ESDIS TCAT Experience

E. Tauer – JPL / PO.DAAC

February 4th, 2016

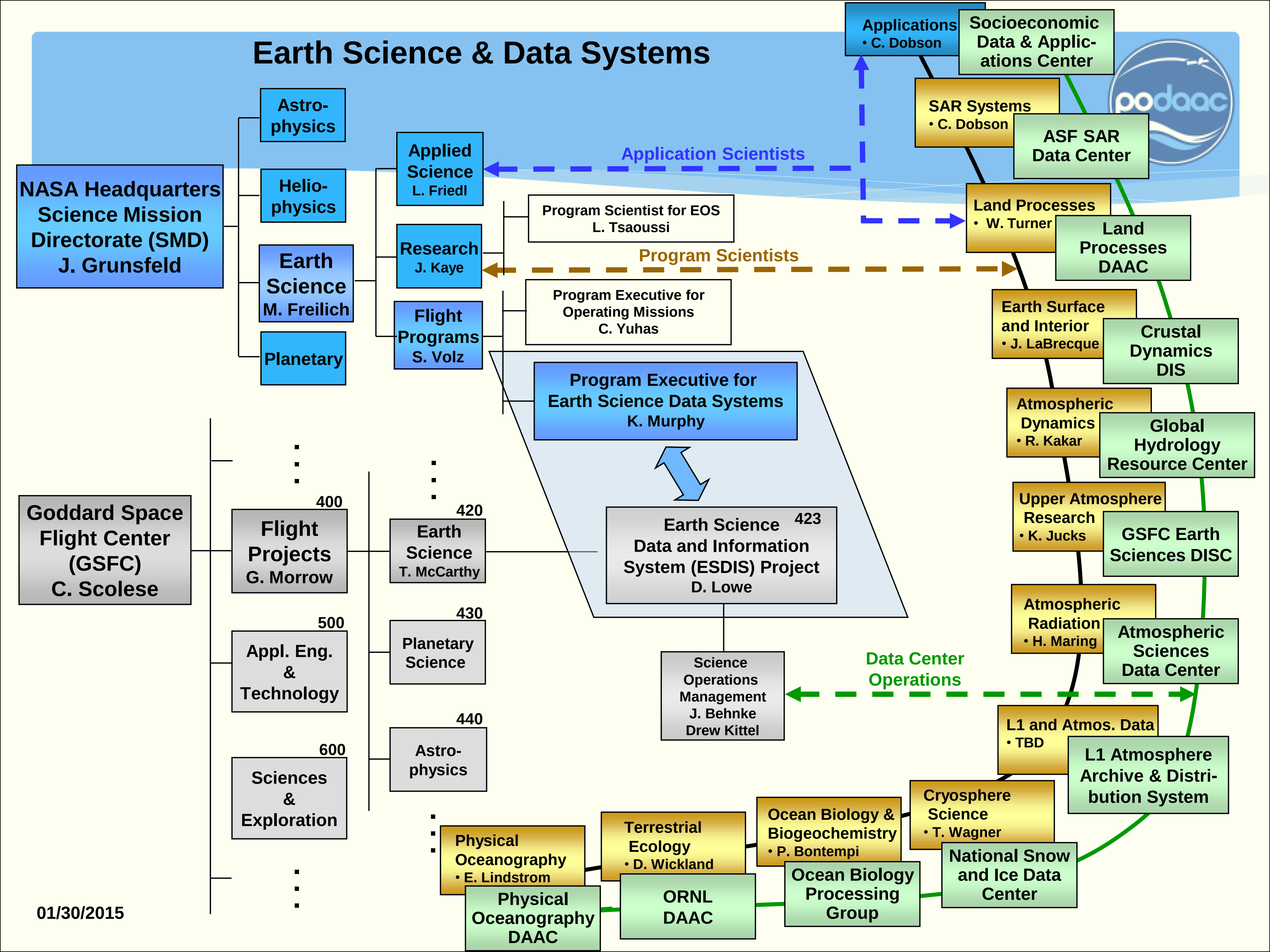
Talking points

* Executive summary
* Context of the review subject: ESDIS DAACs
  * Overview of ESDIS
  * Overview of a DAAC
* TCAT vs. the E&E Review team
* Recommendations (1, 2, and 3)
* Context of the solution space:
  * Breadth of DAAC capabilities
  * Example of a potential response
* A couple lessons learned / parting thoughts

Executive Summary

* Technical Capabilities Assessment Team (TCAT) formed in early 2014
* Analysis ("deep dive") of the DAACs spanned a one year period
* Result: Assigned ESD/ESDIS to sponsor an independent review team
* E&E Review Team formed in April 2015
* Three findings/recommendations submitted to ESD in August 2015
  1) Develop prototypes for using cloud environments
  2) Implement common data services and development strategies
  3) Improve processes related to open source software

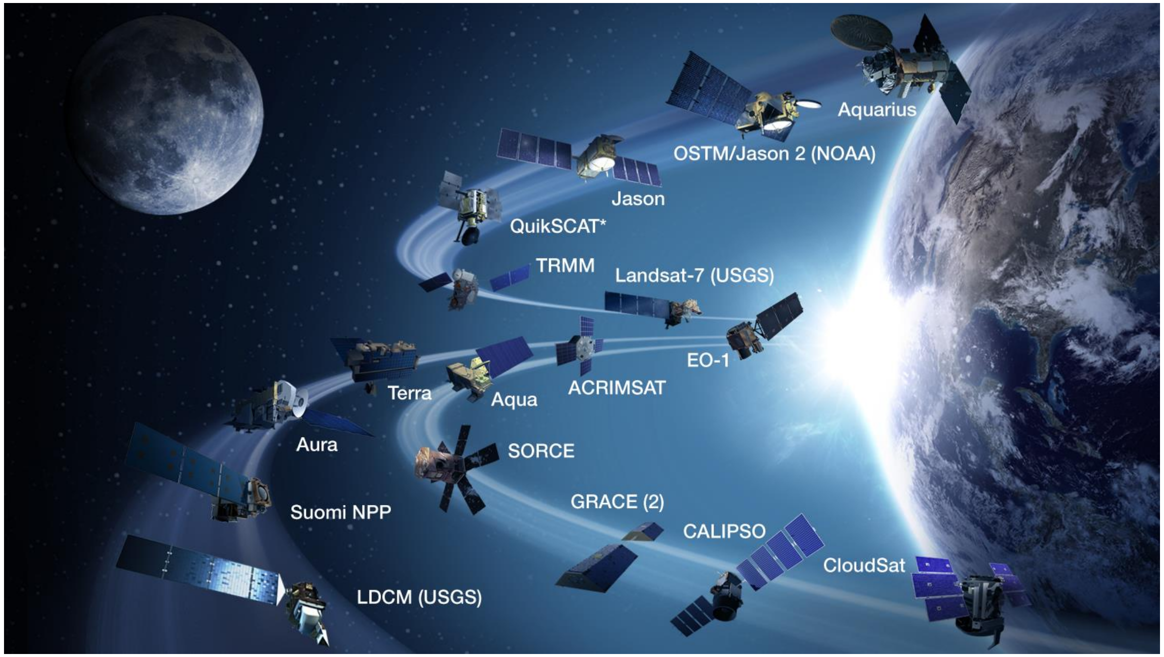
ESDIS and the DAACs


Earth Science & Data Systems
NASA Headquarters Science Mission Directorate (SMD) J. Grunsfeld
Astro-physics
Helio-physics
Earth Science M. Freilich
Planetary
Applied Science L. Friedl
Research J. Kaye
Flight Programs S. Volz
Program Scientist for EOS L. Tsaoussi
Program Executive for Operating Missions C. Yuhas
Program Executive for Earth Science Data Systems K. Murphy
Application Scientists
Program Scientists
Data Center Operations
Applications • C. Dobson
Socioeconomic Data & Applic-ations Center
SAR Systems • C. Dobson
ASF SAR Data Center
Land Processes • W. Turner
Land Processes DAAC
Earth Surface and Interior • J. LaBrecque
Crustal Dynamics DIS
Atmospheric Dynamics • R. Kakar
Global Hydrology Resource Center
Upper Atmosphere Research • K. Jucks
GSFC Earth Sciences DISC
Atmospheric Radiation • H. Maring
Atmospheric Sciences Data Center
L1 and Atmos. Data • TBD
L1 Atmosphere Archive & Distri-bution System
Cryosphere Science • T. Wagner
National Snow and Ice Data Center
Ocean Biology & Biogeochemistry • P. Bontempi
Ocean Biology Processing Group
Terrestrial Ecology • D. Wickland
ORNL DAAC
Physical Oceanography • E. Lindstrom
Physical Oceanography DAAC
podaac
Goddard Space Flight Center (GSFC) C. Scolese
400 Flight Projects G. Morrow
500 Appl. Eng. & Technology
600 Sciences & Exploration
420 Earth Science T. McCarthy
430 Planetary Science
440 Astro-physics
423 Earth Science Data and Information System (ESDIS) Project D. Lowe
Science Operations Management J. Behnke Drew Kittel
01/30/2015

Earth Science at NASA

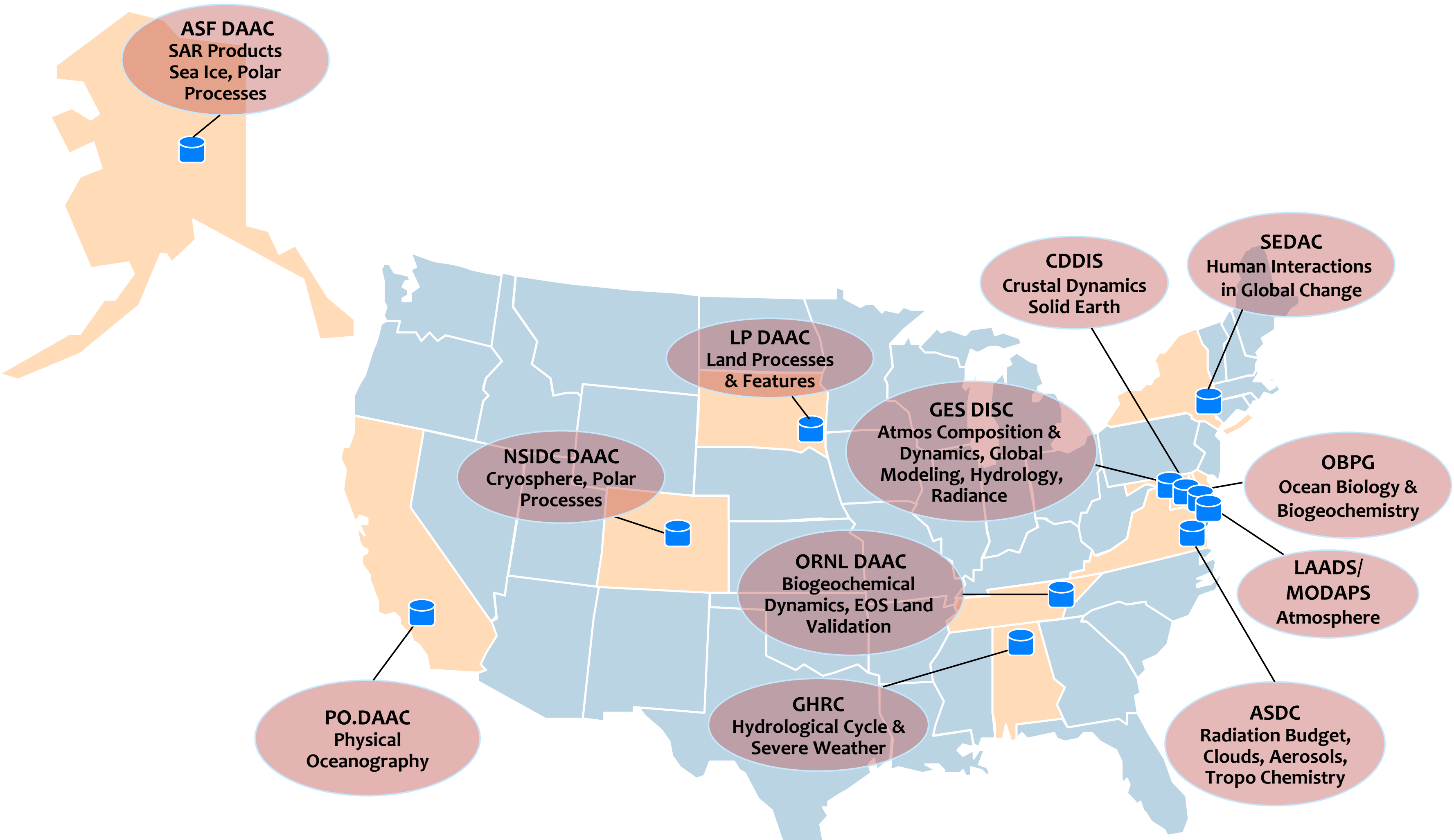
Earth Science Disciplines
Cryospheric science
Land use/land cover change
Terrestrial ecology
Physical oceanography
Biological oceanography
Terrestrial hydrology
Precipitation science
Atmospheric modeling

27 Operating missions
30 Past missions
8 Under development
6 Under study

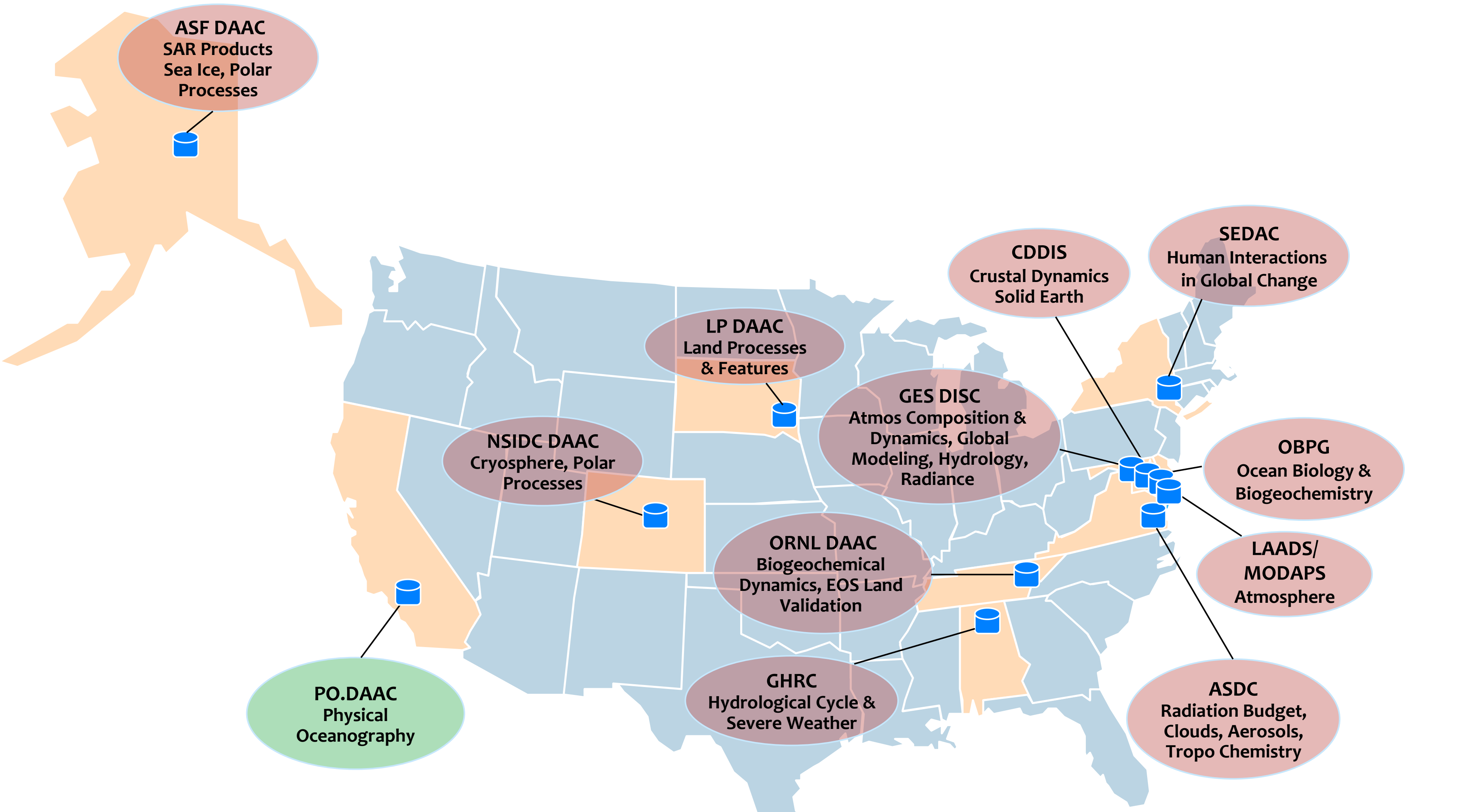


7.4 Petabytes of data (average growth: 5.4 TB/day)
1.5 Million distinct users of EOSDIS data and services

D = Distributed…

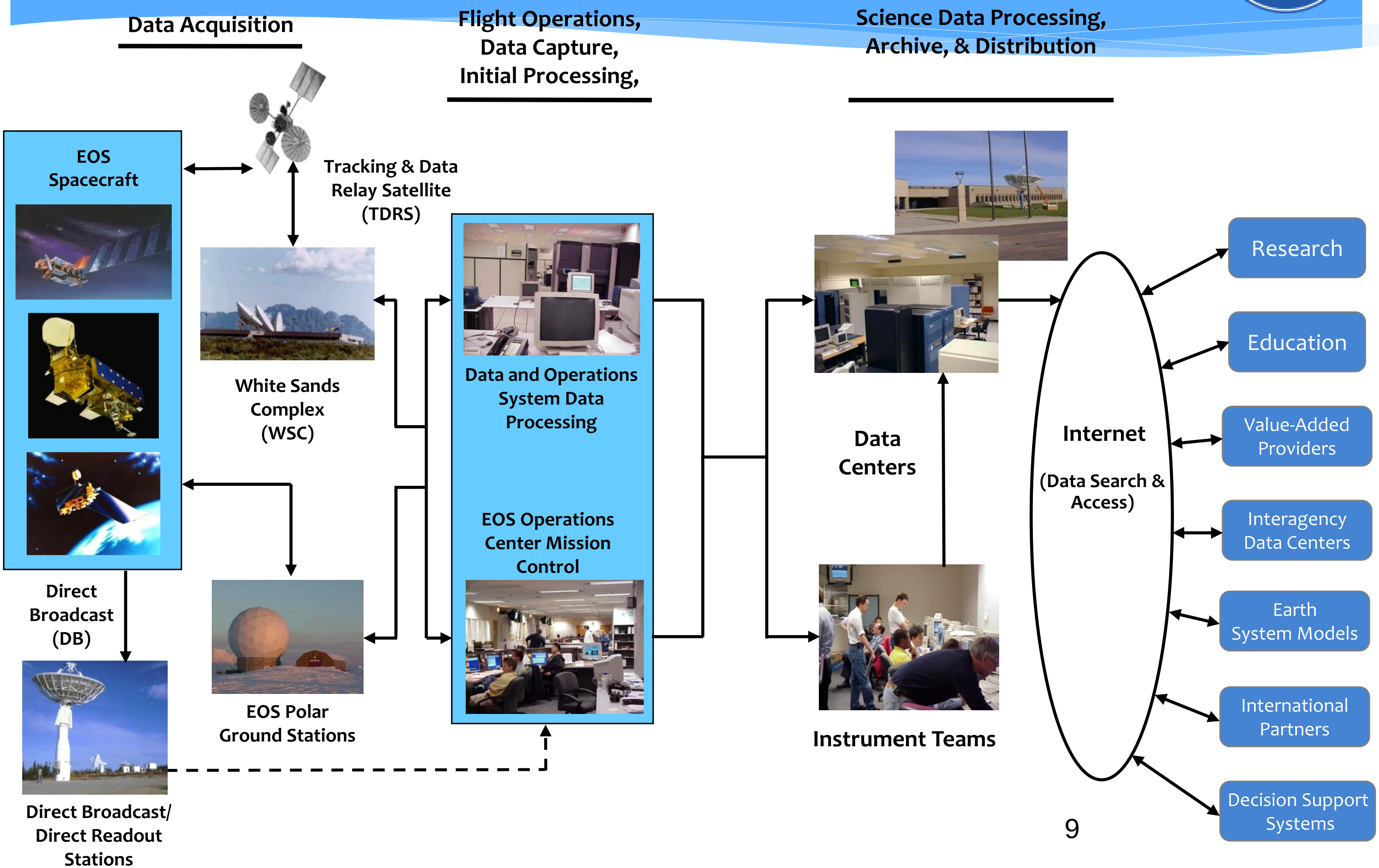


D = Distributed…



Mission Operations

Science Operations



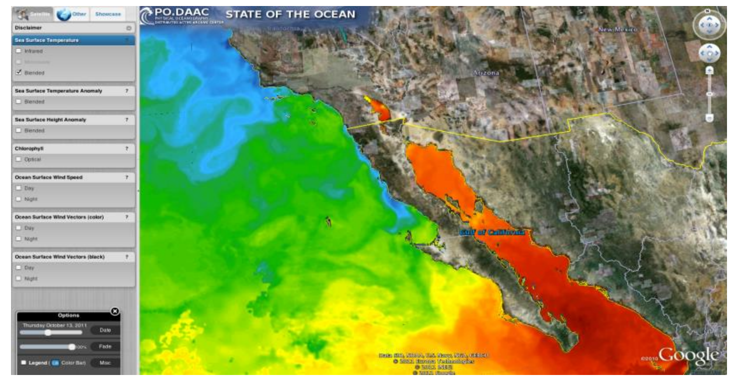
PO.DAAC Functional Areas

Data Management & Stewardship

Preserve NASA's data for the benefit of future generations

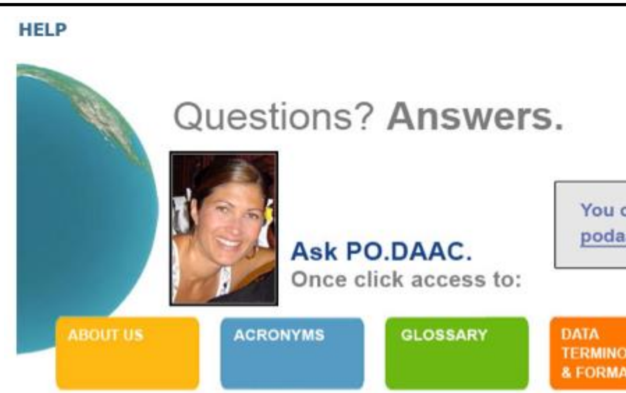
Data Access

Provide intuitive services to discover, select, extract and utilize data

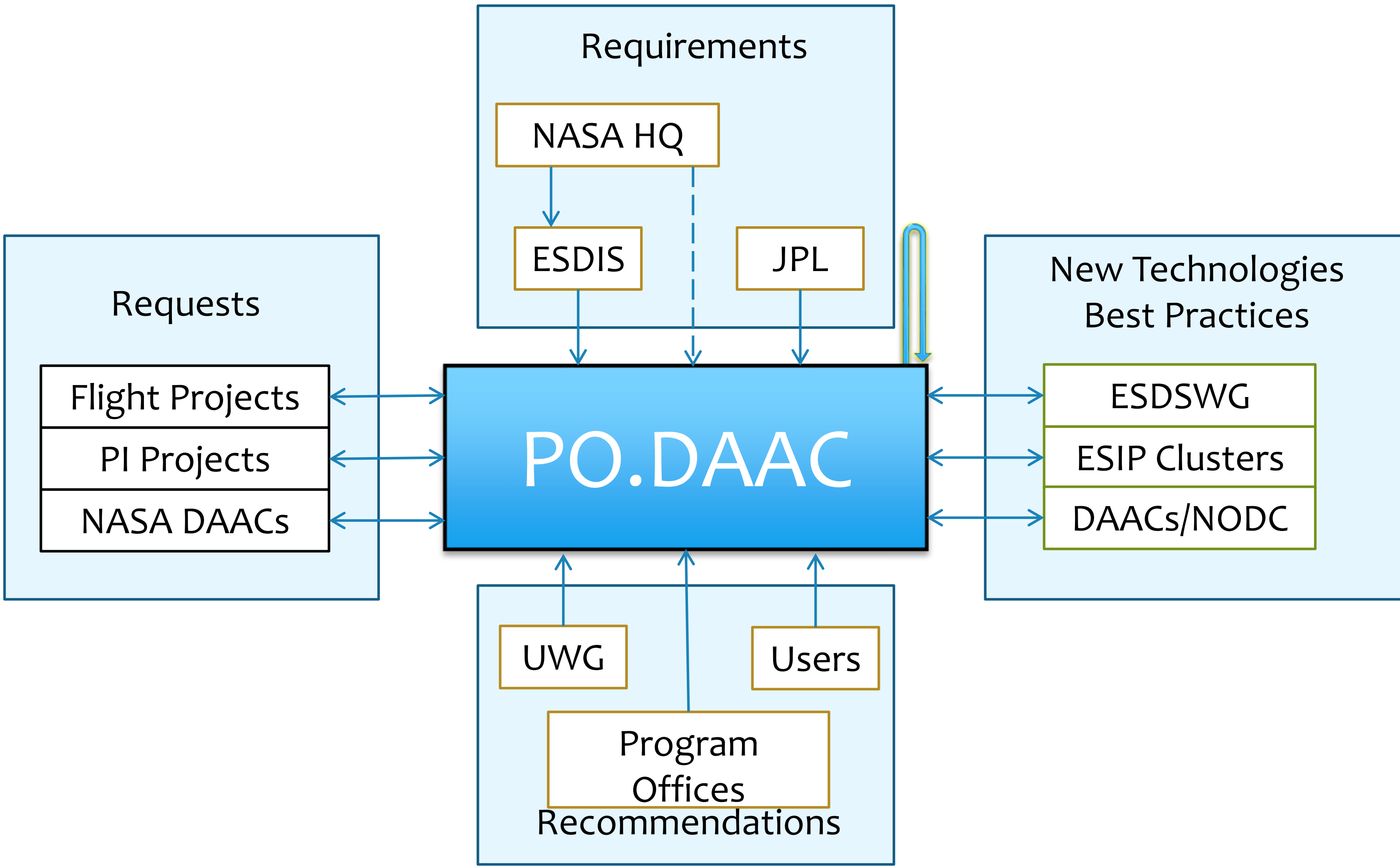


Science Information Services

Provide a knowledgebase to help a broad user community understand and interpret satellite ocean data and related information



PO.DAAC Interfaces and Stakeholders



TCAT and the E&E

Technical Capability Assessment Team (TCAT)

- Technical Capabilities Assessment Team formed in early 2014
  - https://nbat.hq.nasa.gov/tcat/index
- Purpose was to arrive at a more efficient operating model without compromising core functions or necessary technologies
  - Question of core functions (what qualifies) is important
  - Question of what's necessary ("Given the volatility of national space policy") is important
- Did a number of deep dives, most relevant to us: "Long term data management, as a NASA solution"
- TCAT's analysis of the DAACs spanned a one year period
- Initial findings sparked interest, but there were gaps in knowledge that affected the relevance
- E&E Review Team formed to have a closer look

From TCAT to E&E

* ESD/ESDIS was assigned to sponsor an independent review team formed to "study potential efficiencies and enhanced capabilities"
* E&E Review Team formed April 20, 2015
  * *E&E = Evolution, Enhancement, Efficiency (choose any 2)*
* ESDIS/DAAC accomplishments were recognized
  * Strong relationship with science community
  * Consistently meets stewardship responsibility
  * Actively promotes early partnerships with data providers
* Three findings and associated recommendations were reported out to the DAACs on August 17, 2015

Recommendations

Findings & Recommendations #1

Finding: Commercial cloud environments offer potential for storage, processing, and operations efficiencies; improved cross-DAAC collaboration; and potential for new data access and service paradigms. The ESDIS Project and DAACs have initiated preliminary studies of relevant Cloud technology but these have been ad hoc. Substantial additional work, coordination, and investment is needed in this area to identify opportunities and risks.

Recommendation: The ESDIS Project 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 (e.g., assess benefits and impacts of "virtual data center", "virtual collections" and costs for distributing data from cloud environments);
- 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.

Findings & Recommendations #2

Finding: There are a plethora of tools across the DAACs with varying levels of usability. While many are discipline/data collection unique, there is also considerable overlap, redundancy, and duplication of effort. Redundancy of tools leads to user confusion.

Recommendation: The ESDIS Project should develop, execute and report on efforts to:

- 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;
- Develop a common software requirement and development strategy across DAACs;
- Work towards a uniform user experience across DAACs based on design/development best practices;

The objectives are to ensure the right tools are built; eliminate unnecessary duplication; build with reuse in mind; leverage common web services; release software via Open Source and improve system wide usability.

Findings & Recommendations #3

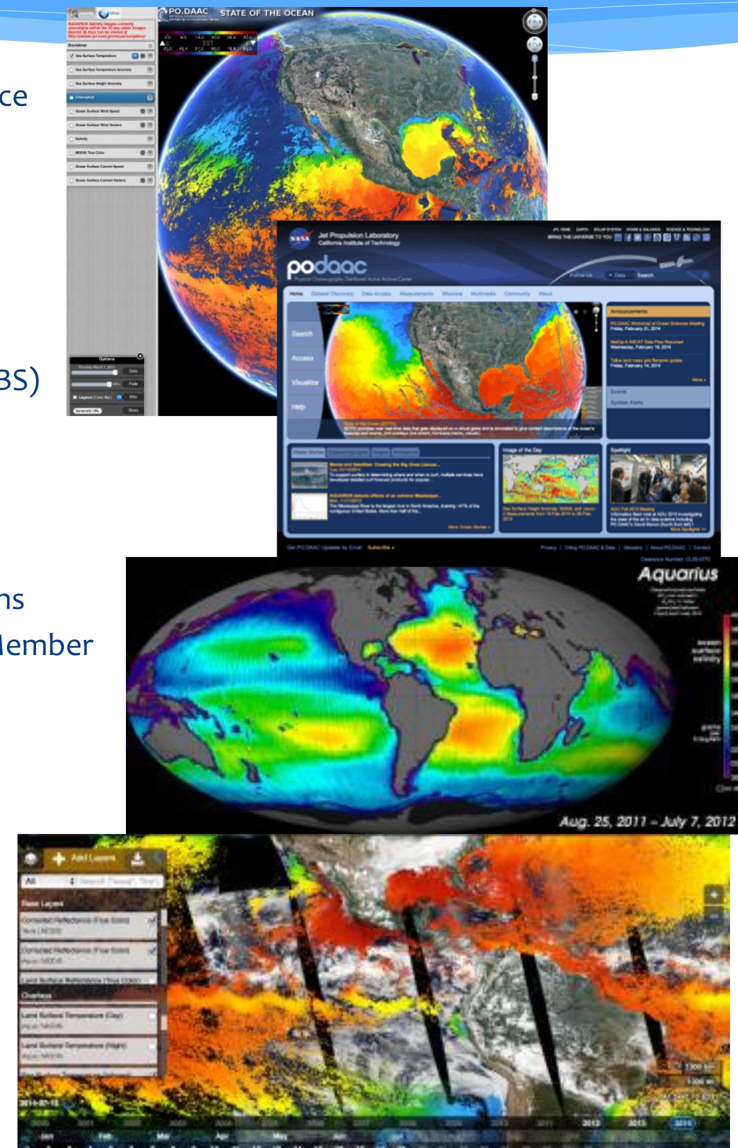
Finding: There is significant potential for increased collaboration across DAACs through open source software development. However the time and effort required to share software outside the original NASA development organizations, by open sourcing, can be prohibitive. Inability to share software increases cost and inhibits reproducibility of data products.

Recommendation: 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.

Solution Space

PO.DAAC: Science and Technology Leadership and Innovation

* **Science Community Leadership**
  * Expertise in gravity data sets, sea surface temperature, salinity, ocean surface topography, ocean currents, ocean winds, ocean circulation
  * Interface with global array of providers, partners, and user groups
  * Vital to both the Oceanographic and Climate communities
  * User and Science information services
* **Technology Leadership**
  * Advance ESDIS core technologies, e.g., Global Imagery Browse Services (GIBS)
  * Data discovery and access tools
    * Extensible Data Gateway Environment (EDGE)
    * Semantic Web for Earth and Environmental Terminology (SWEET)
  * State-of-the-Art Visualization technologies
  * Strong participation in technology proposals to further science investigations
  * Federation of Earth Science Information Partners (ESIP) long-term Voting Member
  * NASA Earth Science Data System Working Groups (ESDSWG)
    * Data-Intensive Architecture (DIA)
    * Data Quality
    * Visualization
    * Digital Object Identifier (DOI)
* **System Engineering/Data Architecture Leadership**
  * Cloud Computing and Big Data
  * End-to-end Data System lifecycles
  * Data Stewardship and Metadata standards
  * Modular processing pipelines for automation and efficiency



Cloud-based Analytics Platform for the Retrieval of Information (CAPRI)

* What is CAPRI (other than an island off the West coast of Italy) ?
  * PO.DAAC-owned HW/SW asset
  * Completely independent of PO.DAAC Operations facility
  * Technology incubator to:
    * Provide key support for PO.DAAC's Number One Priority
    * Empirically demonstrate tradeoffs in building a data analytics platform
    * Benchmark performance
    * Support PO.DAAC-related technology proposals
    * Respond to all 3 TCAT recommendations
* What is currently being done?
  * Use Cases being collected
  * Hardware being installed and configured
  * Experiments being defined

Parting Thoughts

Lessons Learned

* Soul searching is not a bad thing
  * Good to look closely at ourselves once in a while.
  * The recommendations (ultimately) were solid ideas worth examining.

* Innovation is worth fighting for
  * Avoiding redundancy is good; options are good, too.
  * So far, seems a healthy respect for innovation and competition has not been (completely) lost.

* Make sure folks “up the chain” understands who you really are.
  * Many thought DAACs were just a stack of disks.
  * Spent quite a bit of energy reminding people what a DAAC is all about.
  * Seems it finally got through; probably shouldn’t have been necessary to begin with.

BACKUP

DAAC Organizations

	DAAC Name	DAAC Acronym	DAAC Manager
1	Alaska Satellite Facility	ASF	Annette (Nettie) La Belle-Hamer
2	Atmospheric Science Data Center	ASDC	John Kusterer
3	Land Processes	LP DAAC	Chris Doescher
4	National Snow and Ice Data Center	NSIDC	Brian Johnson
5	Oak Ridge National Laboratory	ORNL	Suresh Vannan
6	Physical Oceanography	PO.DAAC	Rob Toaz
7	Socioeconomic Data and Applications Center	SEDAC	Bob Chen
8	Global Hydrology Resource Center	GHRC	Rahul Ramachandran
9	Ocean Biology Processing Group	OBPG	Gene Feldman
10	MODIS Adaptive Processing System Level 1 and Atmosphere Archive and Distribution System	MODAPS LAADS	Ed Masuoka
11	Crustal Dynamics Data Information System	CDDIS	Carey Noll
12	Goddard Earth Sciences Data and Information Services Center	GES DISC	Steve Kempler