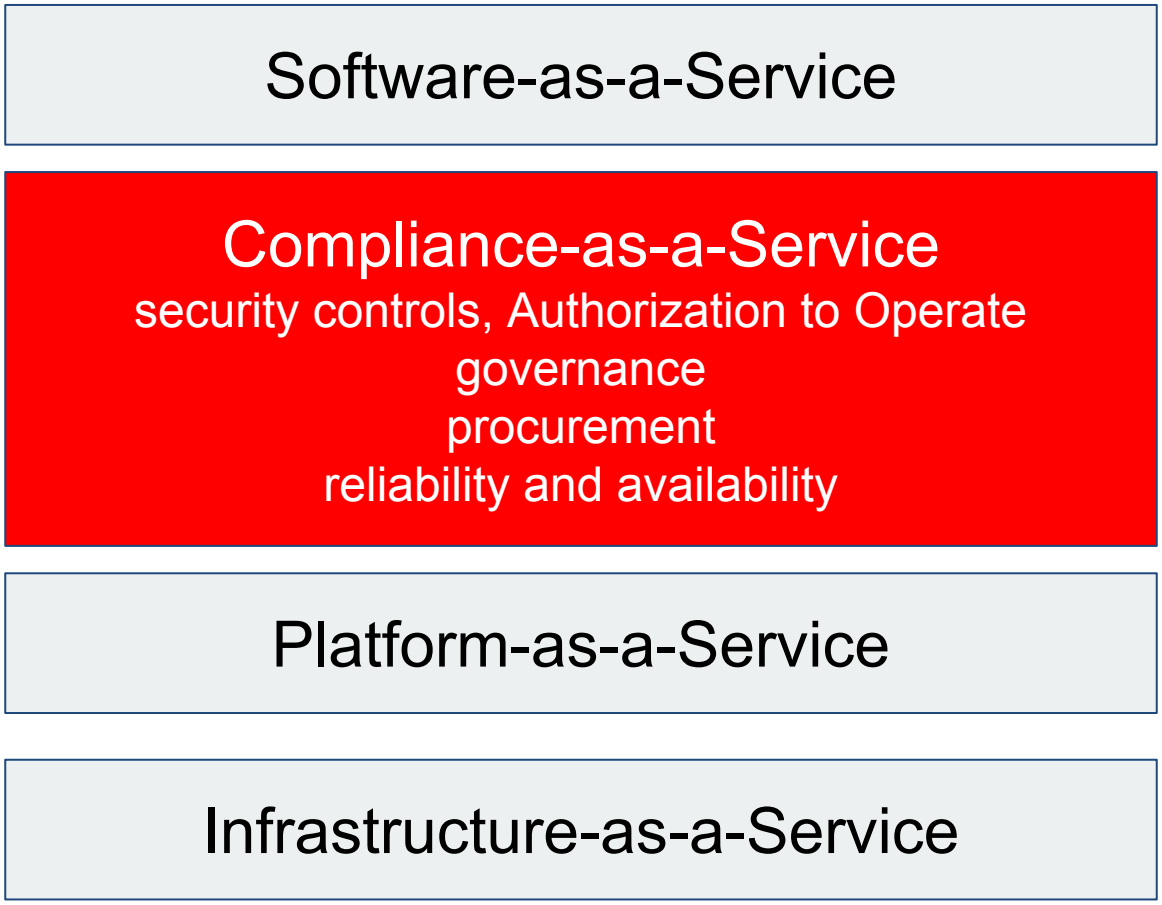
OPeNDAP

The HDF Group

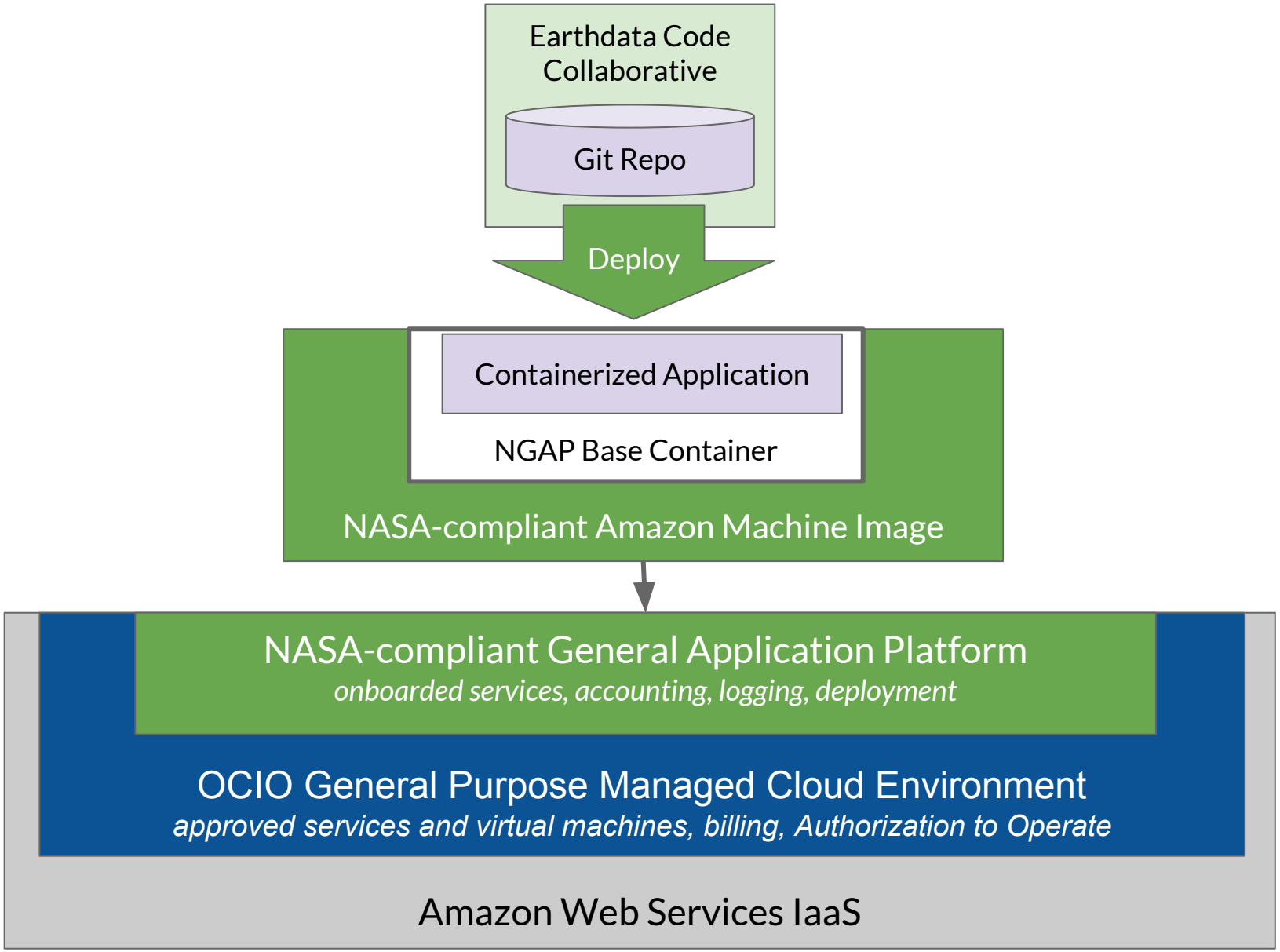
Use-Case 10 aggA	100.00%	5.70%	5.11%
Use-Case 11 aggA	100.00%	11.30%	12.82%
Use-Case 12 aggM	100.00%	27.21%	24.29%
Use-Case 13 aggM	100.00%	51.09%	59.84%
Use-Case 14 aggM	100.00%	16.87%	15.91%
Use-Case 15 aggA	100.00%	4.72%	6.20%
Use-Case 16 aggM	100.00%	32.95%	30.24%
Use-Case 17 aggA	100.00%	10.58%	10.80%
Use-Case 18 multivar multifile M	100.00%	202.71%	316.11%
Use-Case 19 multivar multifile A	100.00%	169.47 %	278.02 %
Use-Case 20 all vars multifile M	100.00%	433.89 %	597.36 %
Use-Case 21 all vars multifile A	100.00%	287.33%	419.11%

NASA Compliant General Application Platform NGAP

“Compliance-as-a-Service”



EOSDIS


Earthdata Code Collaborative
Git Repo
Deploy
Containerized Application
NGAP Base Container
NASA-compliant Amazon Machine Image
NASA-compliant General Application Platform
onboarded services, accounting, logging, deployment
OCIO General Purpose Managed Cloud Environment
approved services and virtual machines, billing, Authorization to Operate
Amazon Web Services IaaS

NGAP Key Takeaways

- We have an ATO through Ames
- We run using nasa.gov hostnames
- We are working on CDM compliance
- This is constantly ongoing work

More technical questions about NGAP?
Ask this guy!

christopher.s.lynnes@nasa.gov

EOSDIS

Summary

- EOSDIS has been operational for > 20 years
- In just the past 5 years, 14 new missions have been added
- Future missions (e.g., SWOT, NISAR) will generate significantly greater data volumes, driving exploration of new strategies for data processing, storage and distribution
- EOSDIS ExCEL Project formed to organize studies and prototypes to help guide decisions on future approaches – involves significant collaboration with stakeholders
- Results are promising, but significant work remains, and both technical and business operations issues are being addressed

EOSDIS is Preparing for the Future

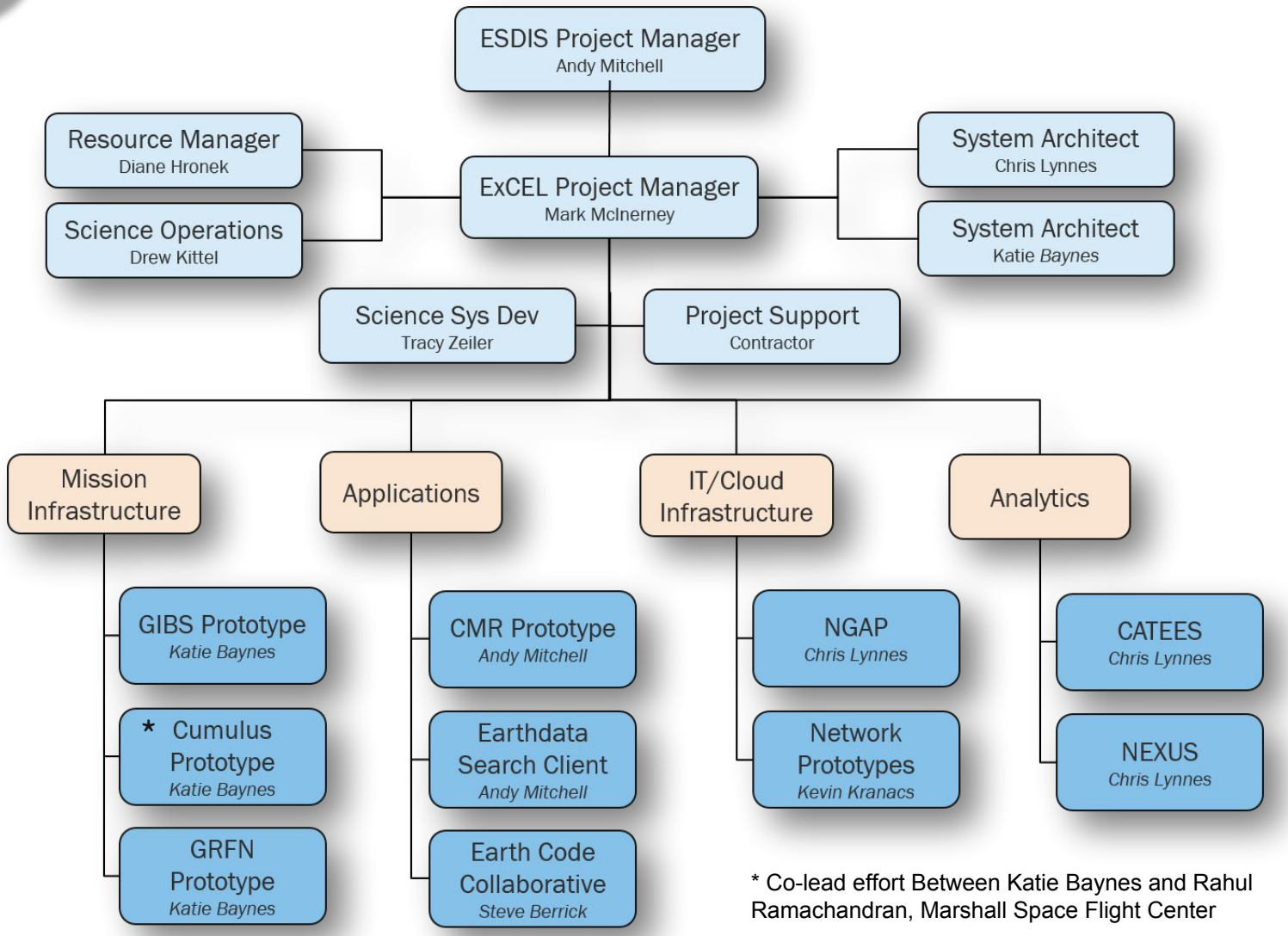
- Preparing for upcoming missions
- Looking towards the cloud
- Streamlining tools
- Continuing commitment to open source

Questions?

kathleen.baynes@nasa.gov

Backup Slides

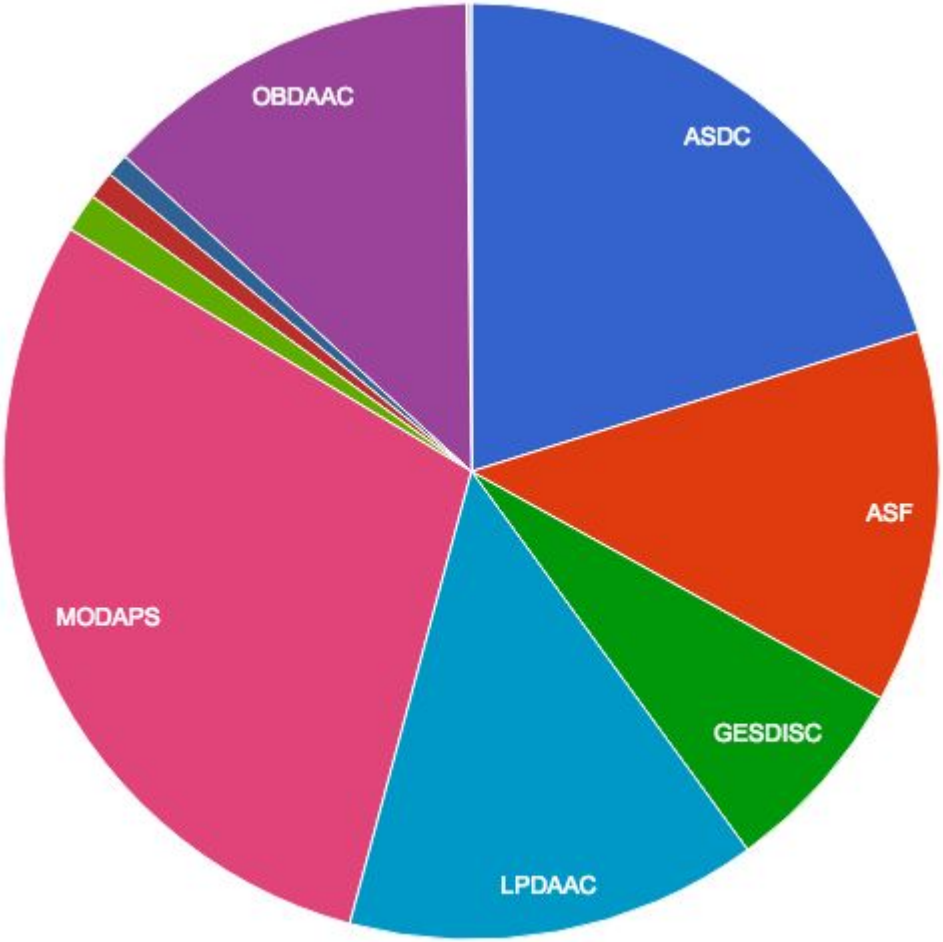
ExCEL Project Organization Chart



* Co-lead effort Between Katie Baynes and Rahul Ramachandran, Marshall Space Flight Center

EOSDIS

Archive Volume by DAAC FY17



EOSDIS