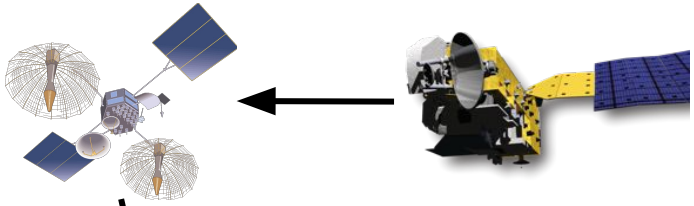
Big Data in the Earth Observing System Data and Information System (EOSDIS)

Chris Lynnes, Katie Baynes, Mark McInerney et al.
Earth Science Data and Information Systems (ESDIS)

Earth Observing System Data and Information System (EOSDIS)



data downlink

EOSDIS

capture and clean

process

archive

distribute

Research

Applications

Education

Users

EDOS

Take Home Message Preview

1. Cloud prototypes are underway to tackle the *Volume* challenge of Big Data…

2. ...But advances in computer hardware or cloud won't help (much) with *Variety*

3. Federation is the key to addressing *Variety*

What is “Big Data”?

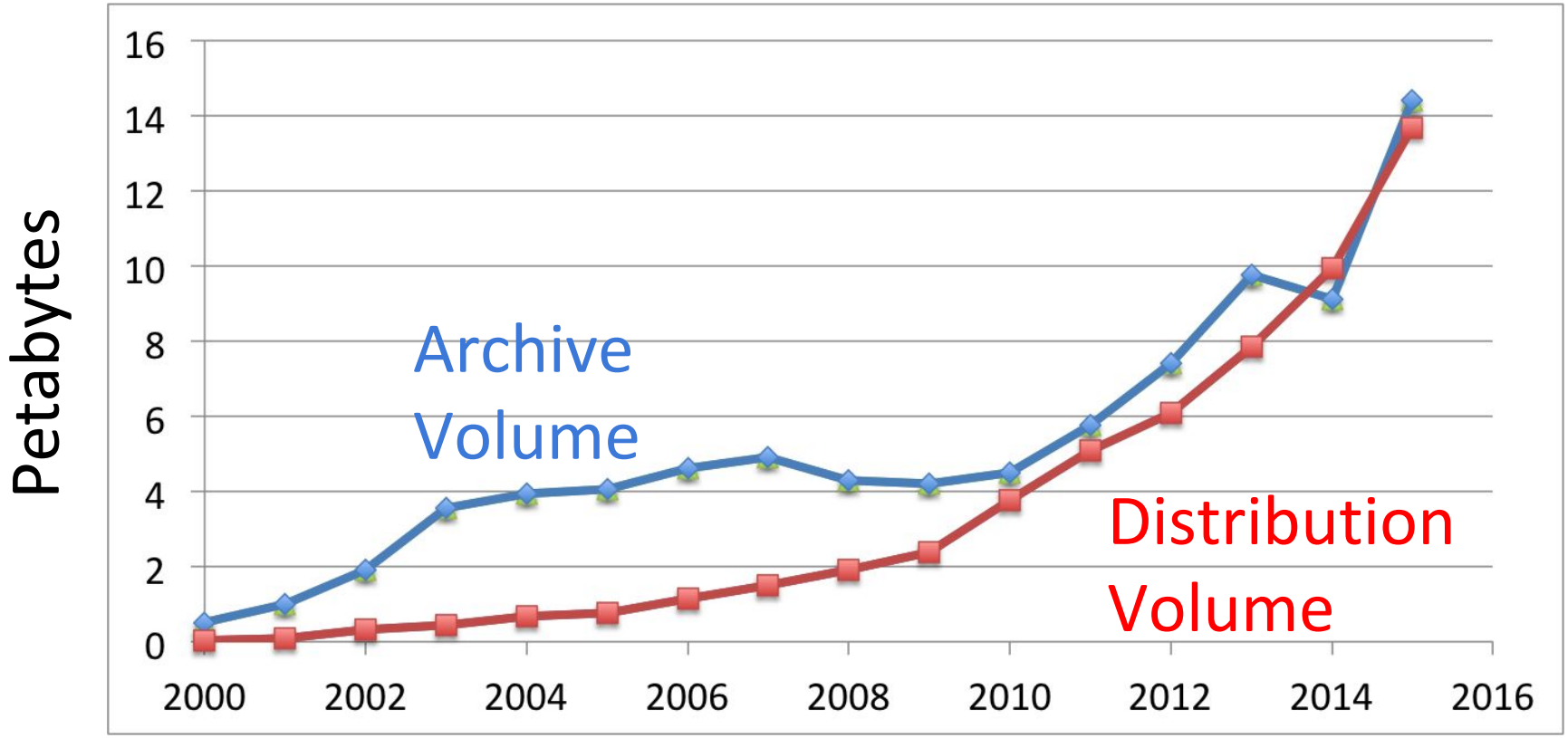
Wikipedia: “data sets that are so large or complex that traditional data processing applications are inadequate to deal with them”

OR

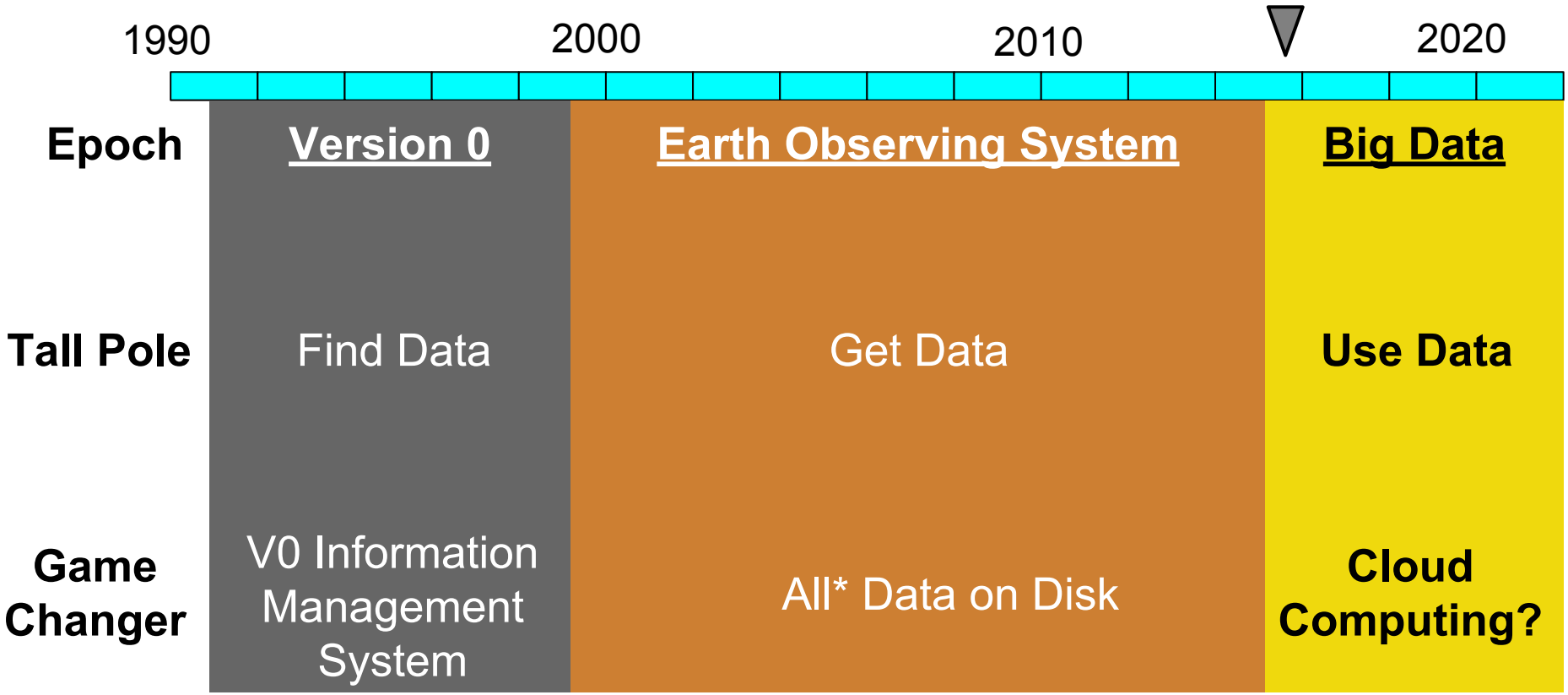
More data than you know what to do with

V is for…

…*Volume*

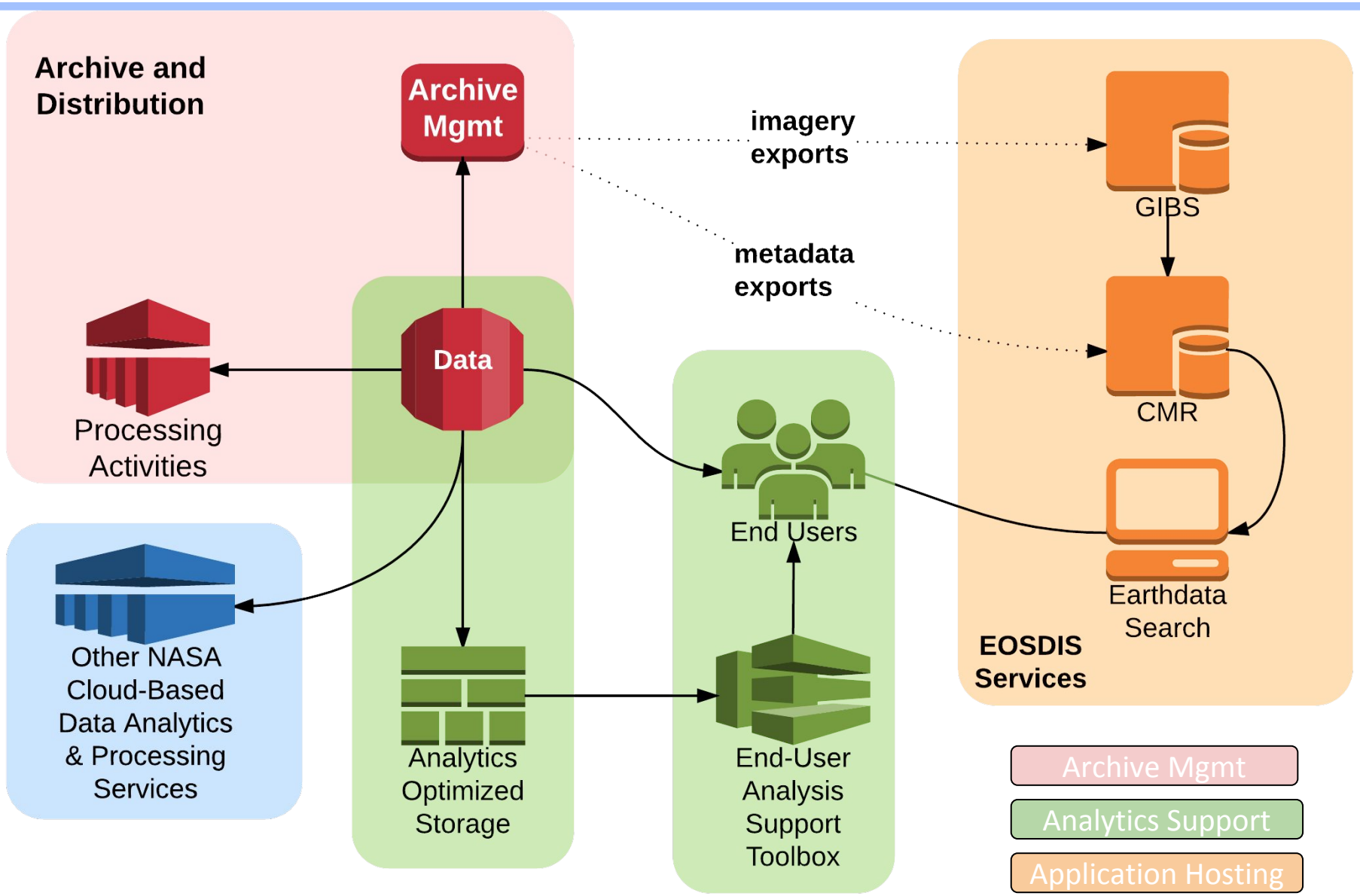


EOSDIS (and Big Data) Evolve Continually

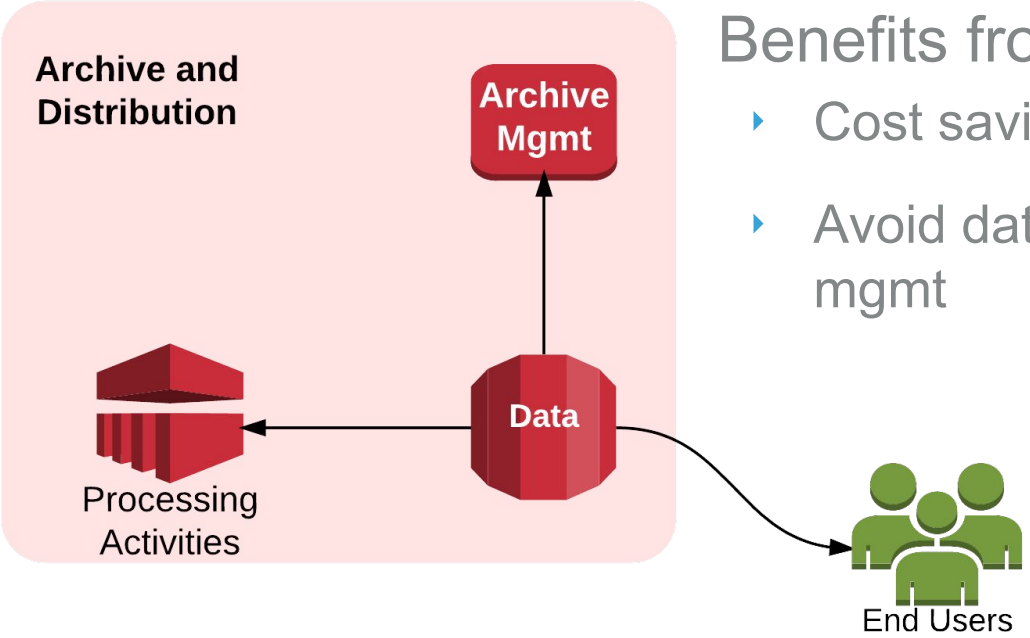


*Almost

EOSDIS Cloud Prototypes



Archive Cloud Prototypes

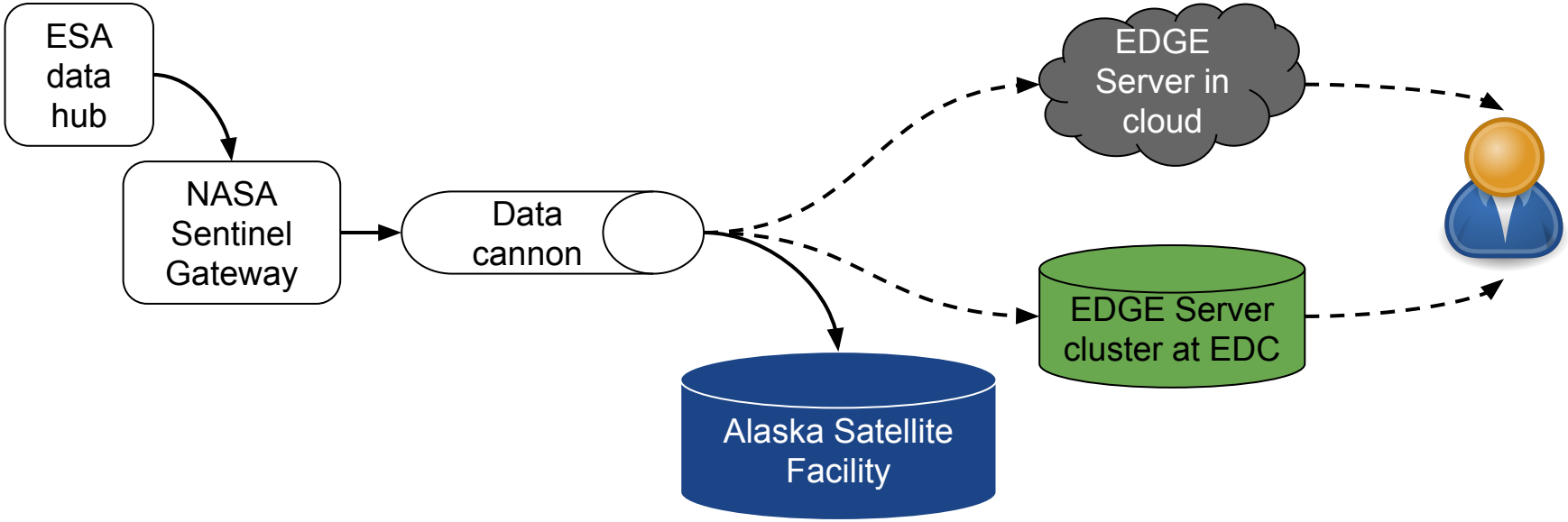


Benefits from Archive in the Cloud

- Cost savings for storage of Big Data?
- Avoid data downloading and local data mgmt

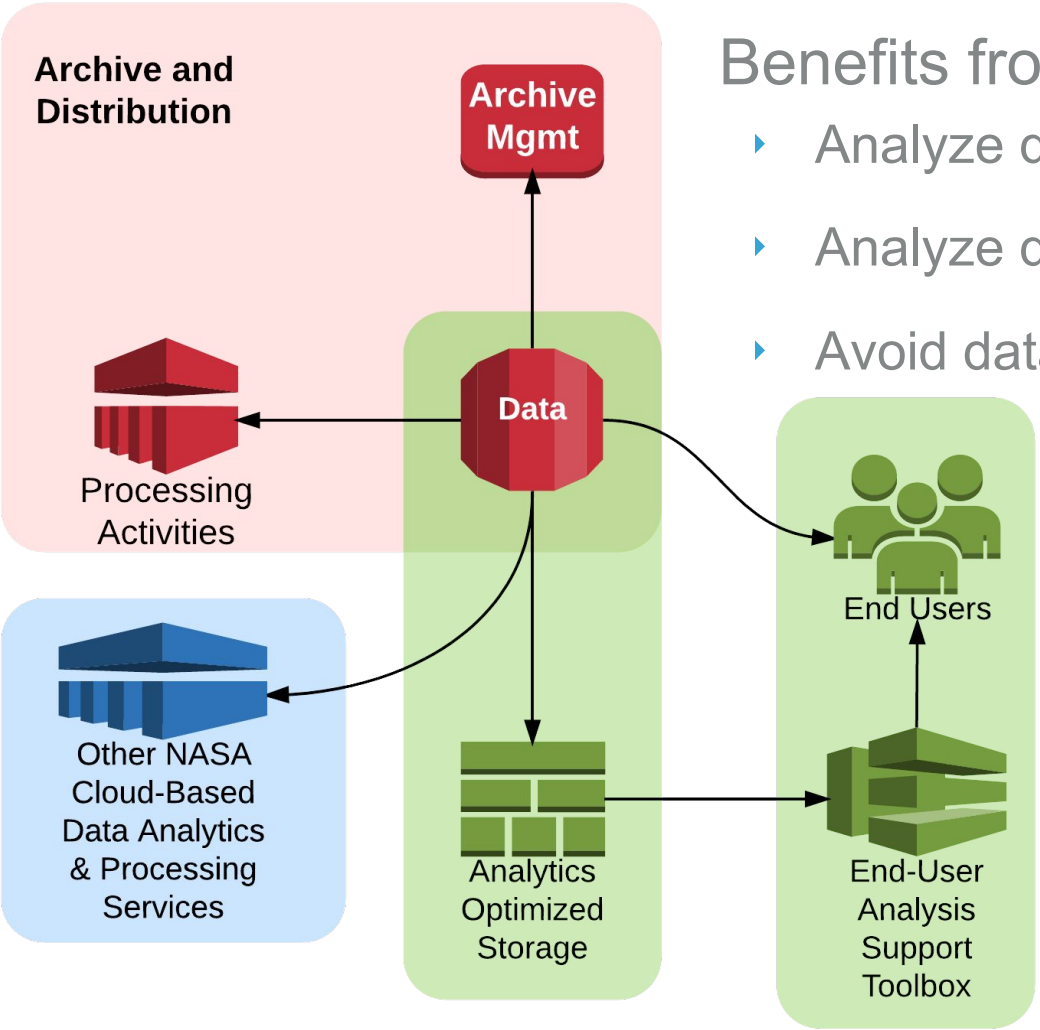
Archive Cloud Prototypes

- Alaska Satellite Facility Web Object Storage prototype
  - Distribute Sentinel radar data from Amazon storage



- Global Imagery Browse Service in the Cloud (Lambda)
- Ingest and Archive management prototype (Lambda)
- NISAR Mission Preparatory Prototype

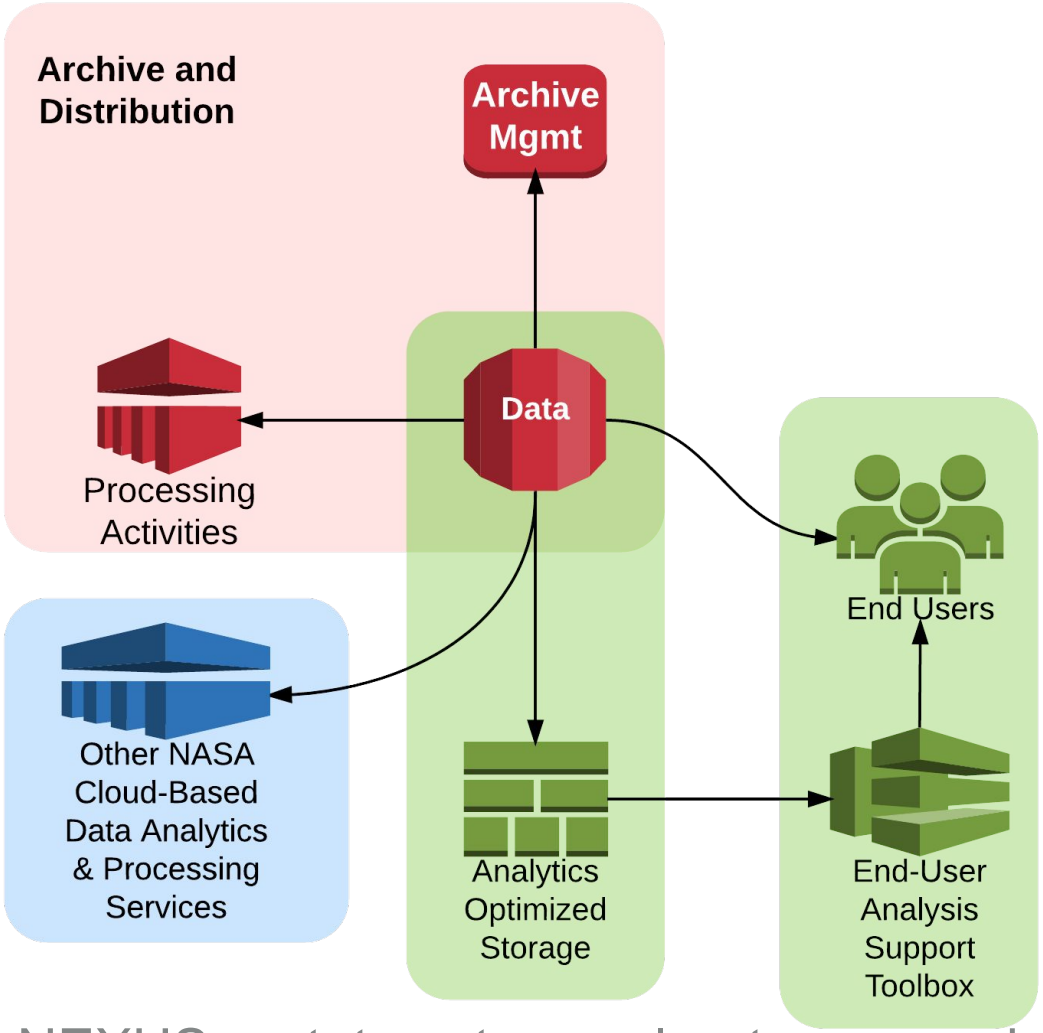
Cloud Analytics Prototypes



Benefits from Cloud Analytics

- Analyze data at scale
- Analyze datasets together easily
- Avoid data downloading and local mgmt

Cloud Analytics Prototypes



Analysis support toolbox to attract users to cloud analytics

- Community open source tools
- DAAC-developed tools
- Cloud analytics examples and recipes

NEXUS prototype to accelerate processing

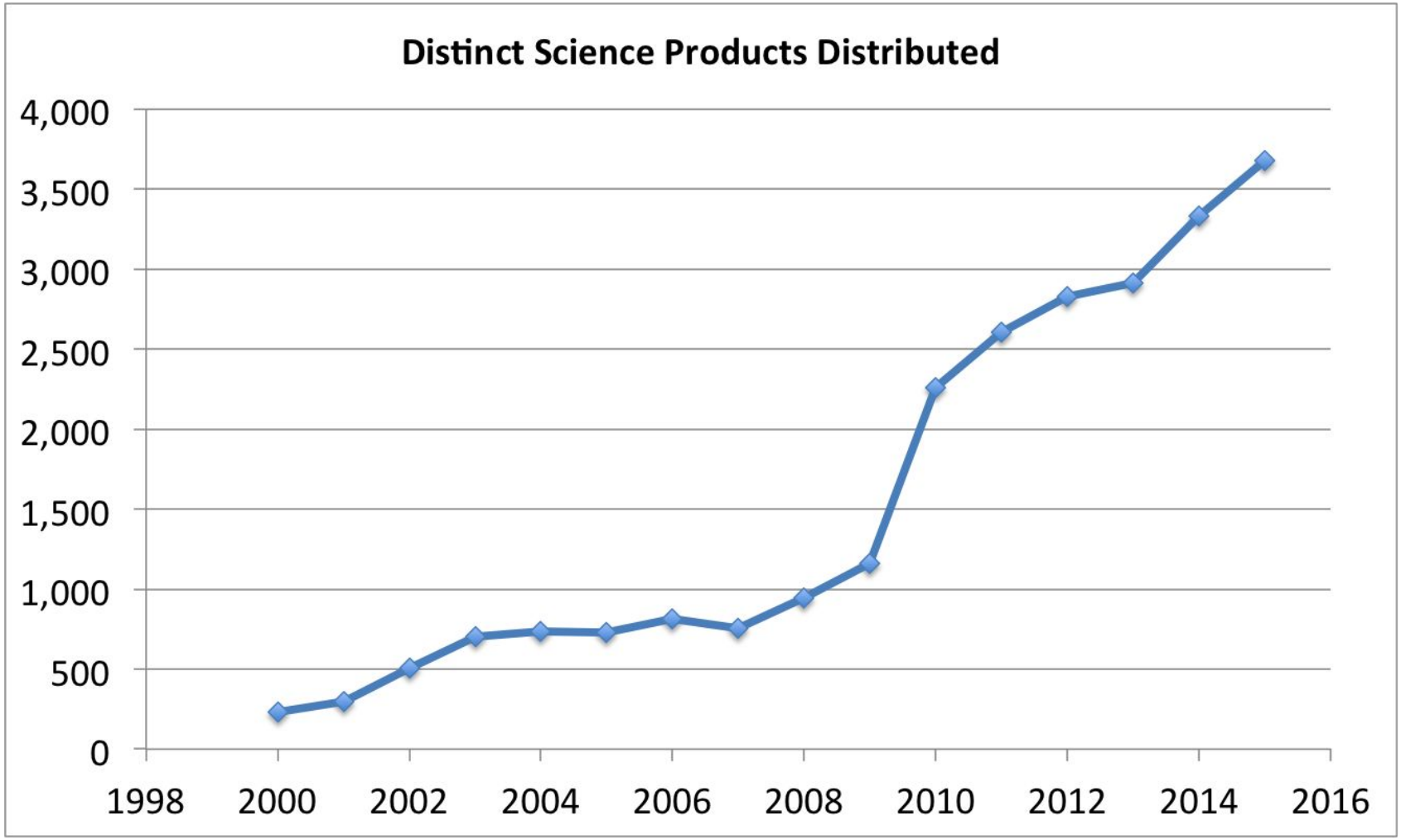
- Spark + Cassandra DB

Terra Incognita

1. Uncapped egress costs
2. Vendor lock-in
3. Future storage costs
4. Security restrictions and Network trust
5. Changes in architecture
6. Changes in procedures

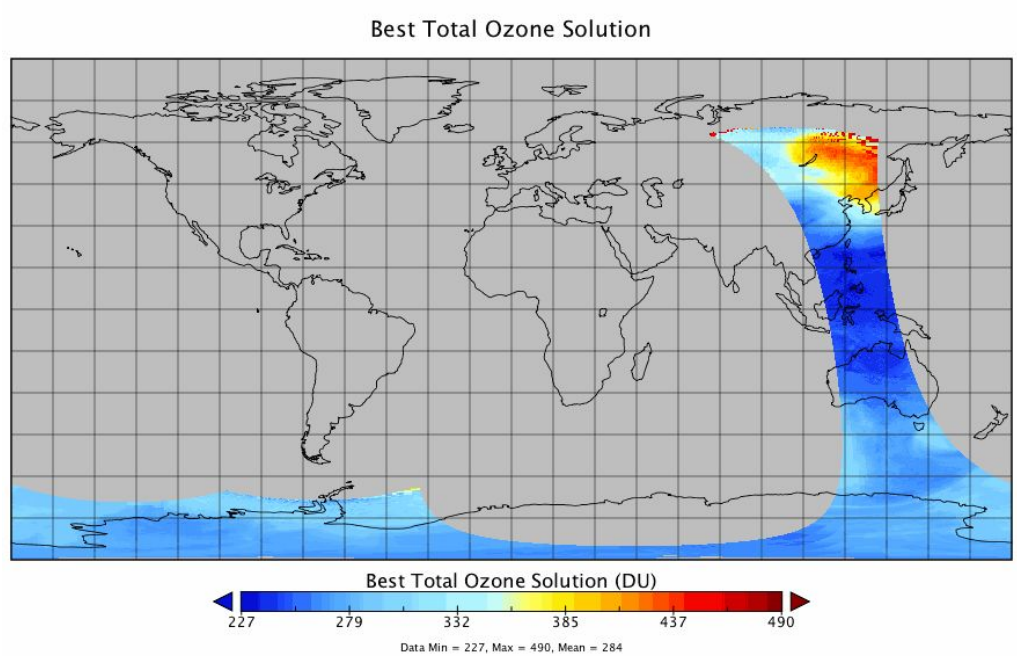
V is for...

...*Variety*

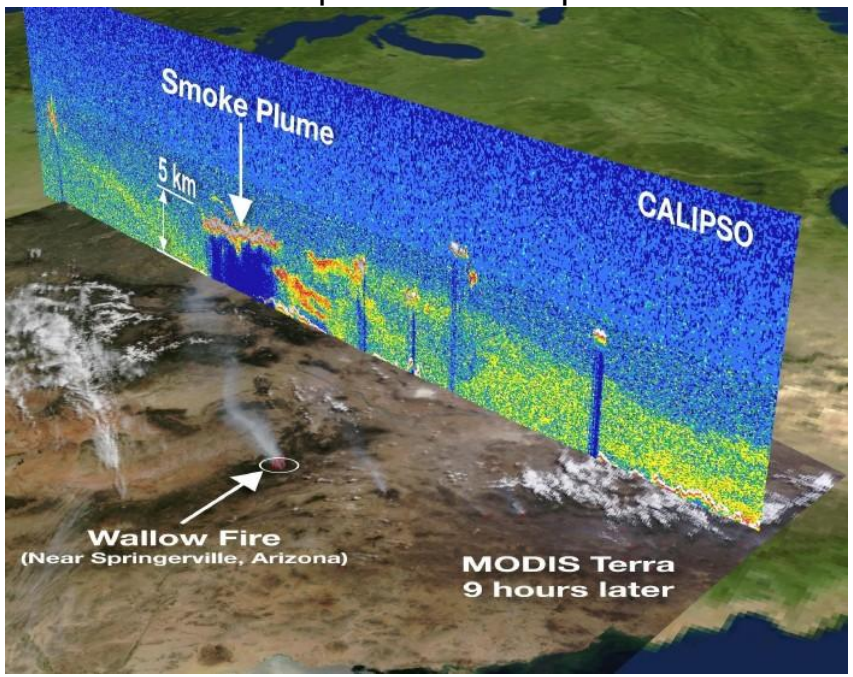


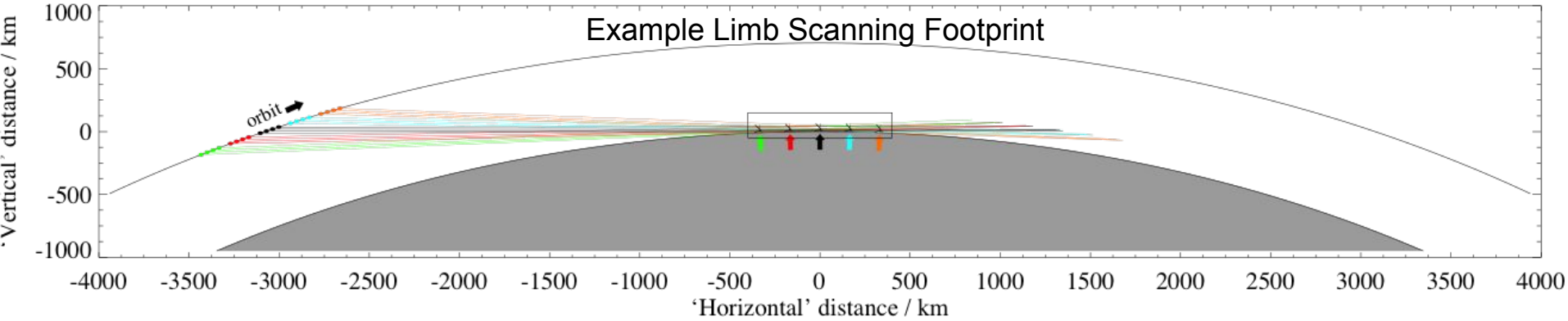
Satellite Instrument "Footprint" Variety

Example Imaging Footprint



Example LIDAR footprint



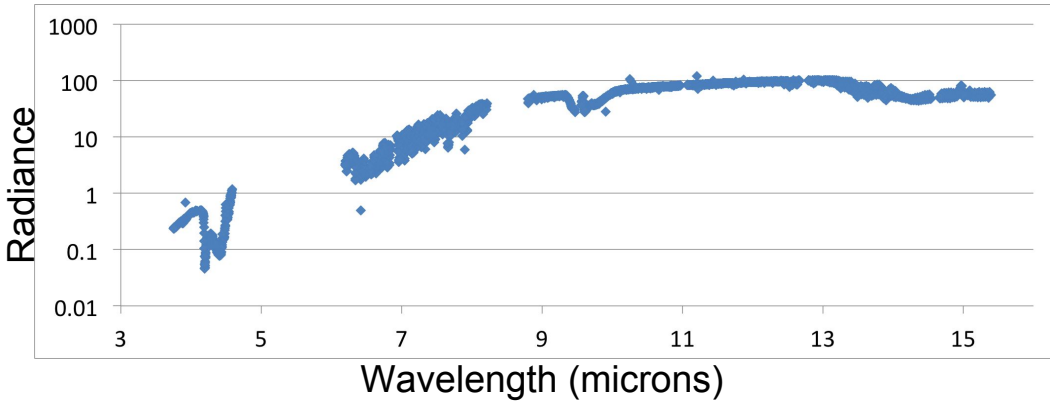


Microwave Limb Scanner (from Algorithm Theoretical Basis Document, Livesey and Wu, 1999)

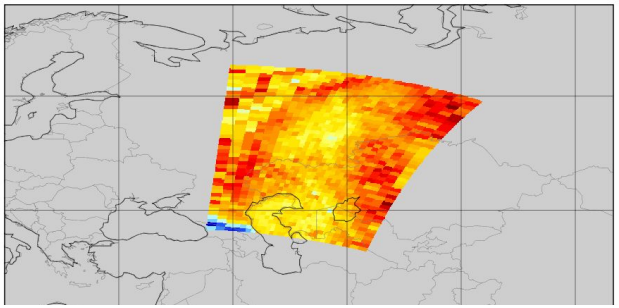
Processing Level Variety

AIRS data for 2011-08-11

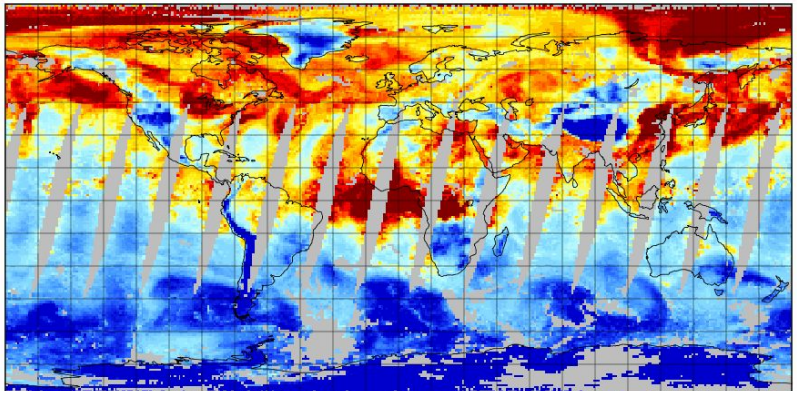
Level 1B
Calibrated radiance at a pixel



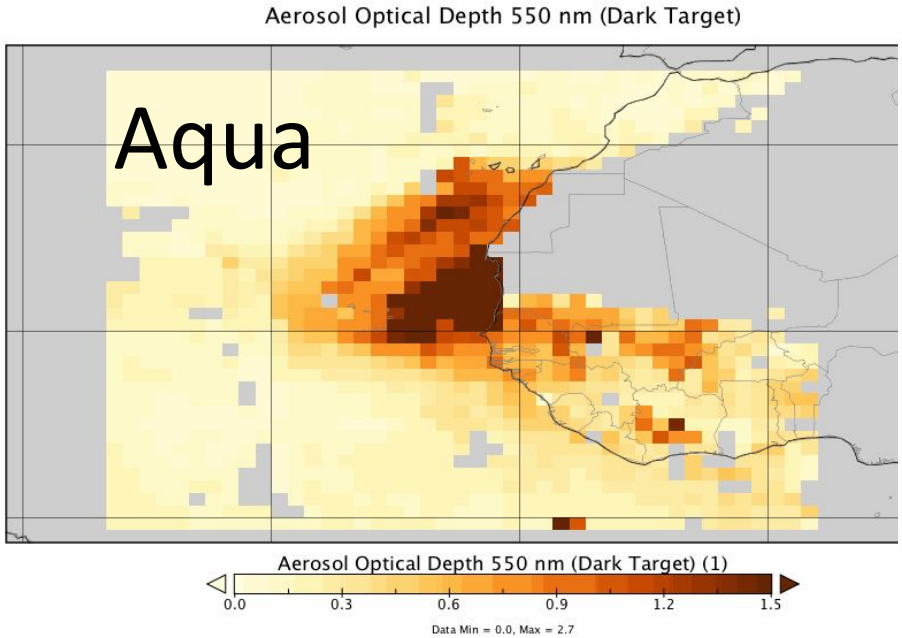
Level 2
Carbon monoxide for one scene

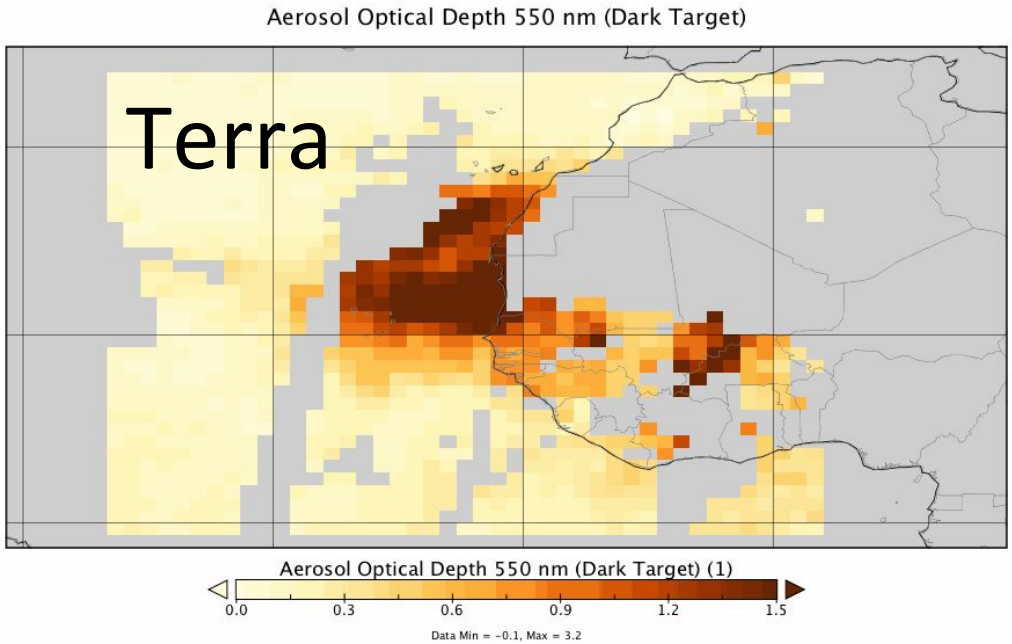


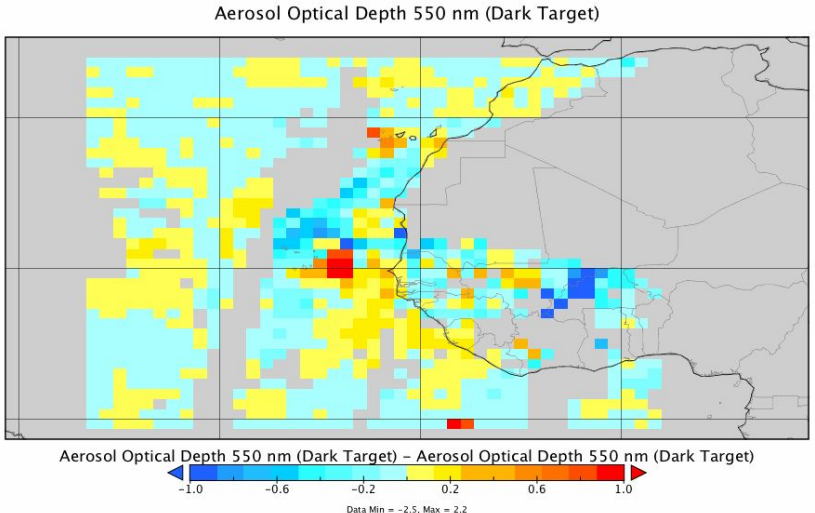
Level 3
Global carbon monoxide for one night



Same Instrument, Different Satellite

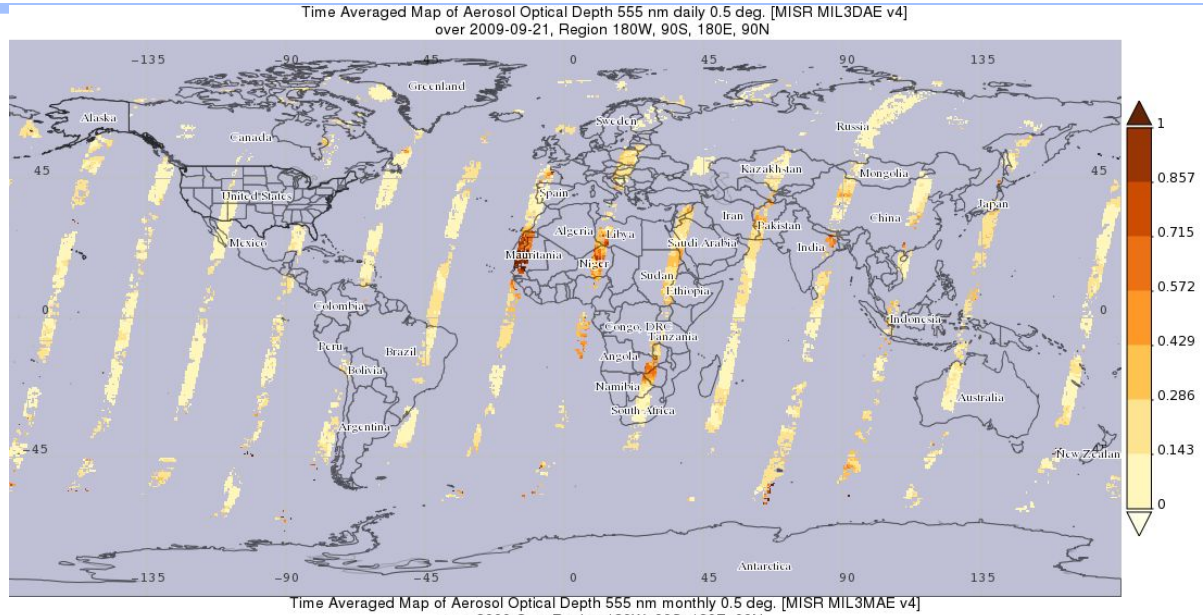






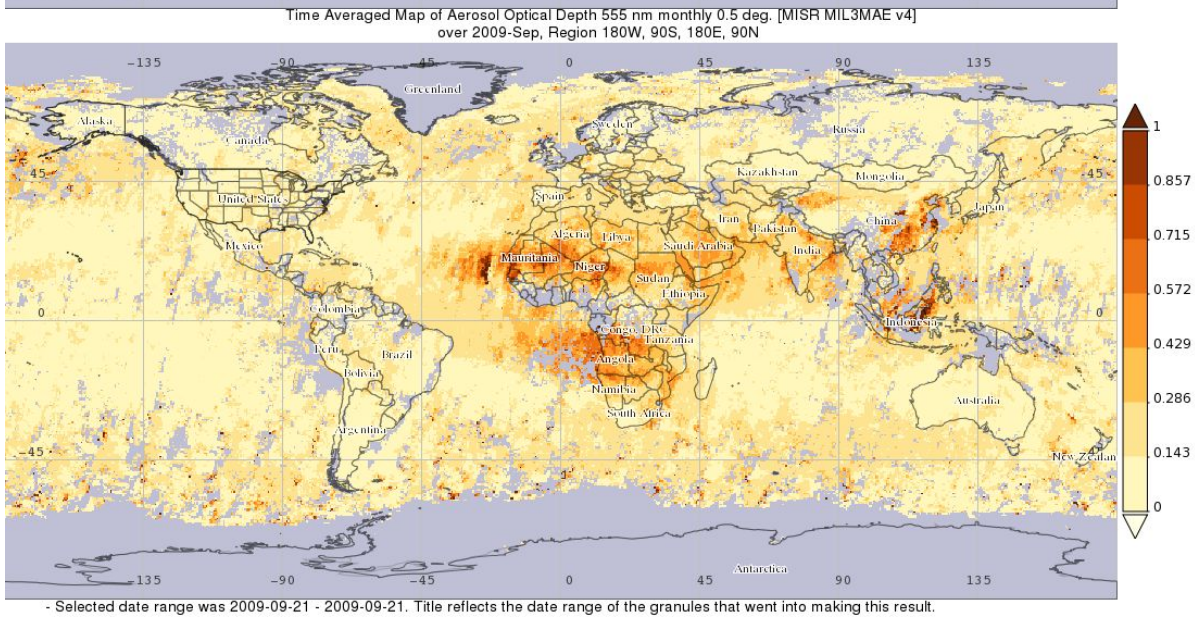
Aqua - Terra

Time Aggregation Variety



Aerosol Optical Depth at 555 nm from Multi-angle Imaging Spectro-Radiometer

Daily

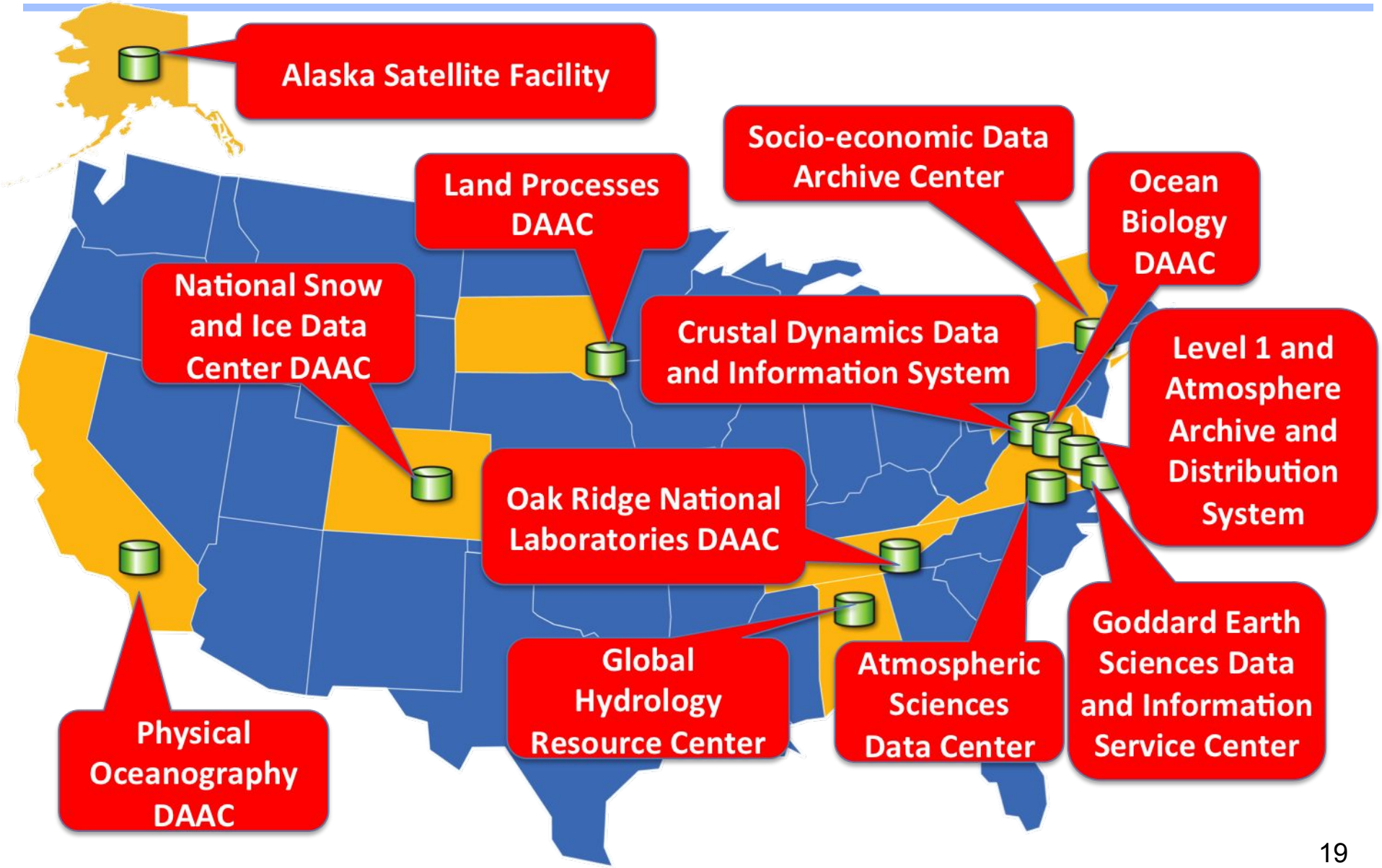


Monthly

Solution to the Variety Problem: Federation

Federation = Distributed Responsibility + Common Frameworks

Discipline-Focused Distributed Active Archive Centers (DAACs)



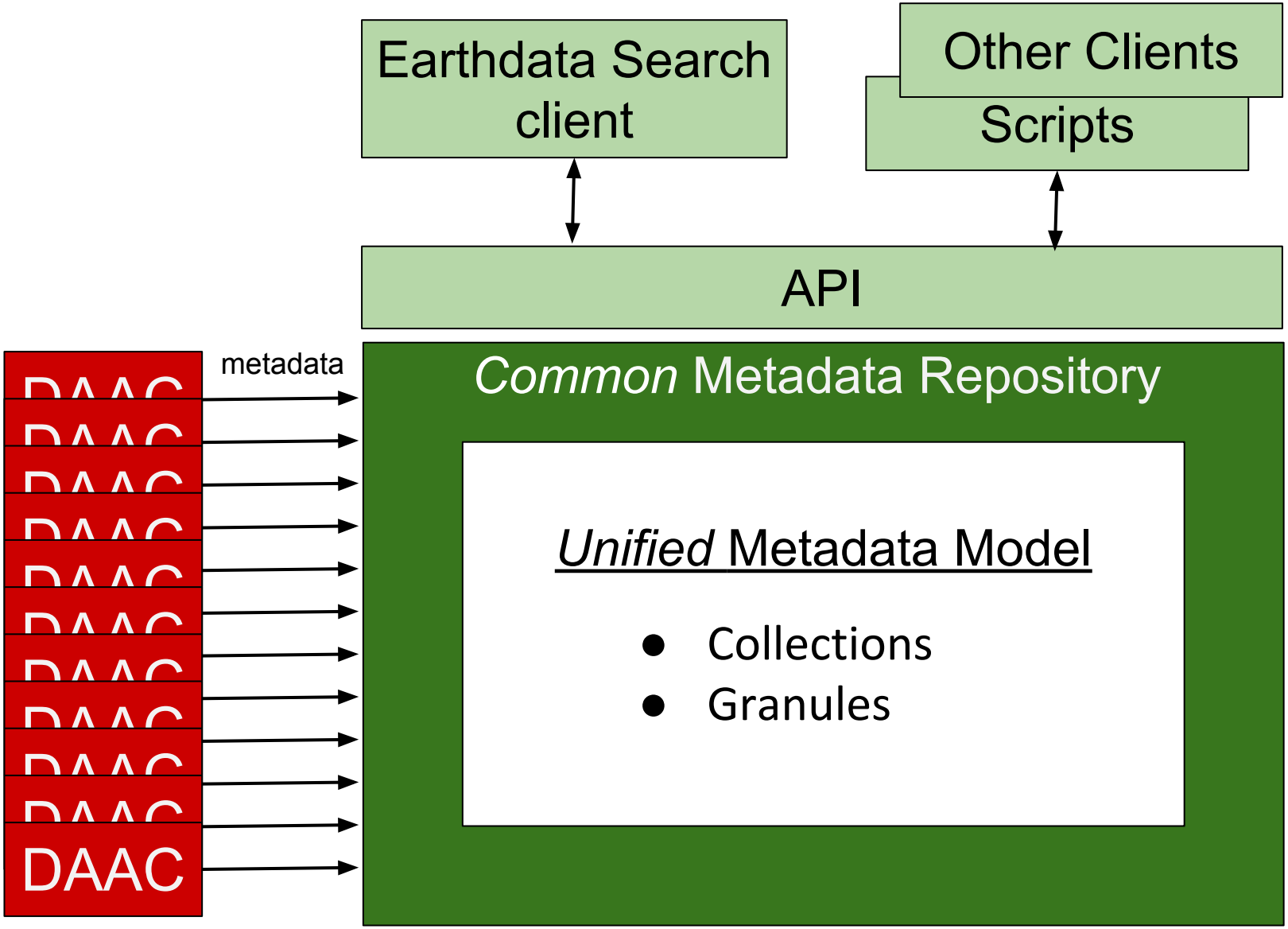
Different DAACs, different "-Spheres of Influence"

DAAC	Atmo	Hydro	Bio	Cryo	Litho	Anthropo
Atmospheric Sciences Data Center	✓					
L1 and Atmosphere Archive & Dist System	✓					
Goddard Earth Sciences DISC	✓	✓				
Global Hydrology Resource Center		✓				
Physical Oceanography DAAC		✓				
Ocean Biology DAAC		✓	✓			
Oak Ridge National Laboratory DAAC			✓			
Land Processes DAAC			✓			
National Snow & Ice Data Center DAAC				✓		
Alaska Satellite Facility				✓	✓	
Crustal Dynamics Data Information System					✓	
Socioeconomic Data Archive Center						✓

Common Frameworks knit EOSDIS together

Role	Implementations
Data Inventory	Common Metadata Repository
Image Repository	Global Imagery Browse System
Data Access	OPeNDAP: Open-source Project for a Network Data Access Protocol
Data Format Standards	• Hierarchical Data Format • network Common Data Form • EOSDIS Standards Office
Collaborative Development	Earthdata Code Collaborative • Stash (git) code repo • Jira tickets • Confluence wiki and forum

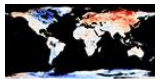
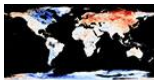
The Common Metadata Repository presents a consistent catalog for discovery of data from multiple DAACs

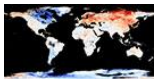


Federation Solution to the Variety Problem

Distributed Responsibility + Common Frameworks = Federated Tools

Federated Tools

Tool	*Role*	*# of DAACs*
Earthdata Search client	🔍	12+
Worldview		8+ ⇒ 10+
Simple Subset Wizard	✂	10
Federated Giovanni	∑ 	5
AppEEARS subset broker	✂	1 ⇒ 2+

🔍 Search
 Visualize
✂ Subset
∑ Analyze

EOSDIS Federation (old school), ca 1993