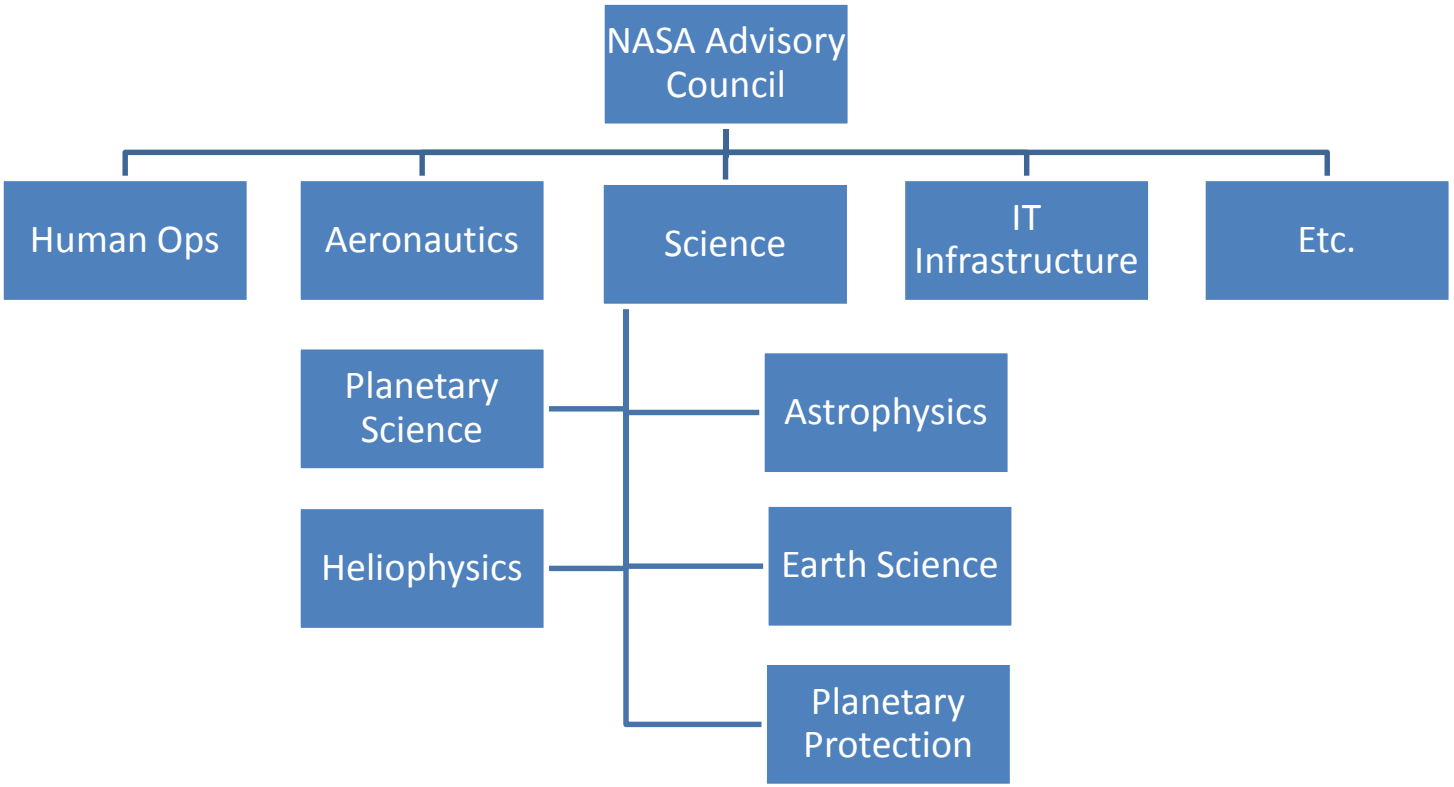
Big Data Task Force
Legacy from
NAC IT Infrastructure Committee

Charles P Holmes
Chair, BDTF
Formerly Vice-Chair, ITIC
February 16, 2016

Mandatory Flow Chart



NAC Structure 2010 - 2013

Last slide of the last presentation
By Chair of the ITIC Presentation to the NAC – Dec 12, 2013

NAC Committee on IT Infrastructure Recommendation #1 July 31, 2013

- Recommendation: The NASA NAC ITIC & Science Committees should collaboratively explore the existing and planned evolution of NASA's science data cyberinfrastructure that supports broad access to data repositories for NASA SMD missions. This exploration should be undertaken in the context of effective practices within NASA, other Federal agencies, as well as industry and research institutions.

Wording Agreed to by Both ITIC and Science Committees
July 31, 2013

Work Will Continue as Big Data Taskforce Under Science Committee

NAC Committee on IT Infrastructure Recommendation #1

- Recommendation: To enable NASA to gain experience on emerging leading-edge IT technologies such as:
  - *Data-Intensive Cyberinfrastructure,*
  - *100 Gbps Networking,*
  - *GPU Clusters, and*
  - *Hybrid HPC Architectures,*

  we recommend that NASA aggressively pursue partnerships with other Federal agencies, specifically NSF and DOE, as well as public/private opportunities.

  We believe joint agency program calls for end users to develop innovative applications will help keep NASA at the leading edge of capabilities and enable training of NASA staff to support NASA researchers as these technologies become mainstream.

NAC Committee on IT Infrastructure

DRAFT* Recommendation #2

- Recommendation: NASA should formally review the existing national data cyberinfrastructure supporting access to data repositories for NASA SMD missions. A comparison with best-of-breed practices within NASA and at other Federal agencies should be made.
- We request a briefing on this review to a joint meeting of the NAC IT Infrastructure, Science, and Education committees within one year of this recommendation. The briefing should contain recommendations for a NASA data-intensive cyberinfrastructure to support science discovery by both mission teams, remote researchers, and for education and public outreach appropriate to the growth driven by current and future SMD missions.

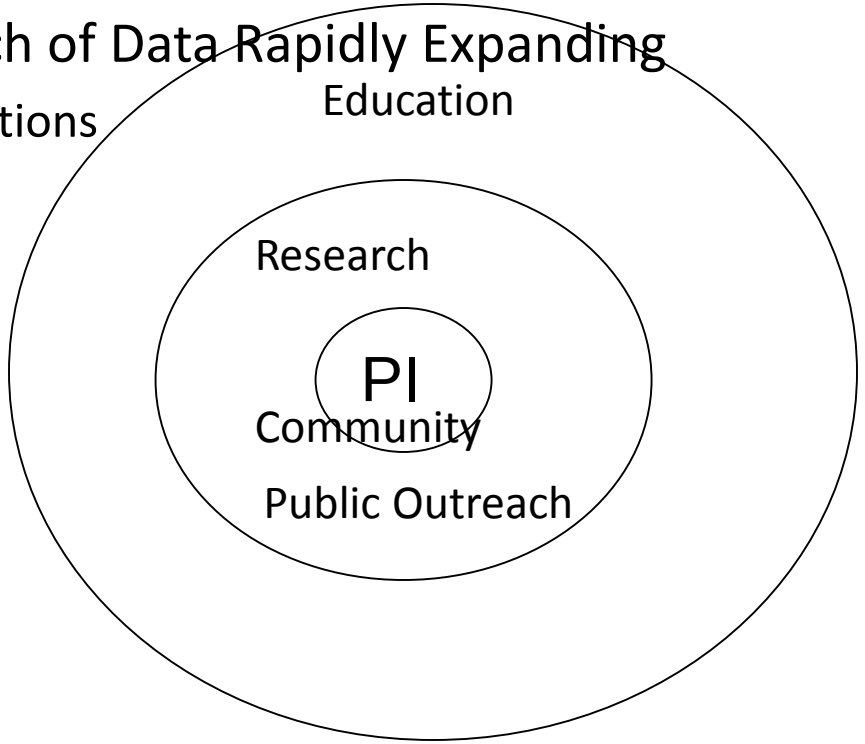
* To be completed after a joint meeting of ITIC, Science, and Education Committees in July 2012 and the final recommendation submitted to July 2012 NAC meeting

NAC Committee on IT Infrastructure Recommendation #2 (continued)

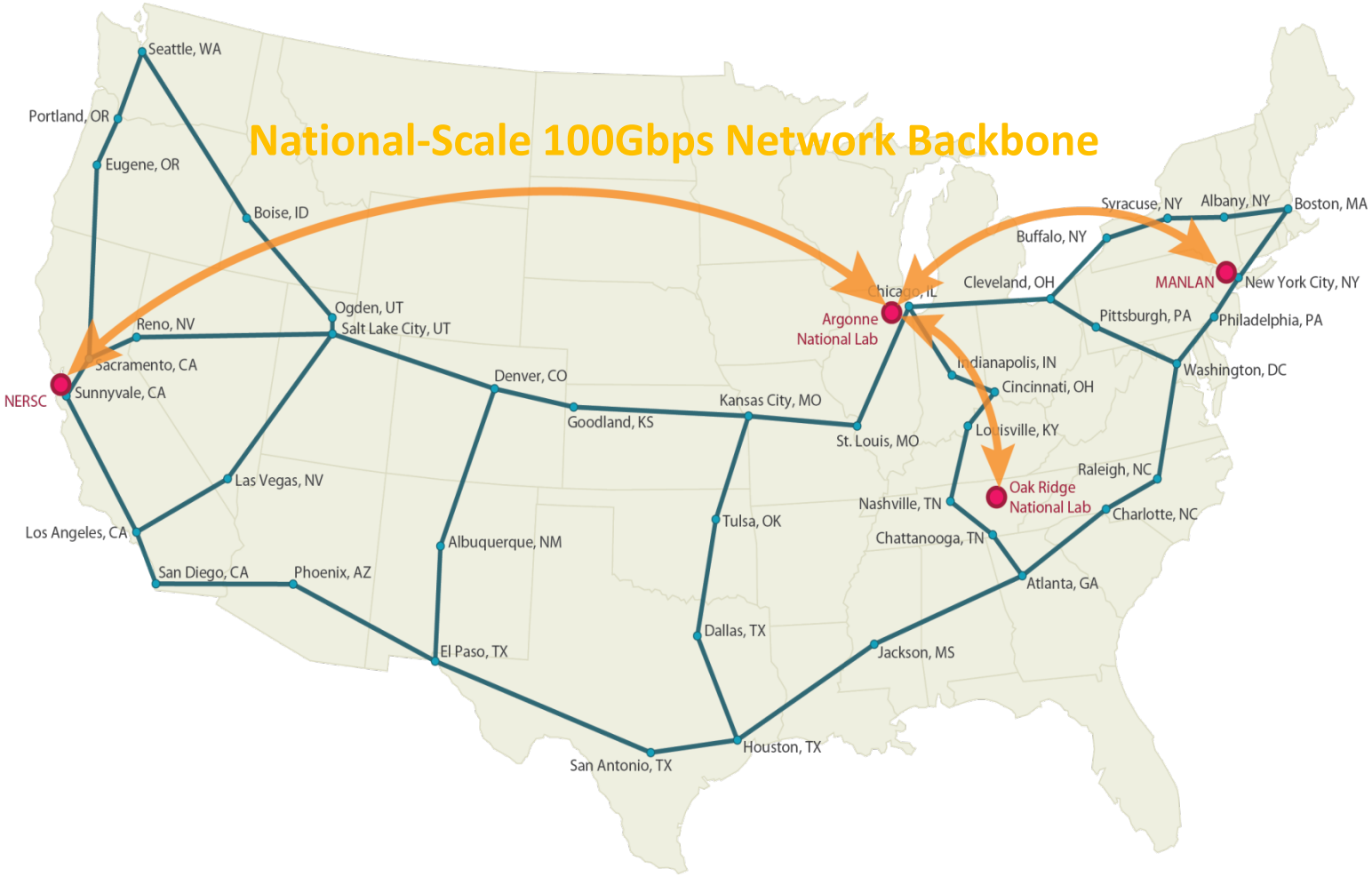
- Major Reasons for the Recommendation: NASA data repository and analysis facilities for SMD missions are distributed across NASA centers and throughout U.S. universities and research facilities.
  - There is considerable variation in the sophistication of the integrated cyberinfrastructure supporting scientific discovery, educational reuse, and public outreach across SMD subdivisions.
  - The rapid rise in the last decade of "mining data archives" by groups other than those funded by specific missions implies a need for a national-scale cyberinfrastructure architecture that can allow for free-flow of data to where it is needed.
  - Other agencies, specifically NSF's Ocean Observatories Initiative Cyberinfrastructure program, should be used as a benchmark for NASA's data-intensive architecture.
- Consequences of No Action on the Recommendation: The science , education, and public outreach potential of NASA's investment in SMD space missions will not be realized .

ITIC Finding

- SMD Data Resides in a Highly Distributed Servers
  - Many Data Storage and Analysis Sites Are Outside NASA Centers
  - Access to Entire Research Community Essential
    - Over Half Science Publications are From Using Data Archives
    - Secondary Storage Needed in Cloud with High Bandwidth and User Portal
  - Education and Public Outreach of Data Rapidly Expanding
    - Images/Videos for Public Relations
    - Apps for Smart Phones
    - Crowd Sourcing



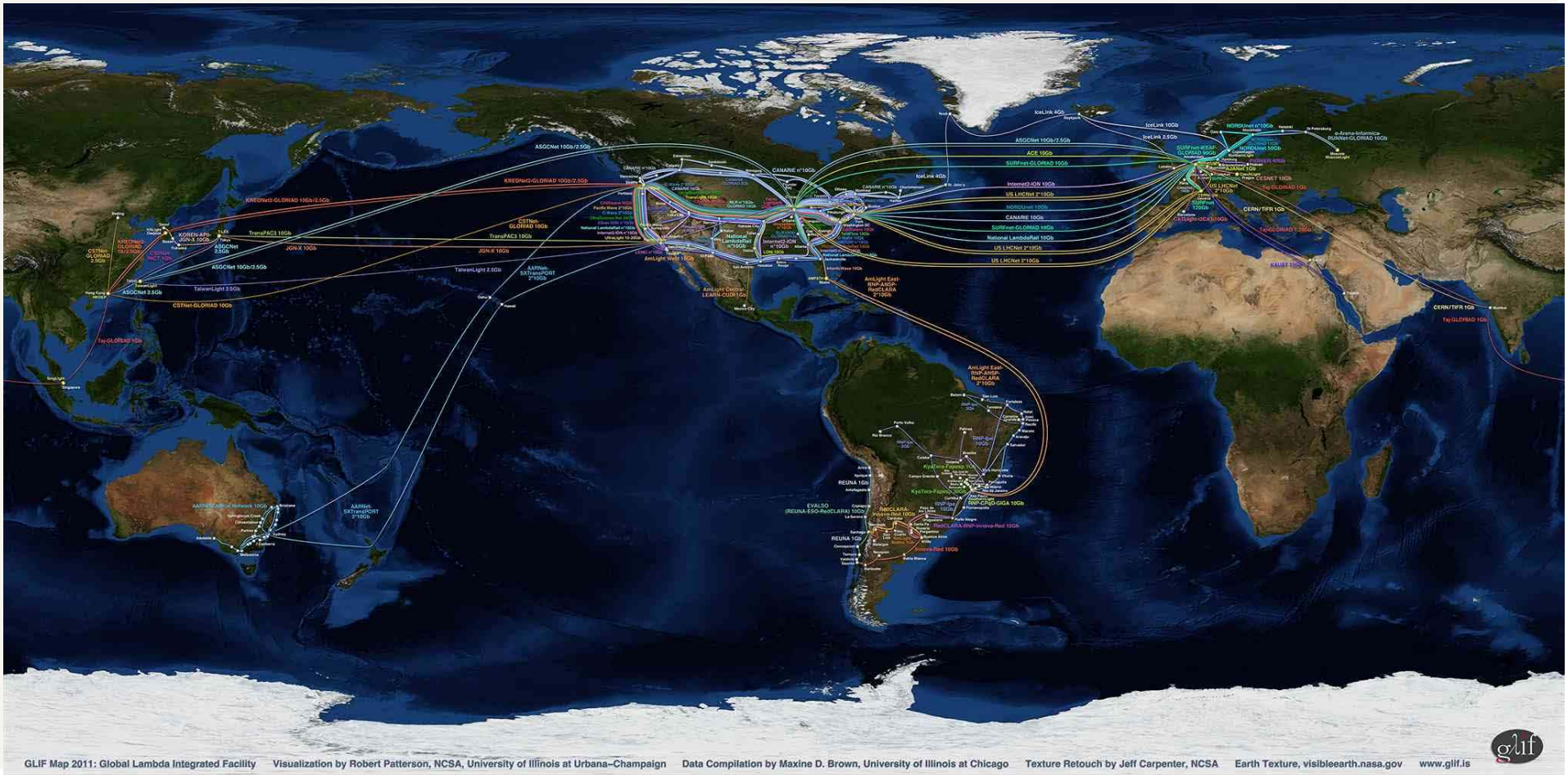
Partnering Opportunities with DOE: ARRA Stimulus Investment for DOE ESnet



Source: Presentation to ESnet Policy Board

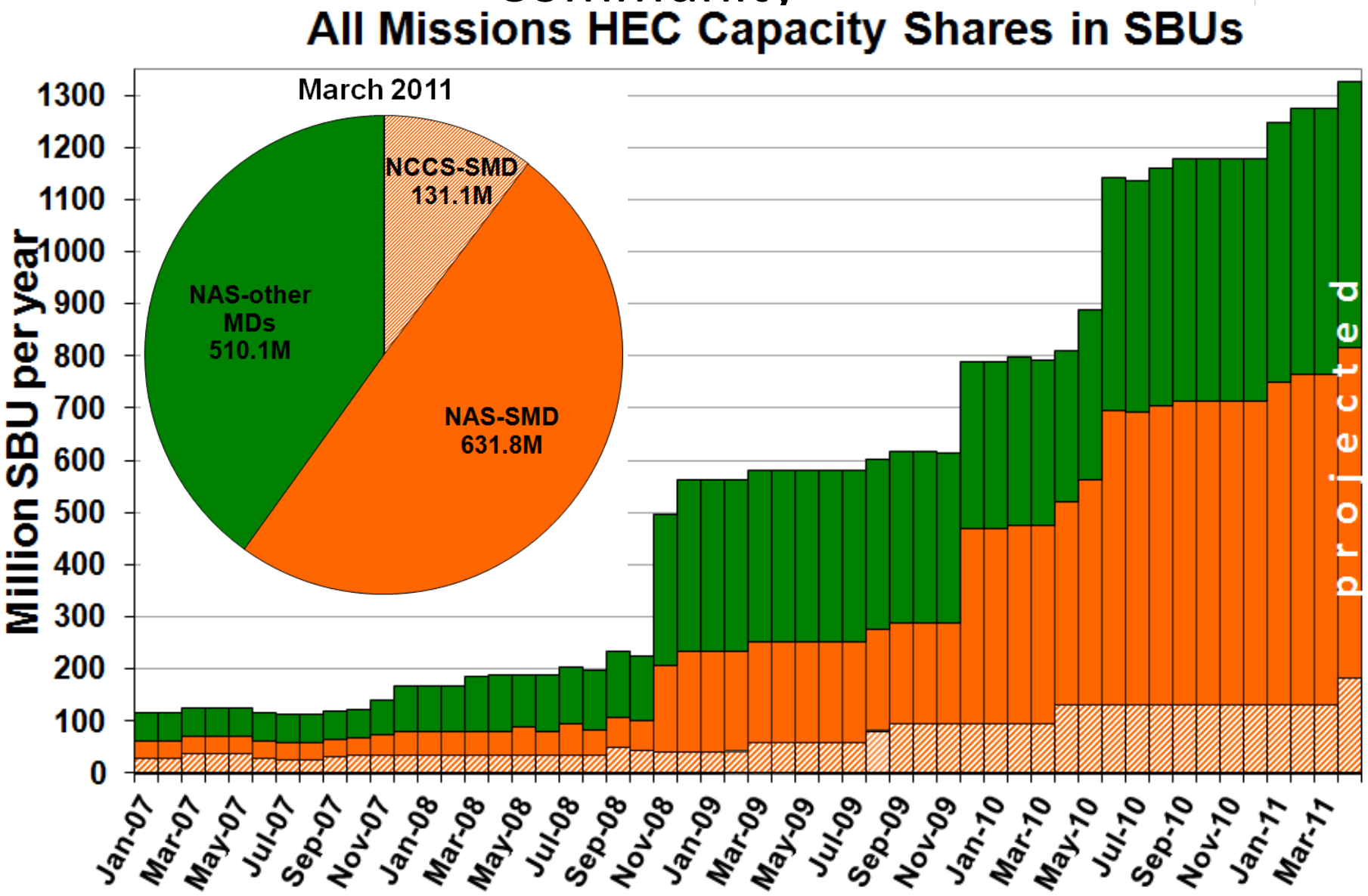
Global Partnering Opportunities:
The Global Lambda Integrated Facility

Research Innovation Labs Linked by 10Gps Dedicated Networks



www.glif.is/publications/maps/GLIF_5-11_World_2k.jpg

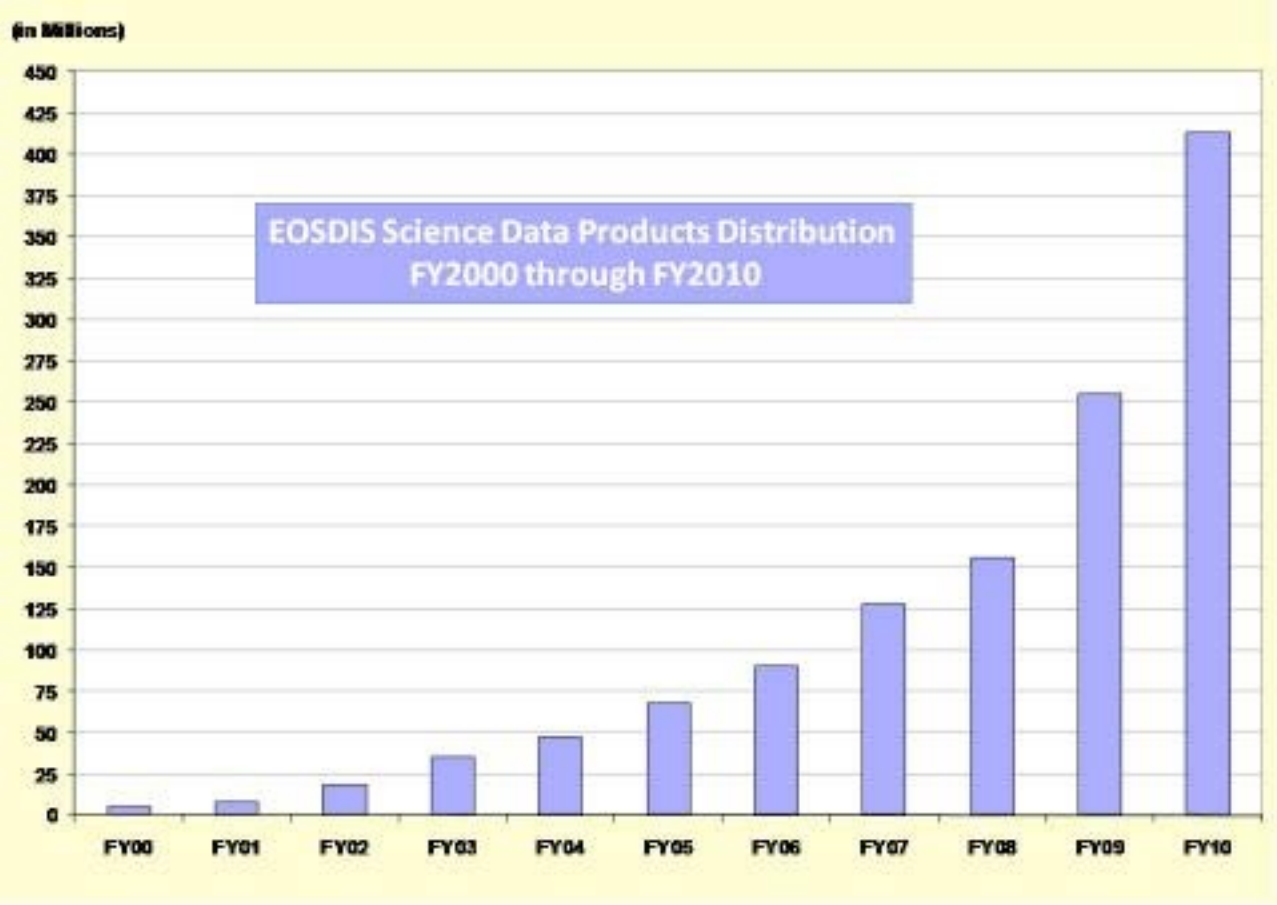
SMD is a Growing NASA HPC User Community



Source: Tsengdar Lee, Mike Little, NASA

EOS-DIS Data Products Distribution

Approaching ½ Billion/Year!



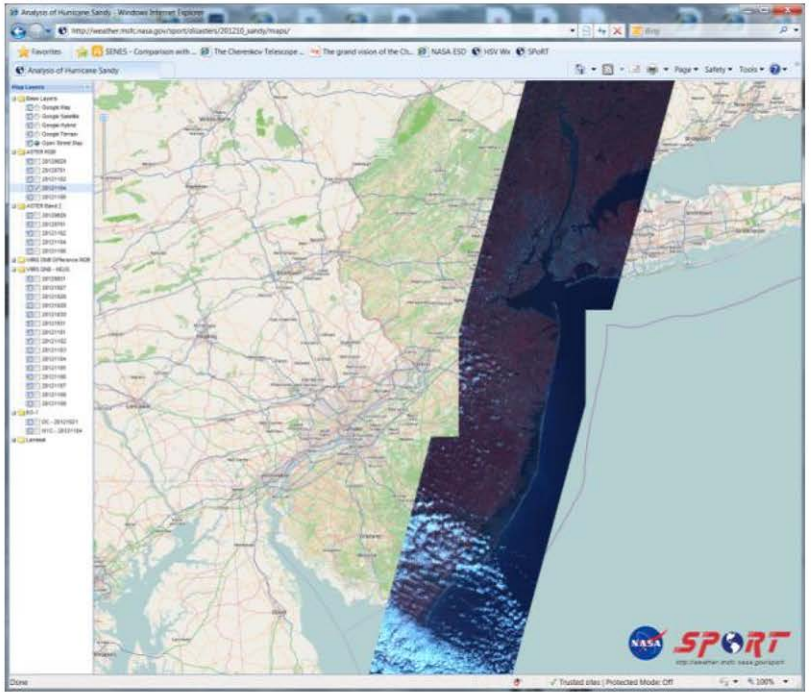
Web Services to Support Disaster Applications

Short-term Prediction Research and Transition Center

Need for access to data and products supporting disaster applications "anytime and from any place"

SPoRT Web Services

- tiled imagery for a "Google Earth" roam and zoom
- web-based applications - tiled web service link
- Android and IPhone "apps"



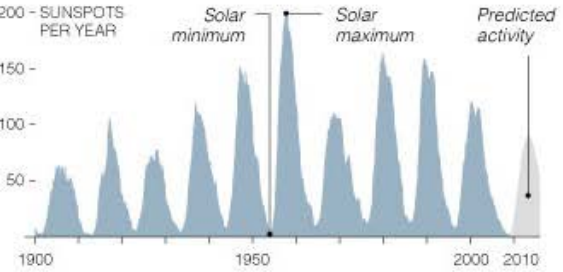
Tiled web service for Hurricane Sandy

SCIENCE ILLUSTRATED | # A Mission to Study the Sun

Scheduled for launching this week, the Solar Dynamics Observatory is the first mission in NASA's Living With a Star program. The craft will observe the Sun in detail, studying how its magnetic field is generated and measuring its extreme ultraviolet radiation. This information will help scientists predict solar activity, which can affect Earth's climate, knock out power grids and disrupt communications.

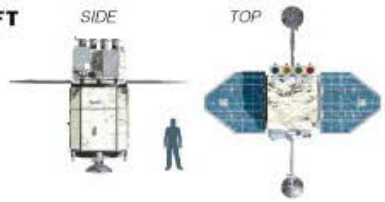
THE SOLAR CYCLE

Solar activity occurs in waves of roughly 11 years. The current lull has been unexpectedly long.



THE SPACECRAFT

S.D.O. will be placed in a geosynchronous orbit. The craft weighs 3,440 pounds and is 14.8 feet tall.



SOLAR PANELS
Produce 1,450 watts of power. Tapered shape avoids blocking antennas.

STAR TRACKER

PROPULSION SYSTEM
S.D.O. carries 3,117 pounds of fuel, enough to keep it in orbit up to 10 years.

THRUSTERS

HIGH-GAIN ANTENNAS
Two antennas transmit 18 megabytes of data per second

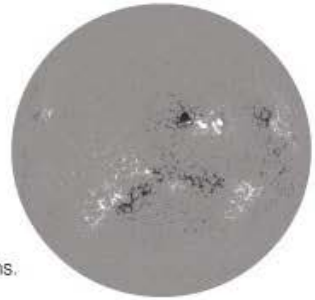
S.D.O.'s SCIENCE INSTRUMENTS

Extreme Ultraviolet Variability Experiment

Will measure variations in the Sun's ultraviolet output, which can fluctuate by factors of 2 to 100 in a matter of minutes.

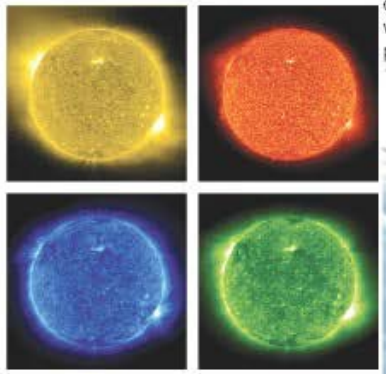
Helioseismic and Magnetic Imager

Will be able to study the flow of material in the Sun's interior by observing waves rippling across the Sun's surface. It will also measure erupting magnetic fields. In the magnetogram at right, black and white areas indicate magnetically intense active regions.



Atmospheric Imaging Assembly

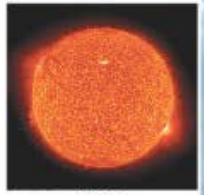
Made up of four multiwavelength telescopes, it will produce high-definition images of the Sun's corona in 8 of 10 selected wavelengths every 10 seconds. Previous spacecraft have been able to study only four wavelengths at different and much slower cadences, as in the images below.



S.D.O. will send back images in higher resolution and at a much higher frame rate than earlier spacecraft, like the Solar and Heliospheric Observatory (SOHO).

NUMBER OF PIXELS

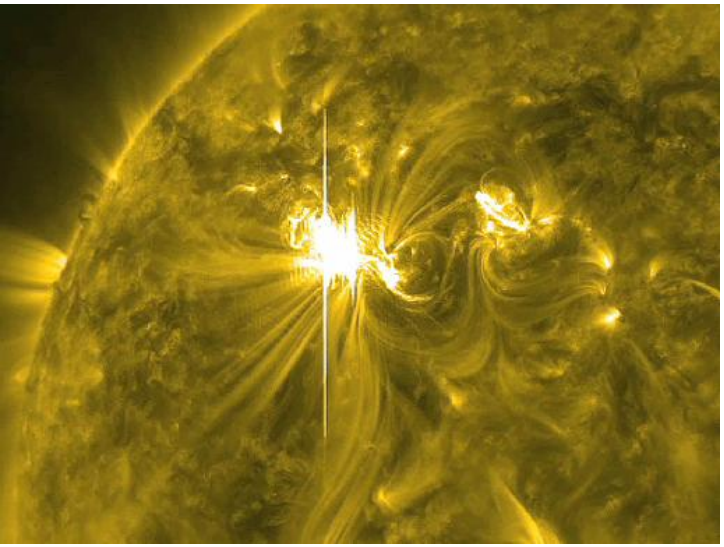
SOHO
1024 x 1024



S.D.O.
4096 x 4096

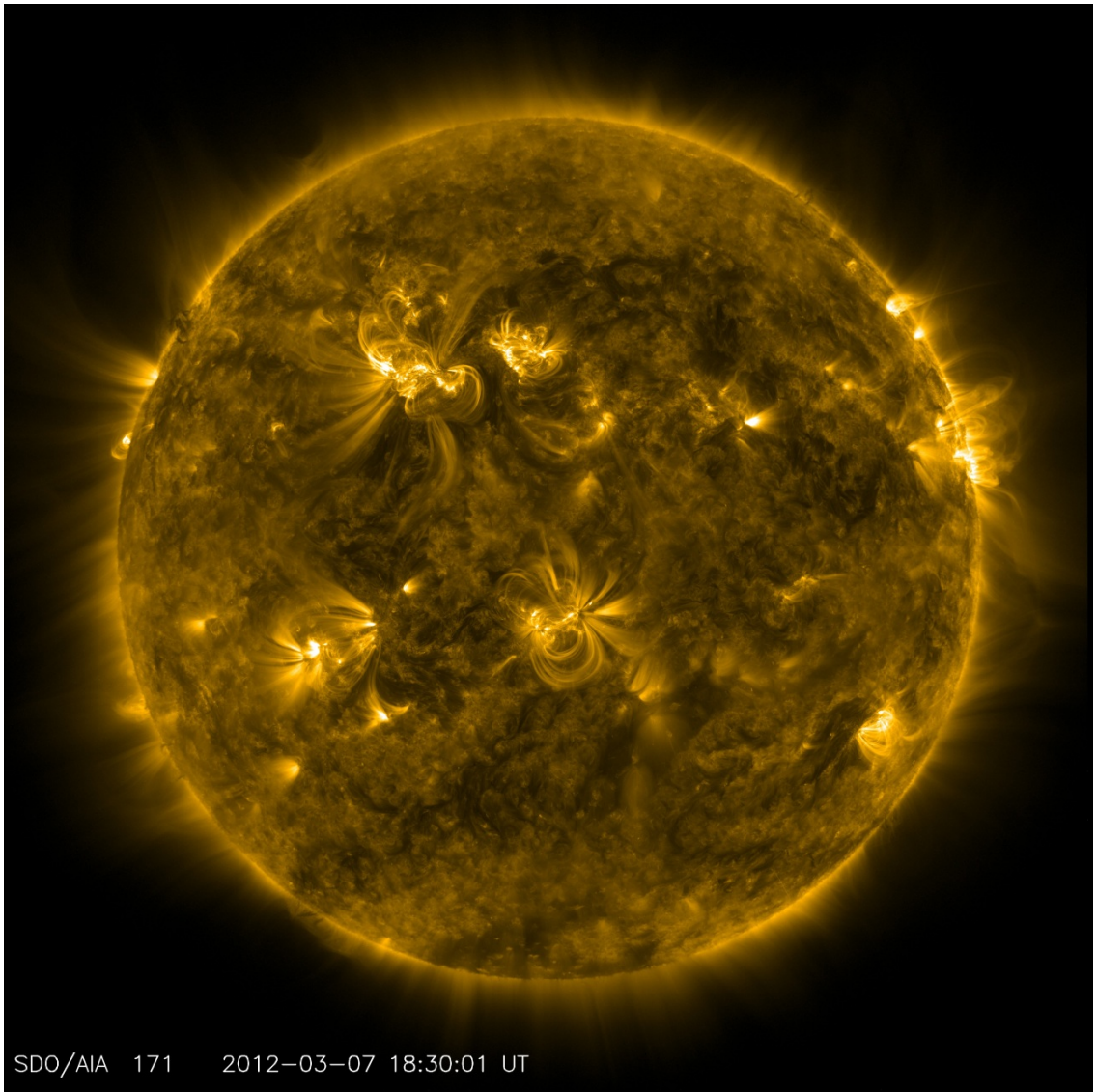
Solar Dynamics Observatory
4096x4096 AIA Camera – 57, 600 Images/Day

JSOC is Archiving ~5TB/day From 6 Cameras
Leads to over 1 Petabyte per year!



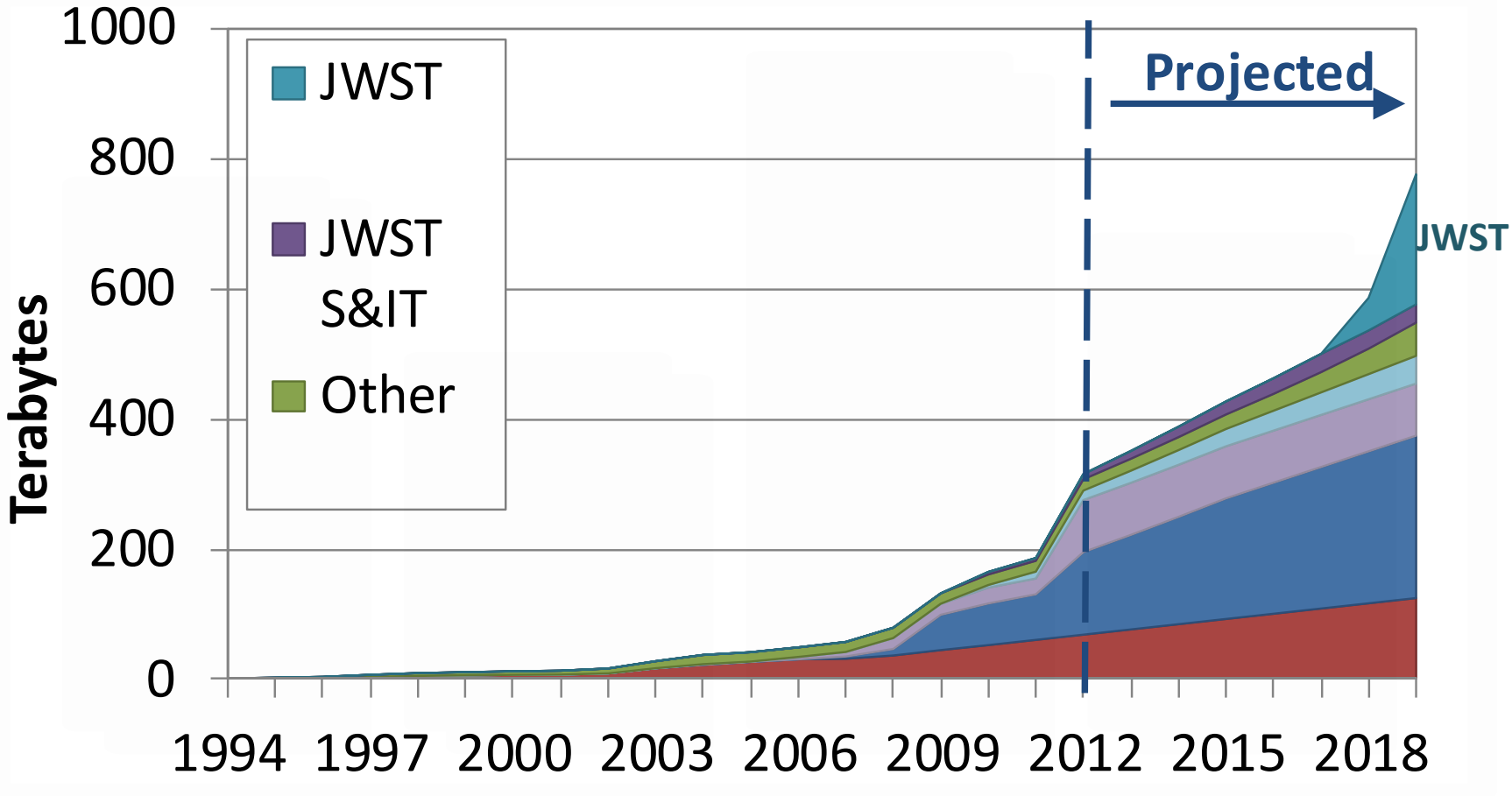
March 6, 2012 X5.4 Flare from Sunspot AR1429 Captured by the Solar Dynamics Observatory (SDO)
in the 171 Angstrom Wavelength

Credit: NASA/SDO/AIA



Multi-Mission Data Archives at STSI Will Continue to Grow - Doubling by 2018

Cumulative Petabyte Over 20 Years



Finding #1 Presented at March, 2012 NAC

- **To enable new scientific discoveries, in a fiscally constrained environment, NASA must develop more productive IT infrastructure through "frugal innovation" and "agile development"**
  - Easy to use as "flickr"
  - Elastic to demand
  - Continuous improvement
  - More capacity for fixed investment
  - Adaptable to changing requirements of multiple missions
  - Built-in security that doesn't hinder deployment

We Found A Two Examples at Goddard Moving in this Direction

32 of the 200+ Apps in the Apple iStore that Return from a Search on “NASA”

Crowdsourcing Science: Galaxy Zoo and Moon Zoo Bring the Public into Scientific Discovery

More than 250,000 people have taken part in Galaxy Zoo so far.
In the 14 months the site was up Galaxy Zoo 2 users helped us make over 60,000,000 classifications. Over the past year, volunteers from the original Galaxy Zoo project created the world's largest database of galaxy shapes.

www.galaxyzoo.org

Finding #1

- The U.S. government has issued several new guidance and directives on open data:
  - OSTP February 22, 2013 Increasing Access to the Results of Federally Funded Scientific Research
  - OSTP March 29, 2013 Big Data is a Big Deal
  - Presidential Exec Order May 9, 2013 Open Data Policy-Managing Information as an Asset

White House Big Data Initiative

- National Science Foundation
- National Institutes of Health
- Department of Defense
- Department of Energy
- U.S. Geological Survey

ITIC Finding #1 Nov. 2012

- While the CSO appears to be doing an amazing job managing the communications requirements and responsibilities for the agency, it may be time for them to develop the business case for acquiring dedicated fiber-optic pathways in support current and future high data-volume traffic: e.g. interfacing to NASA's supercomputers.

Re-organization of the NASA Advisory Council – (Memo signed April 28, 2014)

The NASA Administrator shall establish the following Council committees, subcommittees, and task forces:

- Aeronautics Committee.
- Human Exploration and Operations Committee.
- Science Committee.
  - Astrophysics Subcommittee.
  - Earth Science Subcommittee.
  - Heliophysics Subcommittee.
  - Planetary Protection Subcommittee.
  - Planetary Science Subcommittee.
  - Ad Hoc Task Force on Big Data.
- Technology, Innovation, and Engineering Committee.
- Institutional Committee.
- Ad Hoc Task Force on Science, Technology, Engineering, and Mathematics (STEM) Education.

Timeline

- ITIC in existence – April 2010 – Dec 2013
- NAC reorganized – April 2014
  - Science Committee to have a Big Data Task Force
- BDTF Terms of Reference signed Jan 8, 2015
- BDTF members appointed Dec 2015
- SMD appoints Exec. Sec. who solicits feed back from the Committee members and subcommittees
- 1st meeting of BDTF – Feb 16, 2016

Last slide of the last presentation
By Chair of the ITIC Presentation to the NAC – Dec xx, 2013

NAC Committee on IT Infrastructure Recommendation #1 July 31, 2013

- Recommendation: The NASA NAC ITIC & Science Committees should collaboratively explore the existing and planned evolution of NASA's science data cyberinfrastructure that supports broad access to data repositories for NASA SMD missions. This exploration should be undertaken in the context of effective practices within NASA, other Federal agencies, as well as industry and research institutions.

Wording Agreed to by Both ITIC and Science Committees
July 31, 2013

Work Will Continue as Big Data Taskforce Under Science Committee

Need I say more?