NASA's Earth Science Data Systems

Kevin Murphy
Earth Science Division
Science Mission Directorate
NASA HQ – Washington D.C.

Agenda

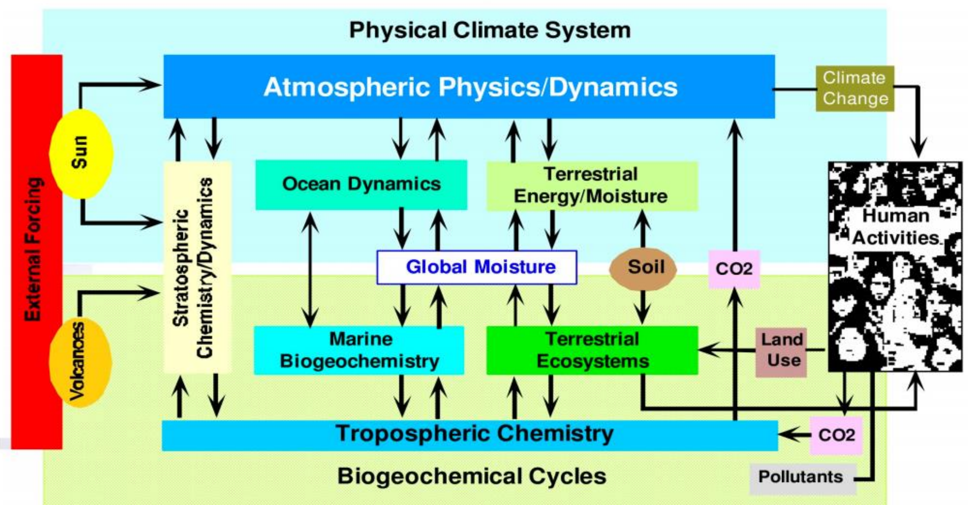
- NASA's Earth Science Data System (ESDS)
- Evolving Capabilities
- Future Architectures (Discussion)

NASA'S EARTH SCIENCE DATA SYSTEM (ESDS)

NASA's Earth Observations Support Science and Society

- Earth observations
  - improve understanding of the physical, chemical and biological processes – and interactions – involved in the evolution of Earth systems
  - quantify and aid in the modeling of Earth systems at a range of spatial and temporal scales
- Improved understanding is vital for securing natural resources, and to mitigate and adapt to environmental change

The Bretherton Diagram



(from Earth System Science: An Overview, NASA, 1988)

NASA's data systems play a vital role in supporting earth science research and making the science advancements (data and information) available to everyone for research and societal benefit.

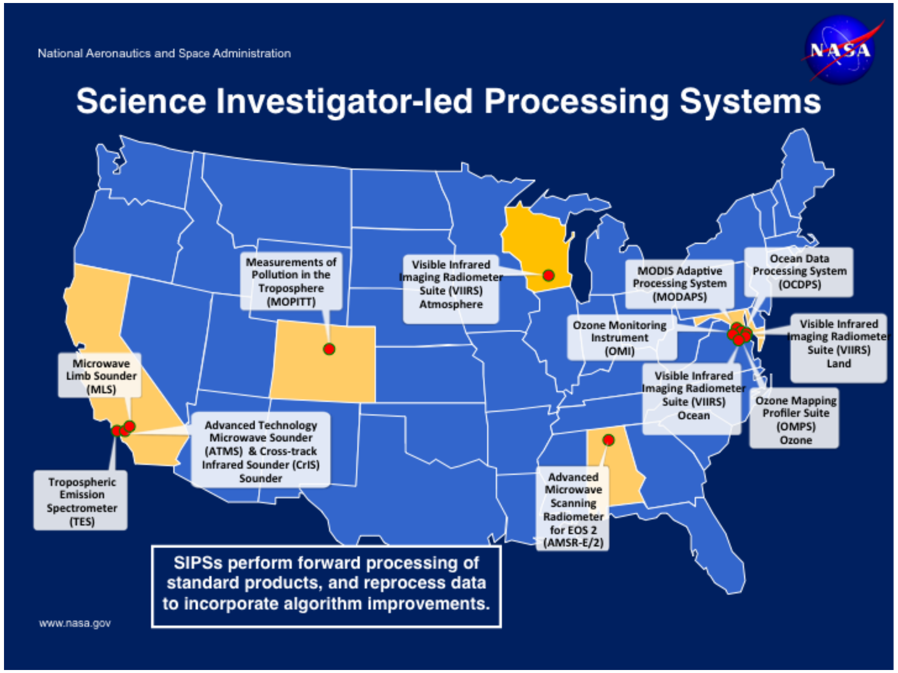
Earth Science Data Systems (ESDS) Program is Responsible for:

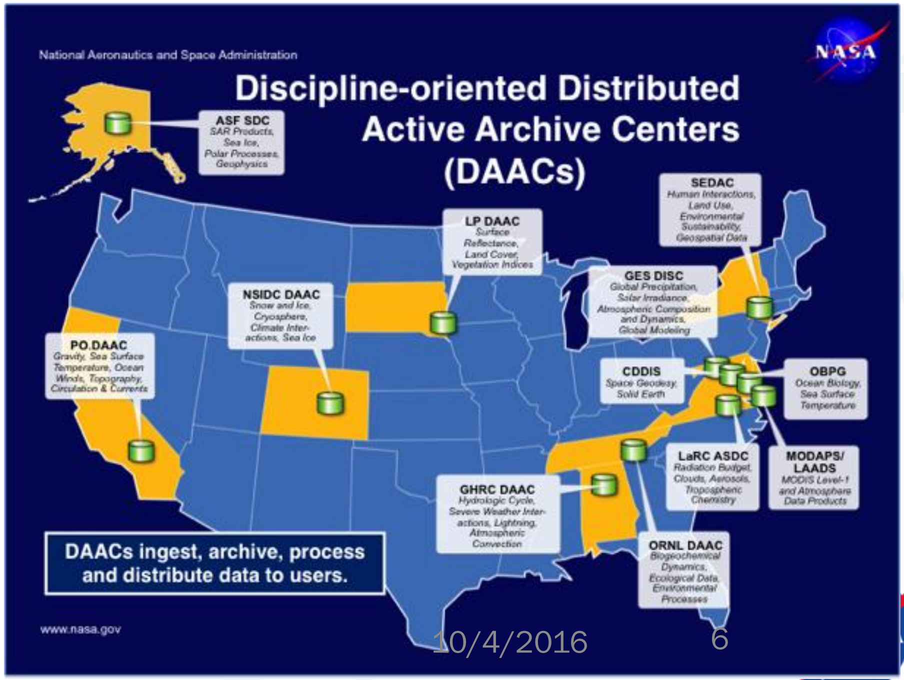
- Actively managing NASA's earth science data as a national asset.
- Developing unique data system capabilities optimized to support rigorous science investigations and interdisciplinary research.
- Processing (and reprocessing) instrument data to create high quality earth science data records.
- Upholding NASA's policy of full and open sharing of all data, tools and ancillary information for all users.
- Engaging members of the earth science community in the evolution of data systems.

Data Systems must ensure the long-term preservation of unique measurements and act as the Rosetta Stone among science disciplines.

NASA'S EOSDIS

- Earth Science mission data are **processed** by Science Investigator-led Processing Systems (SIPS) designed for missions and measurements
- Stewardship of Earth Science Data is conducted by Distributed Active Archive Centers (DAACs) that provide **knowledgeable curation** and **science-discipline-based support**
- NASA provides **high bandwidth** network connectivity to support production data flows and community access to data, including access to **near real time data**

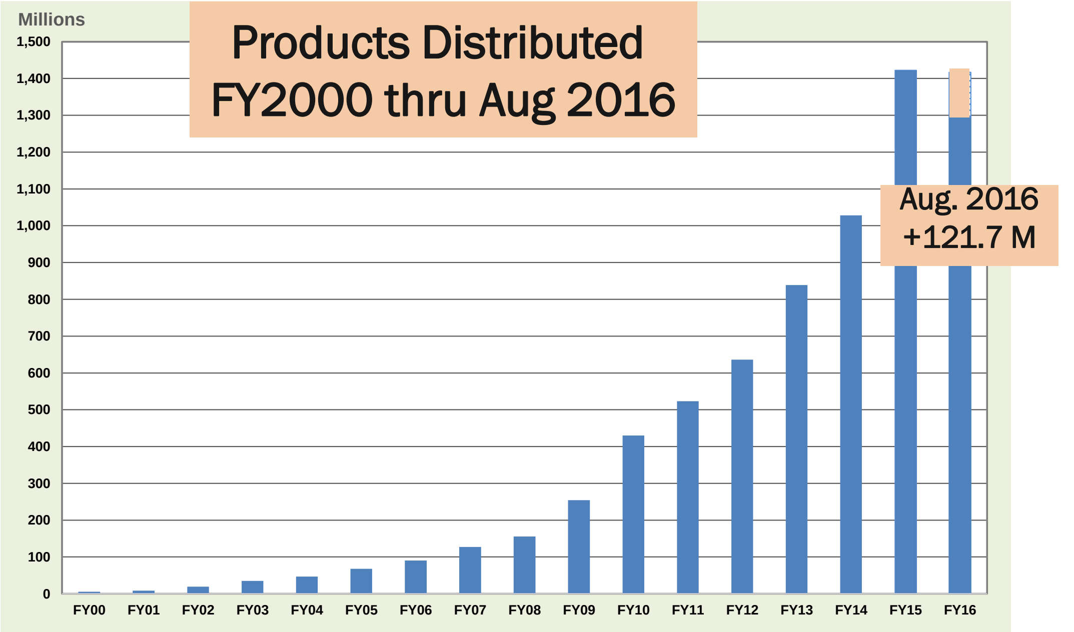


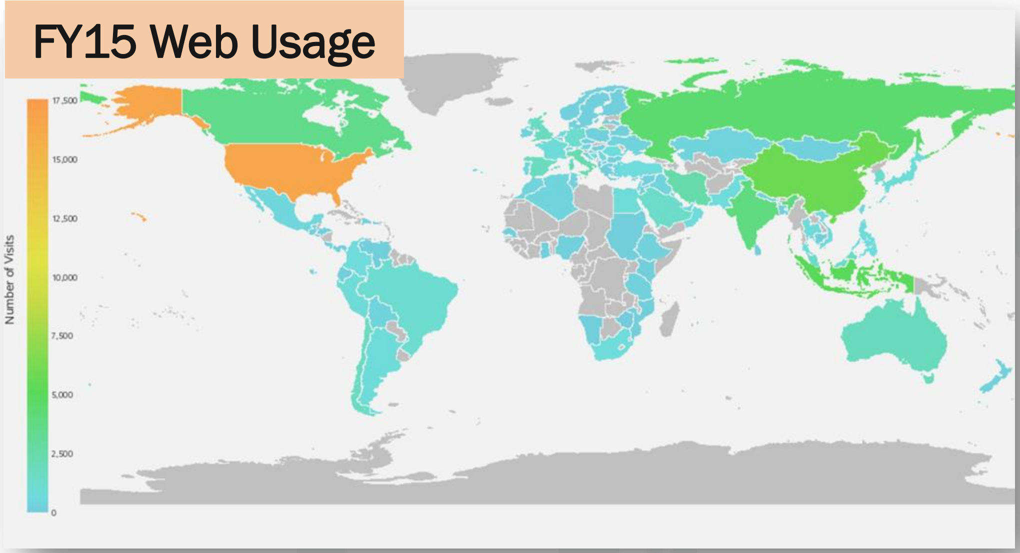


EARTH SCIENCE DATA SYSTEMS

ESDS Metrics

FY2015 Metrics	
Unique Data Products	9,462
Distinct Users of EOSDIS Data and Services	2.6 M
Web Site Visits of 1 Minute or more	2.4 M
Average Daily Archive Growth	16 TB/day
Total Archive Volume (as of Sept. 30, 2014)	14.6 PB
End User Distribution Products	1.42 B
End User Average Daily Distribution Volume	32.1 TB/day





https://earthdata.nasa.gov/about/system-performance

EOSDIS Common Services/Tools

- **Earthdata:** The EOSDIS website https://earthdata.nasa.gov will increase visibility to the interdisciplinary use of data and demonstrate how data are used.

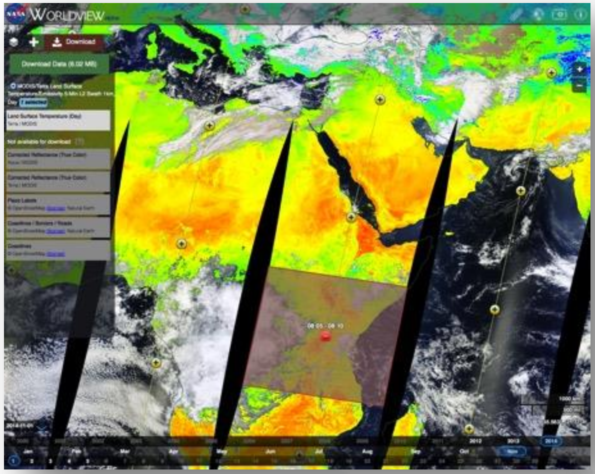
- High Performance Data Search and Discovery
  - **Common Metadata Repository (CMR):** Provide sub-second search and discovery services across the Sentinel and other EOSDIS holdings.
  - **Earthdata Search Client:** Data search and order tool https://search.earthdata.nasa.gov

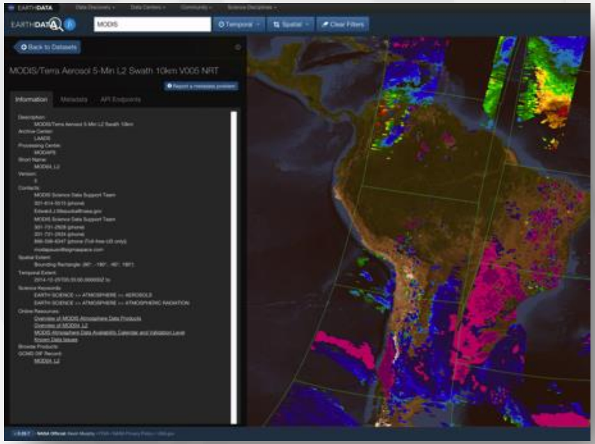
- Imagery and Data Visualization Tools
  - **Global Imagery Browse Services (GIBS)**: full resolution imagery in a community standards-based set of imagery services
  - **Worldview**: highly responsive interface to explore GIBS imagery and download the underlying data granules https://earthdata.nasa.gov/labs/worldview/
  - ***Giovanni**: Quick-start exploratory data visualization and analysis tool

- **EOSDIS Metrics System (EMS):** collects and reports on data ingest, archive, and distribution metrics across EOSDIS

- **ESDIS Standards Office:** assists in formulating standards policy for NASA's ESDS, coordinates standards activities within ESDIS, and provides technical expertise and assistance to standards related tasks

- **User Support Tool (UST):** user relationship management and issue resolution (Kayako)





EARTH SCIENCE DATA SYSTEMS

Open Sourcing Core EOSDIS Software

- ESDS instituted an Open Source policy in Fall 2015 for all new software development projects and existing core software components
  - Common Metadata Repository
  - Earthdata search client
  - Global Imagery Browse Services/Worldview
  - Giovanni

#1 Competitive Awards and Sustained Investments

#2 Programmatic Review

EVOLVING CAPABILITIES

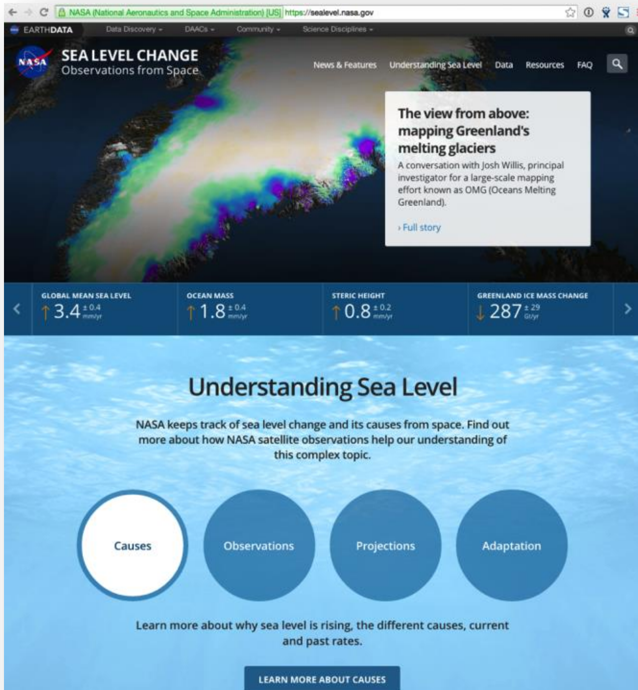
Continuous Evolution

#1 - Data system and science products are **evolved** through a combination of **competitive awards and strategic investments**.

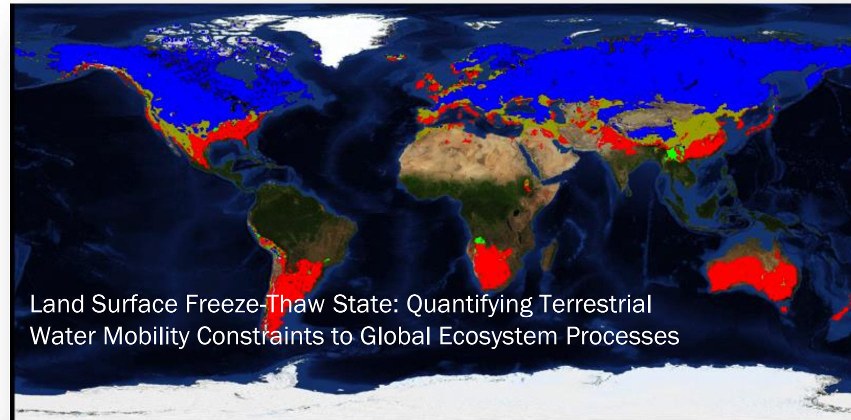
- Competitive awards provide a constant source of innovation to improve data products, advance systems and nurture interdisciplinary tools.
- Sustained investments increase efficiency and interoperability of EOSDIS.
  - Infrastructure - Networks, Virtualization, Storage, Backups
  - Systems - Global Imagery Browse System, Land and Atmosphere Near-real time Capability, Common Metadata Repository

#2 - Significant evolutionary activities are initiated as a result of program reviews.

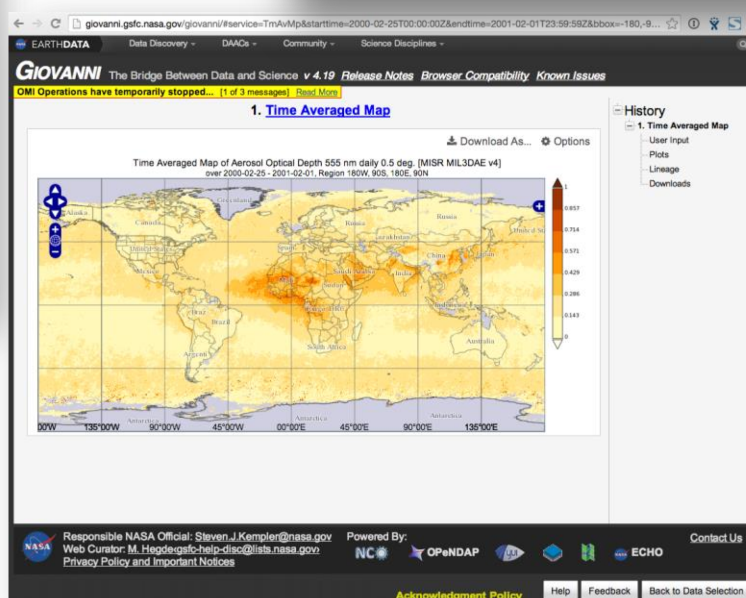
#1- Examples of Competitive Awards



https://sealevel.nasa.gov/



Land Surface Freeze-Thaw State: Quantifying Terrestrial Water Mobility Constraints to Global Ecosystem Processes



Federated Giovanni for Multi-Sensor Data Exploration

EARTH SCIENCE DATA SYSTEMS

#2 - 2015 Program Review

- In 2015, NASA proactively tasked the Earth Science Division (ESD)/ Earth Science Data Information Systems (ESDIS) to sponsor an *independent* holistic review (with appropriate stakeholders including ESD R&A, ESD Applications, ESDIS, Science community leaders and Mission stakeholders, as well as a senior representative from the NASA CIO community) to **study potential efficiencies and enhanced capabilities** from a variety of perspectives:
  - Based on science discipline
  - Based on optimizing common data operation tasks across DAACs

- **Consider advancing current efforts to achieve efficiencies across DAACs, including cloud computing, open source software, and dataset interoperability**.

#2 – Program Review Recommendations

1. ESDS should develop, implement and report on the outcome of **prototypes to explore the advantages, risks, and costs of using commercial cloud environments for**:
   - Storage and data transfer
   - Processing (e.g., advantages and effects of elastic computing on cost and speed for processing and reprocessing);
   - Improved data access (including exploration, selection, reduction, data analytics), and the feasibility associated with enabling data analytics/on-demand processing (e.g., bringing user processing/algorithms closer to the data). An important goal would be enabling user work across multiple large data sets managed by different DAACs without the need to transmit data over networks.

2. The ESDIS Project should develop, execute and report on efforts to:
   - **Develop a common software requirement and development strategy** across DAACs;
   - **Work toward a uniform user experience across DAACs** based on design/development best practices;
   - **Implement common data services and APIs** for uniform search, discovery, access to data (common APIs should augment necessary discipline specific tools);
   - **Consolidate and simplify duplicative software tools**;

3. **NASA should develop more streamlined approaches and processes to allow Open Source software development throughout the agency.** Adequate resources need to be applied to ensure that NASA-developed software can be open sourced on a timeline that facilitates collaboration and reuse.

Started March 2016

#2 – Near and Longer-Term Cloud Prototypes

- Near-Term – within 18 months
  - Evaluate feasibility of operating DAACs and EOSDIS core services in Amazon Web Services (AWS).
    - #2.1 NASA-Compliant General Application Platform (NGAP) Cloud Software Infrastructure
    - #2.2 Core DAAC Functions + Processing
    - #2.3 Application Hosting
    - #2.4 Commercial Cloud Partnerships

- Longer-Term (pending feasibility of near-term prototypes)
  - Tools for improved data access for exploration, selection, reduction and data analytics
  - On-demand processing (i.e., bringing user processing/algorithms closer to the data)

Terra Incognita

1. Vendor Lock-in
2. Future storage costs
3. Uncapped egress costs
4. Security restrictions
5. Network trust

EARTH SCIENCE DATA SYSTEMS

NASA

Commercial Cloud Partnerships

- Collaborative effort between Amazon, Google, Microsoft and NASA

- Goal:
  – Enable efficient access and transfer of NASA's Earth science data to different cloud infrastructures to facilitate new data driven applications and foster new user communities

Recommendations to improve downloads of large Earth observation data

Contributors:

Amazon: *Jamie Baker, Jamie Kinney, Ariel Gold, Jed Sundwall, Mark Korver*

Google: *Allison Lieber, William Vambenepe, Matthew Hancher, Rebecca Moore, Tyler Erickson*

NASA: *Kevin Murphy, Rahul Ramachandran, Chris Lynnes, Katie Baynes*

Microsoft: *Josh Henretig, Brant Zwiefel, Heather Patrick-Ahlstrom*

Draft Modified - July 20, 2016 Rahul Ramachandran
Draft Modified – Aug 2, 2016 Kaylin Bugbee
Final Version – August 8, 2016 Rahul Ramachandran

Future Architectures

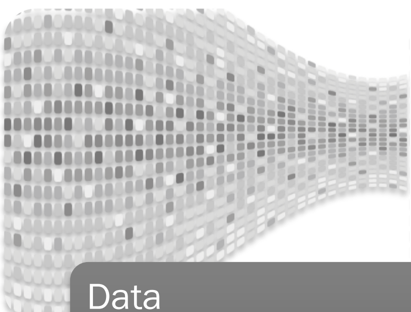
DISCUSSION

What could a future data system architecture look like?

- EOSDIS works well, but can we do better?
  - *Can we evolve NASA archives to better support interdisciplinary earth science?*
  - *Will a commercial cloud environment allow holdings to be more interactive and accessible to a wider group of users?*
  - *How will data from multiple agencies, international partners and the private sector be combined to study the earth as a system?*
    - *GOES-R, CubeSats, Copernicus…*

Increasing need for open science

- The foundation of the scientific method is the ability to repeat experiments
  - Data generation and analysis are increasingly tied to complex, difficult to access algorithms
  - High impact journals increasingly require access to data and source code as a requirement for publication
- Open data, open source code and methodologies improve:
  - Development speed and quality through collaboration
  - Reusability of tools
  - Transparency of underlying methodologies
  - Integrity of the peer review process



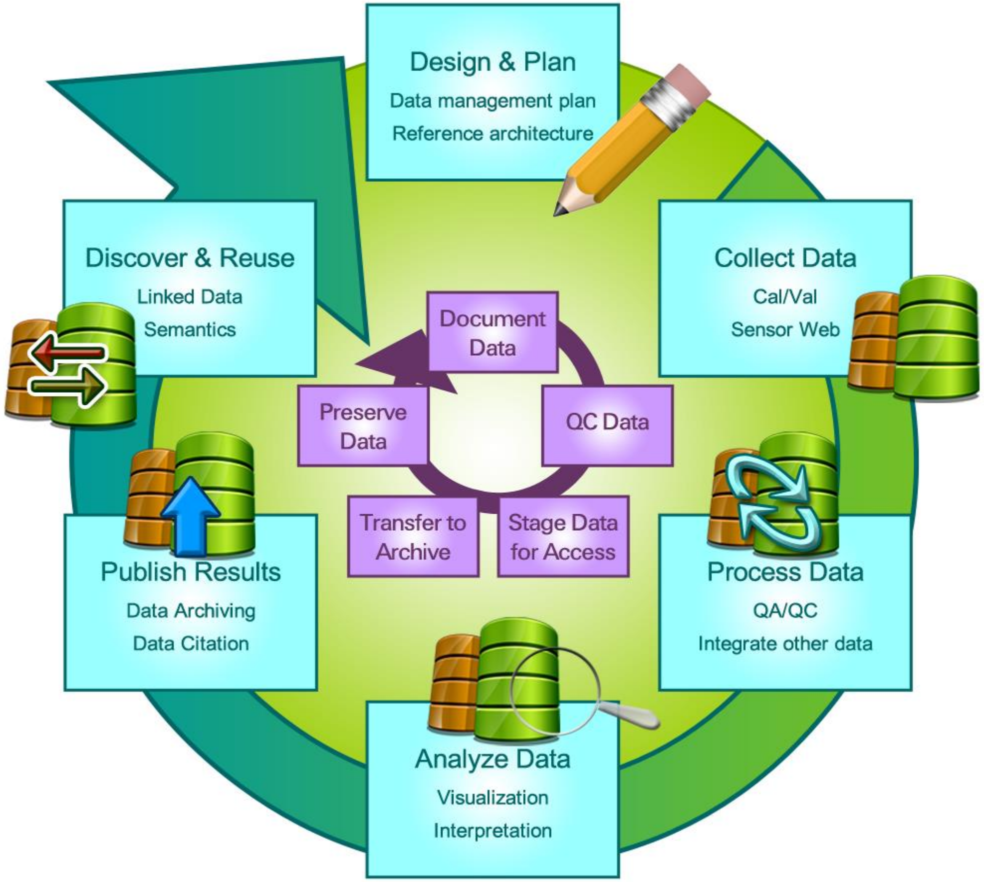
Data
- Free and accessible
- Interoperable through open standards

Source Code
- Open licensing
- Continuous agile development

Methodology
- Reusable work flows
- Collaborative environments

Data Systems in 2026

- In 2026 the Earth Science Data System program could be to make the *knowledge generation process frictionless* for researchers, address existing gaps/hurdles, seamlessly integrate new evolving technology, and also enable new research capabilities
- Drivers:
  - New data streams
  - End user expectations
  - Policy requirements
  - New technology



The Research Knowledge Generation Lifecycle. The inner cycle is the foundational data lifecycle, which is an integral aspect of the outer knowledge generation lifecycle.

EARTH SCIENCE DATA SYSTEMS

BACKUP

Earth Science Data and Information Policy

Since 1994

- NASA commits to the **full and open sharing of Earth science data** obtained from NASA Earth observing satellites, sub-orbital platforms and field campaigns with all users **as soon as such data become available**.
- There will be **no period of exclusive access** to NASA Earth science data. Following a post-launch checkout period, all data will be made available to the user community. Any variation in access will result solely from user capability, equipment, and connectivity.
- NASA will make available all **NASA-generated standard products** along with the **source code** for algorithm software, **coefficients**, and **ancillary data** used to generate these products.
- All NASA Earth science missions, projects, and grants and cooperative agreements shall include **data management plans** to facilitate the implementation of these data principles.
- NASA will enforce a **principle of non-discriminatory data access so that all users will be treated equally**. For data products supplied from an international partner or another agency, NASA will restrict access only to the extent required by the appropriate Memorandum of Understanding (MOU).
- **http://science.nasa.gov/earth-science/earth-science-data/data-information-policy/**

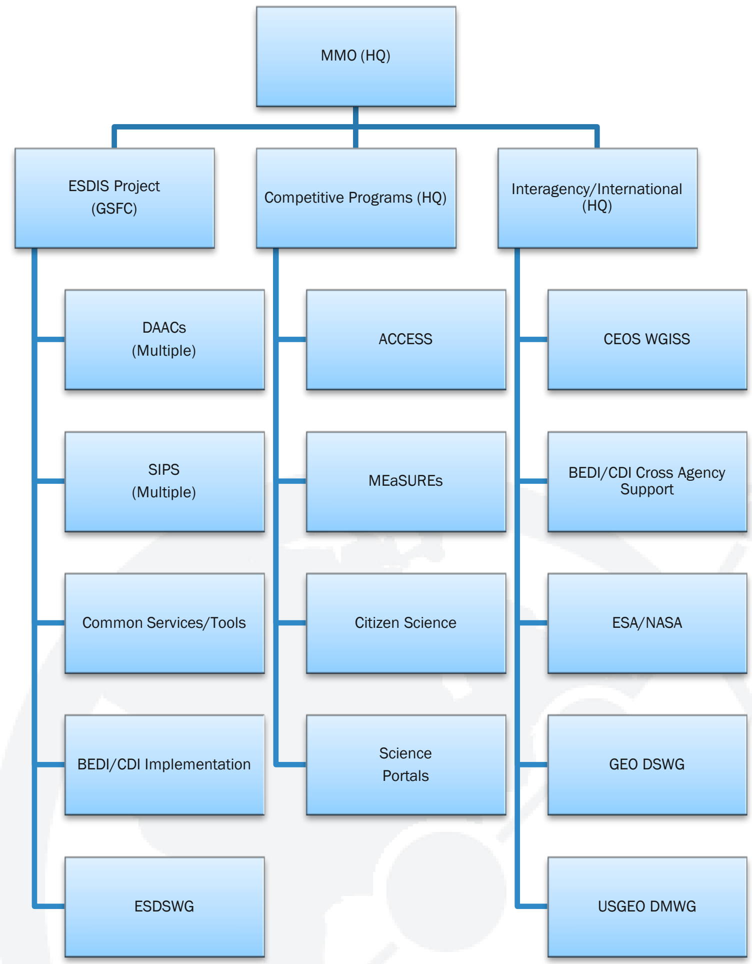
Full and Open data are the pillar of repeatable science – Including software and algorithms used to generate the products and perform analysis

Major Components of the Earth Science Data Systems Program (MMO)

- **Earth Observing System Data and Information System (EOSDIS)** – Managed by the ESDIS Project
  - Science Investigator-led Processing Systems (SIPS)
  - Distributed Active Archive Centers (DAACs)
  - Earth Science Data System Working Group (ESDSWG)

- **Competitive Programs**
  - Making Earth System Data Records for Use in Research Environments (MEaSUREs)
  - Advancing Collaborative Connections for Earth System Science (ACCESS)
  - Citizen Science for Earth System Science
  - Science Portals

- **International and Interagency Coordination and Development**
  - CEOS Working Group on Information Systems and Services (WGISS)
  - Office of Science and Technology Policy Climate Data Initiative (CDI) and Big Earth Data Initiative (BEDI)
  - NASA-European Space Agency (ESA) Bilateral
  - Group on Earth Observation (GEO) Data Sharing Working Group(DSWG)
  - USGEO Data Management Working Group (DMWG)



Earth Science Data System Goals

DAAC and Core System Perspective

- Set the standard for the archival and delivery of science-quality data related to planet Earth.
  - *No data loss with resilient backups.*
  - *Easily add new data products at lower cost (long-tail, aircraft, in situ).*
  - *Data available in all systems (download, metadata, visualization, analysis, etc.) as soon as it is ingested.*

- Provide consistent harmonized services for data, metadata and services to facilitate interdisciplinary earth science investigations.
  - *Standard levels of service including system performance (operator and user perspectives), uptime, and data set interoperability.*
  - *Metadata consistently available through all interfaces and continually updated.*
  - *Quickly deploy new technologies across the enterprise (process/infrastructure/features).*

- Ensure data is documented precisely and fully, and access to data and services are useful and useable by a wide community of users.
  - *Ubiquitous DOIs and linkages to citations and articles.*
  - *Translations among science nomenclatures.*
  - *Quality and error characterizations (especially format and projection conversions).*
  - *Engaging earth science and computer/developer content on websites and tools guided by UI/UX*

- Improve data system capabilities through continual evolution and community engagement.
  - *Adopt and develop new technologies to increase efficiency of archival, distribution and services.*
  - *Open source **ALL** software.*
  - *Work with Science Teams and PIs, UWGs, ESDSWG, ESIP and presentations at conferences.*
  - *ACCESS Program (..reimagined..), Citizen Science and competitively selected science portals.*

Understanding User Needs and Assessing Performance

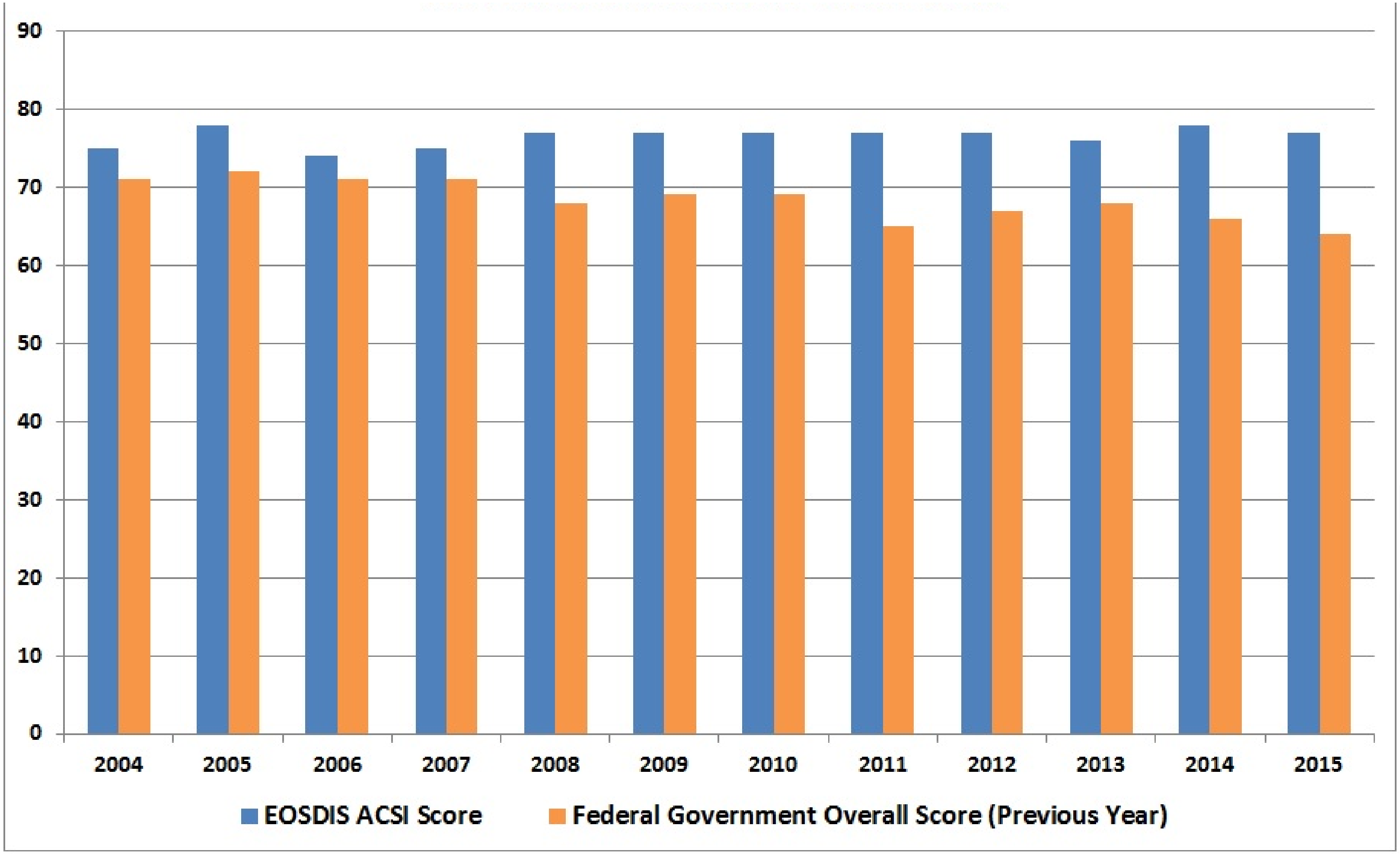
- DAAC User Working Groups – Provide assessments and recommendations based on unique DAAC mission requirements

- DAAC Customer Satisfaction
  - Annual Online survey of all DAAC users to evaluate satisfaction and measure performance
  - Performed by CFI Group, the American Customer Satisfaction Index (ACSI) is the #1 national indicator of customer satisfaction for more than 225 companies and 130 Federal programs

- EOSDIS Metrics System – collects metrics on ingest, archive and distribution for evaluation of system performance.
  - Enables ESDIS to characterize use of the EOSDIS, and report to NASA Headquarters and OMB.

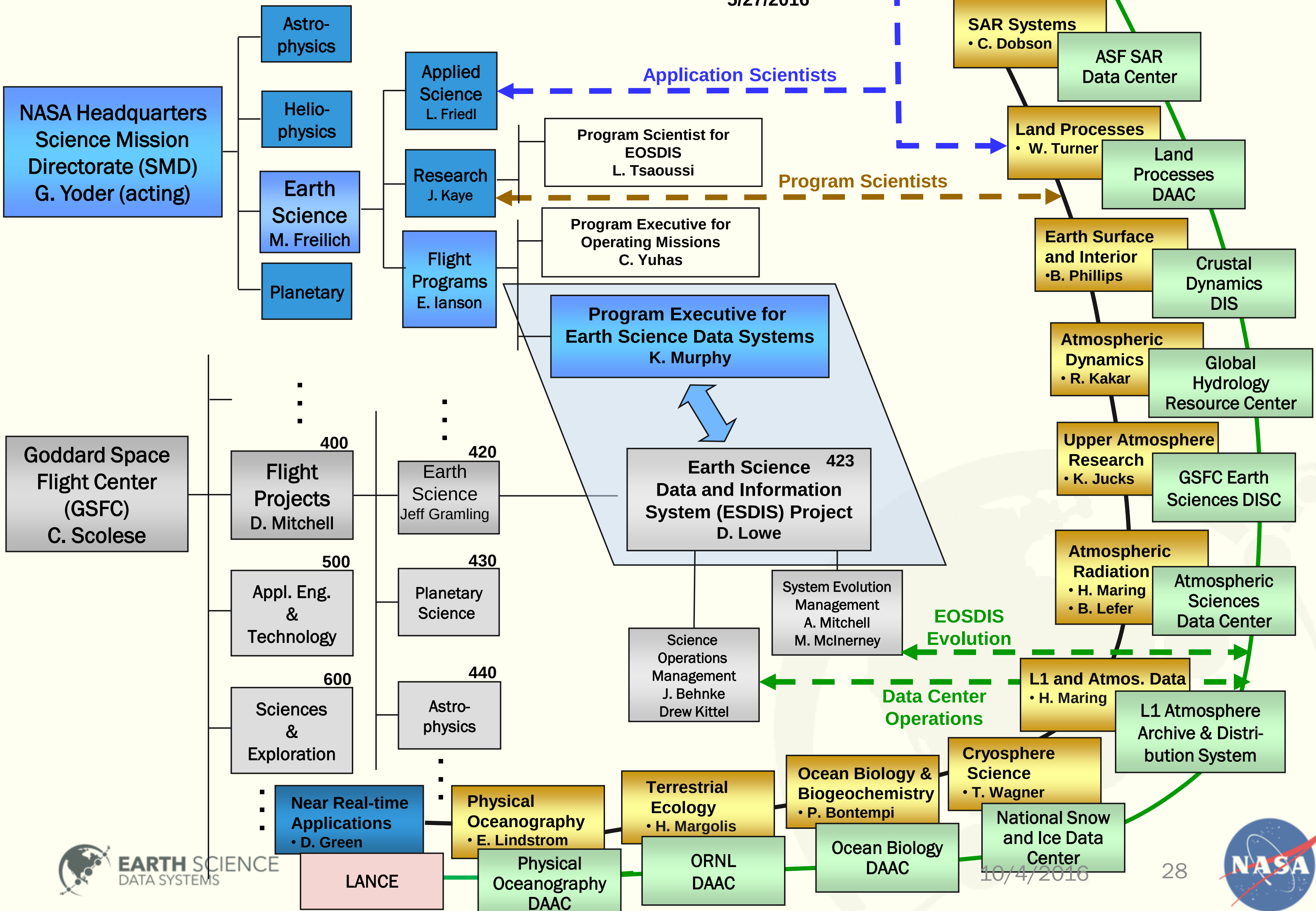
- User Services Working Group – DAAC User Services personnel work together to best service science communities
  - User feedback – via Kayako
  - Personal interaction with users

EOSDIS ACSI Scores Consistently Exceed the Federal Government Overall Score

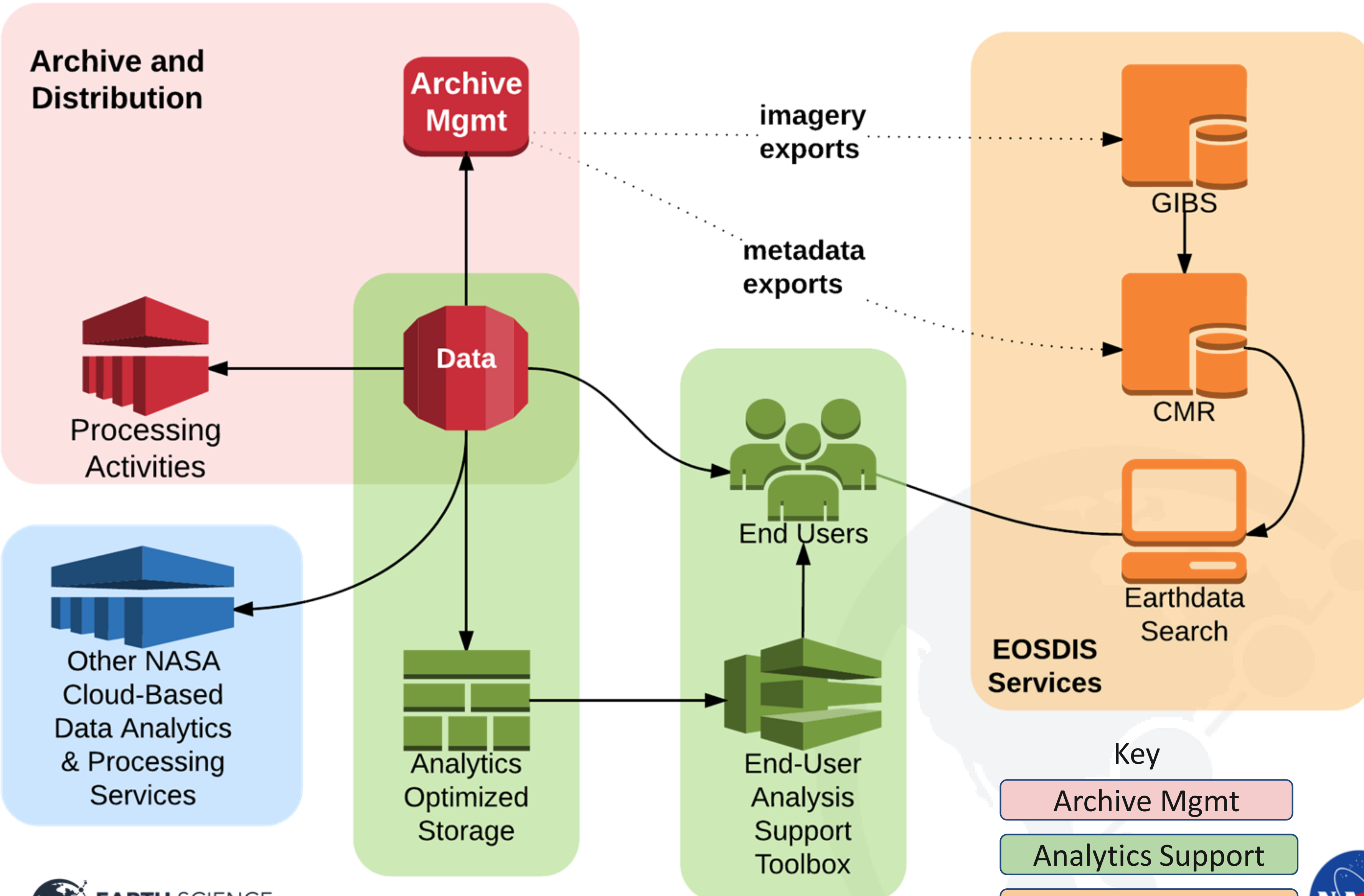


Earth Science & Data Systems

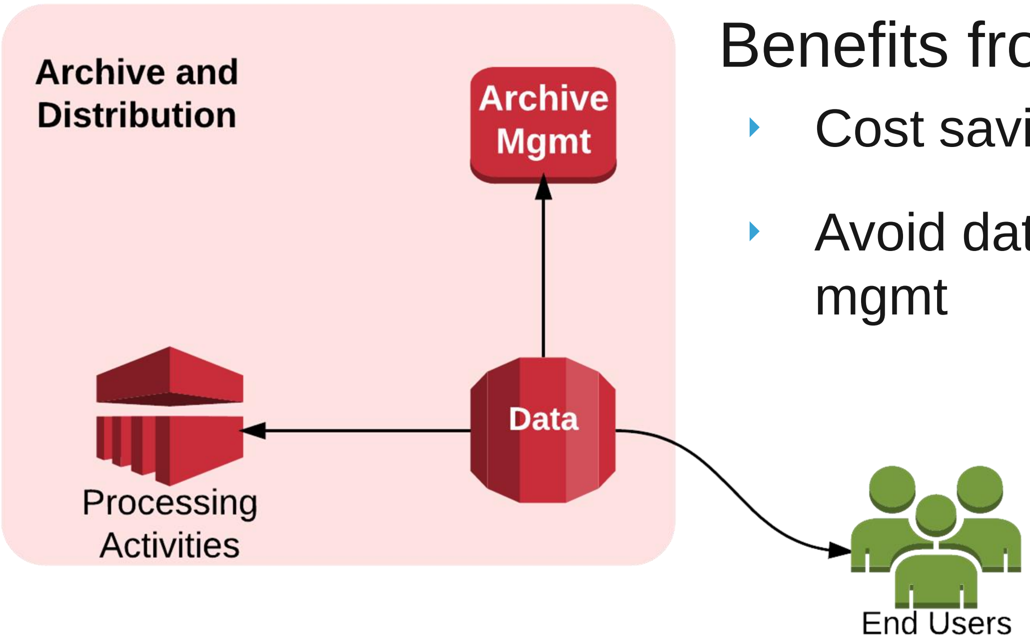
5/27/2016



Cloud Prototypes



Archive Cloud Prototypes

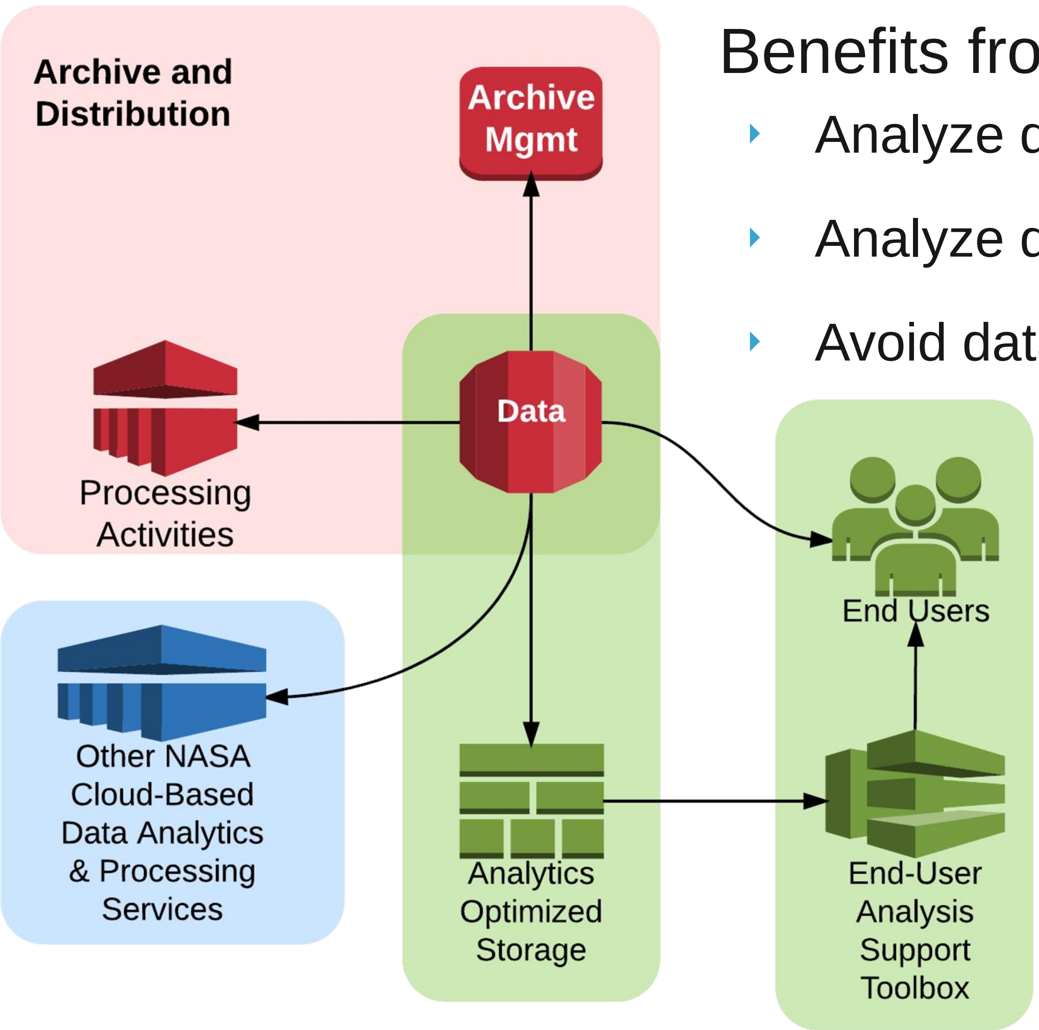


Benefits from Archive in the Cloud

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

- Alaska Satellite Facility Web Object Storage prototype
  - Distribute Sentinel radar data from Amazon storage
- Global Imagery Browse Service in the Cloud
- Ingest and Archive management prototype

Cloud Analytics Prototypes



Benefits from Cloud Analytics

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

Analysis support toolbox to attract users to cloud analytics

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