Information Technology Infrastructure Committee (ITIC)

Briefing to the
Astrophysics Science Subcommittee

November 2012

Larry Smarr
Chair ITIC

ITIC Committee Members

Membership

- *Dr. Larry Smarr (Chair), Director- California Institute of Telecommunications and Information Technology, UC San Diego*
- *Dr. Charles Holmes (Vice-Chair), Retired- NASA HQ Heliophysics Program*
- *Mr. Alan Paller, Research Director- SANS Institute*
- *Dr. Robert Grossman, Professor- University of Chicago*
- *Dr. Alexander Szalay, Professor- Johns Hopkins University*
- *----------*

New Members

- *Dr. Mark Boster; President-ImpaQ Solutions, LLC*
- *Hon. Mark Forman, former associate director of IT and e-government, OMB*
- *Mr. Joel Mambretti, Director, Intl. Center for Advanced Internet Research, NW Univ.*
- *Dr. Ed Lazowska, Gates Professor & Chair , Dept of Computer Science, UWash*
- *Dr. Pete Beckman, Dir., Exascale Technology and Computing Institute, Argonne NL*
- *Mr. John Muratore, former NASA engineer & Program Manager, now with Space X*

- *Mr. Jason Gillis (Exec Sec), Special Assist. to CIO, NASA*

NAC Committee on IT Infrastructure
DRAFT* Recommendation #2 [March 2012]

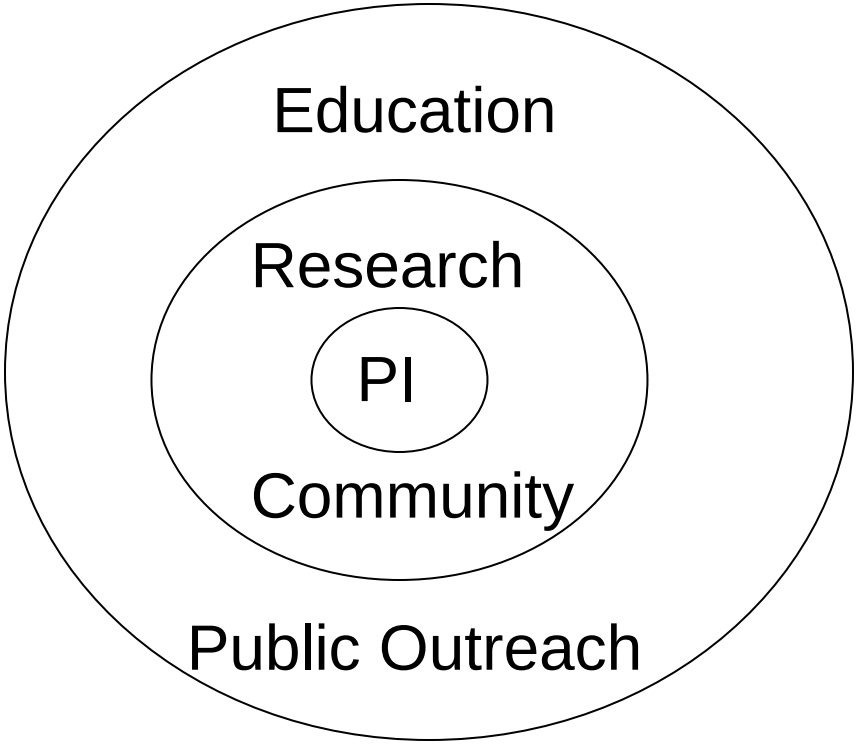
- **Recommendation:** NASA should formally review the existing national data cyberinfrastructure supporting access to data repositories for NASA SMD missions. A comparison with best-of-breed practices within NASA and at other Federal agencies should be made.
- We request a briefing on this review to 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

Good Progress, Stay Tuned for Next NAC Meeting

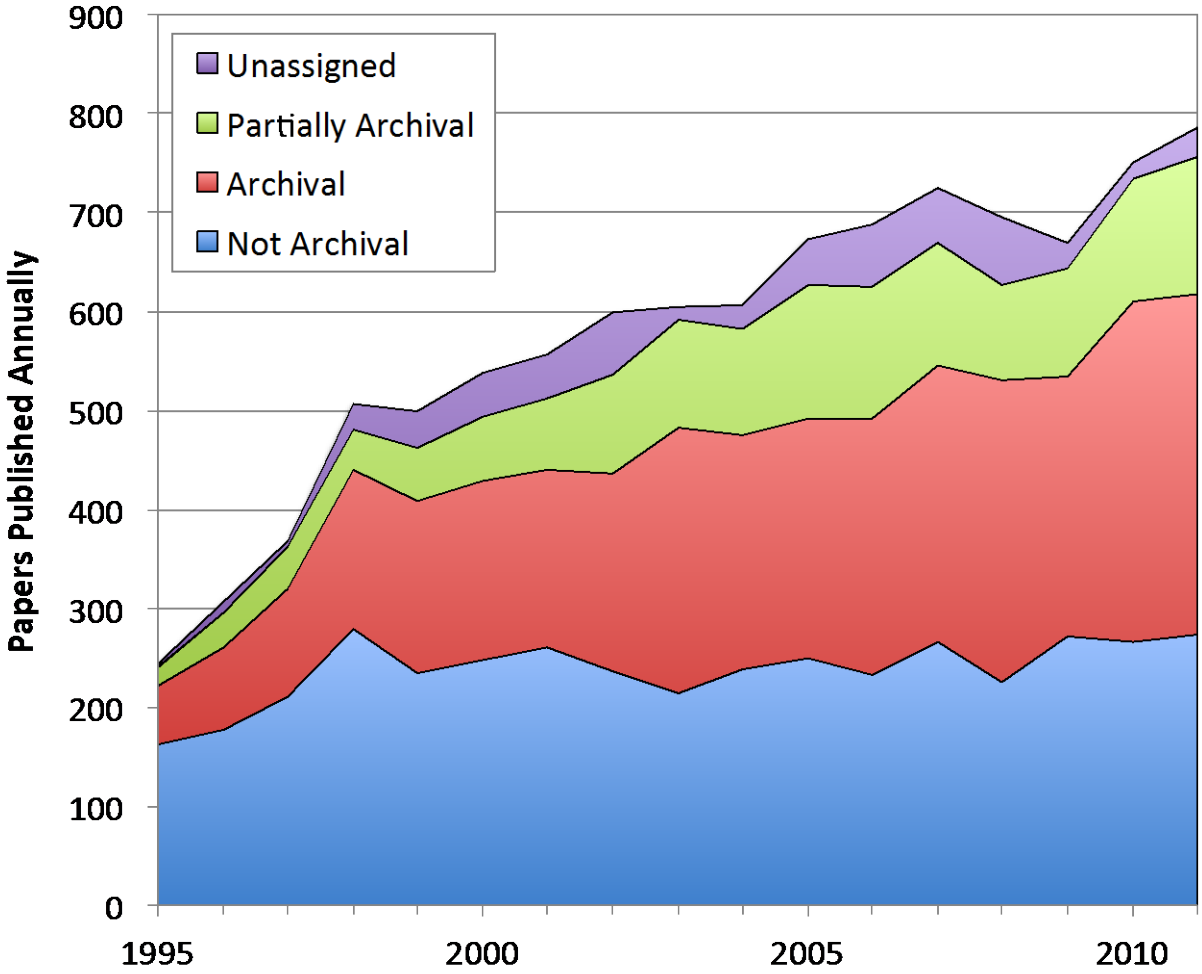
ITIC Finding

- **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 for Public Relations
    - Apps for Smart Phones
    - Crowdsouring

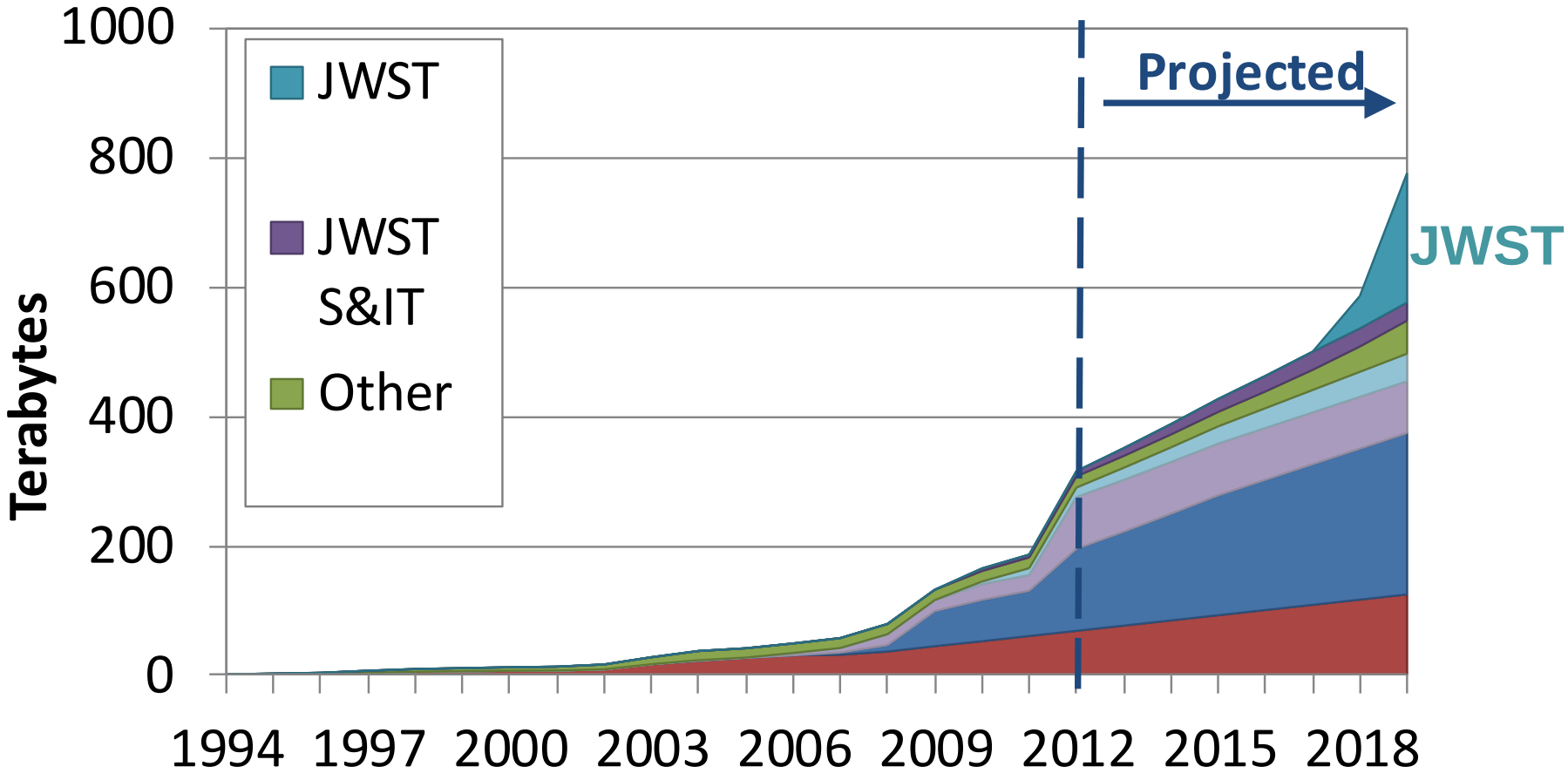


Majority of Hubble Space Telescope Scientific Publications Come From Data Archives

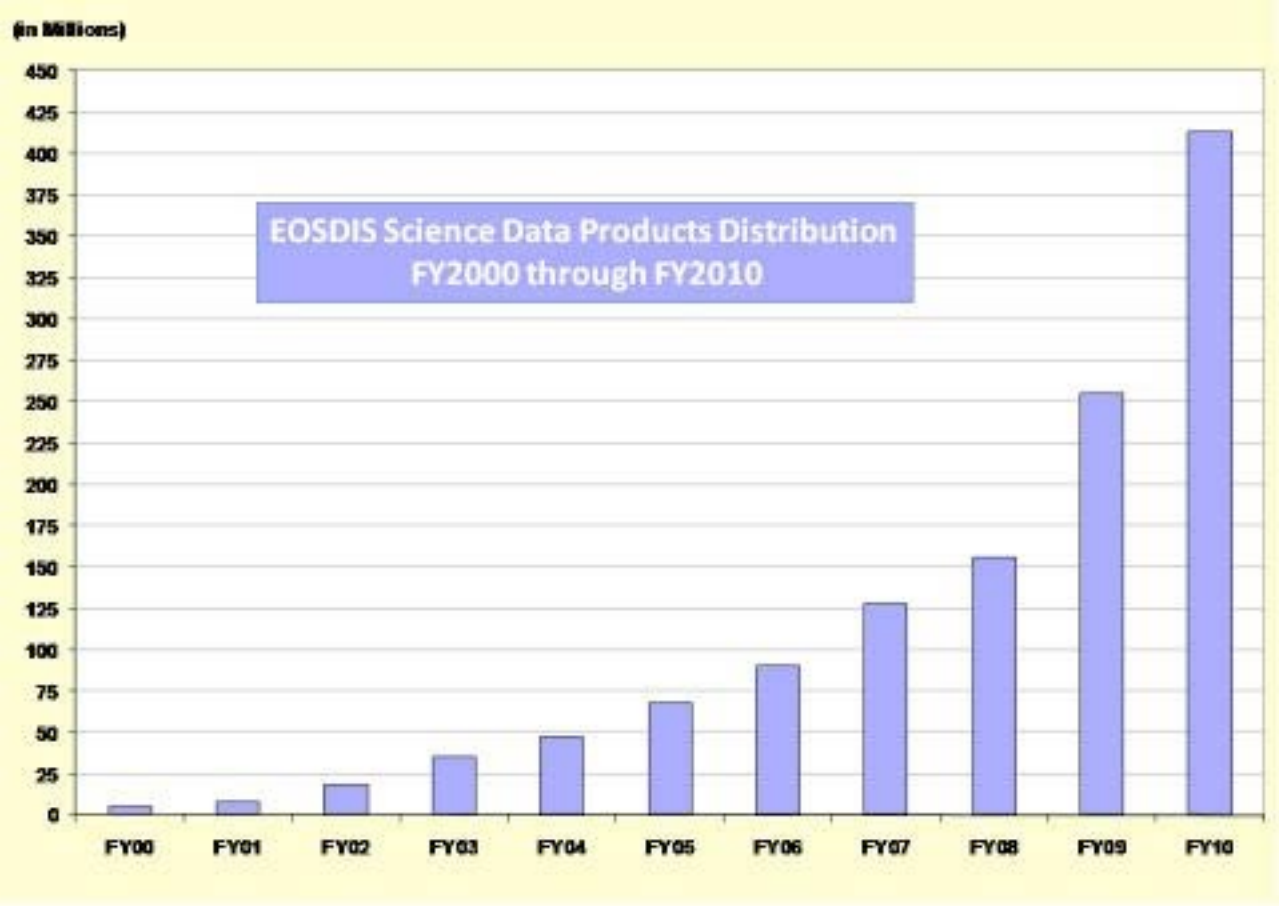
In 2011 there were over 1060 papers written using data archived at MAST



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



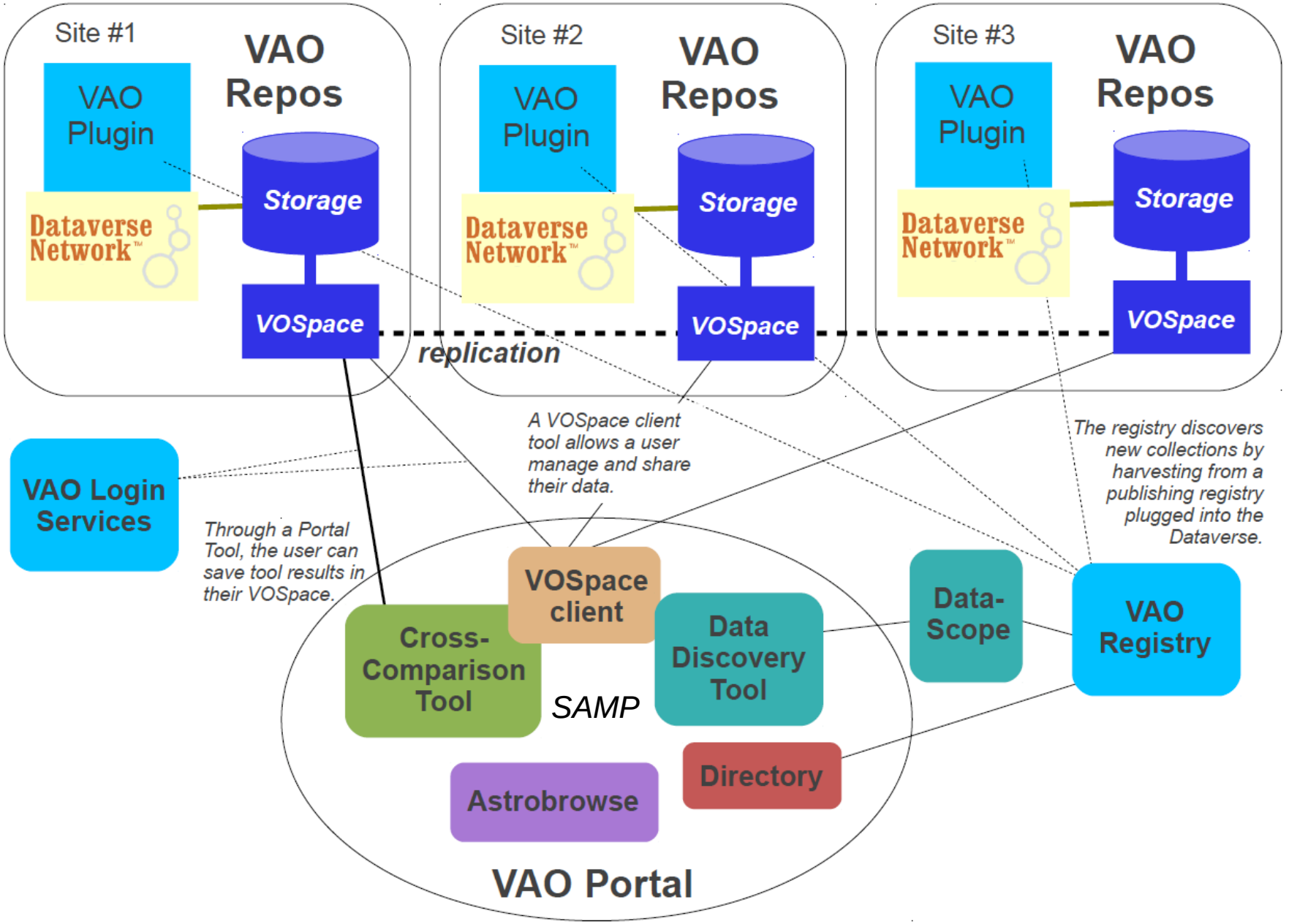
EOS-DIS Data Products Distribution Approaching ½ Billion/Year!



The Virtual Observatory

- **The VO is foremost a data discovery, access, and integration facility**
- **International collaboration on metadata standards, data models, and protocols**
  - Image, spectrum, time series data
  - Catalogs, databases
  - Transient event notices
  - Software and services
  - Distributed computing (authentication, process management)
  - Application inter-communication

- **International Virtual Observatory Alliance established in 2001, patterned on WorldWideWeb Consortium (W3C)**


Site #1
VAO Repos
VAO Plugin
Dataverse Network™
Storage
VOSpace
Site #2
VAO Repos
VAO Plugin
Dataverse Network™
Storage
VOSpace
Site #3
VAO Repos
VAO Plugin
Dataverse Network™
Storage
VOSpace
replication
A VOSpace client tool allows a user manage and share their data.
The registry discovers new collections by harvesting from a publishing registry plugged into the Dataverse.
VAO Login Services
Through a Portal Tool, the user can save tool results in their VOSpace.
VOSpace client
Cross-Comparison Tool
SAMP
Data Discovery Tool
Data-Scope
VAO Registry
Directory
Astrobrowse
VAO Portal

NASA Space Images Are Widely Viewed by Public

Dark Matter Blob Should Not Exist, But There It Is

New Hubble observations puzzle astronomers.

NATIONAL GEOGRAPHIC Daily News

Dark matter, galaxies, and hot gas merge in the core of the galaxy cluster Abell 520 in a composite image.
Image courtesy M.J. Jee/U.C. Davis, A. Mahdavi/SFSU, and NASA/ESA/CFHT/CXO

http://news.nationalgeographic.com/news/2012/03/120306-dark-matter-galaxies-mystery-space-science/

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.

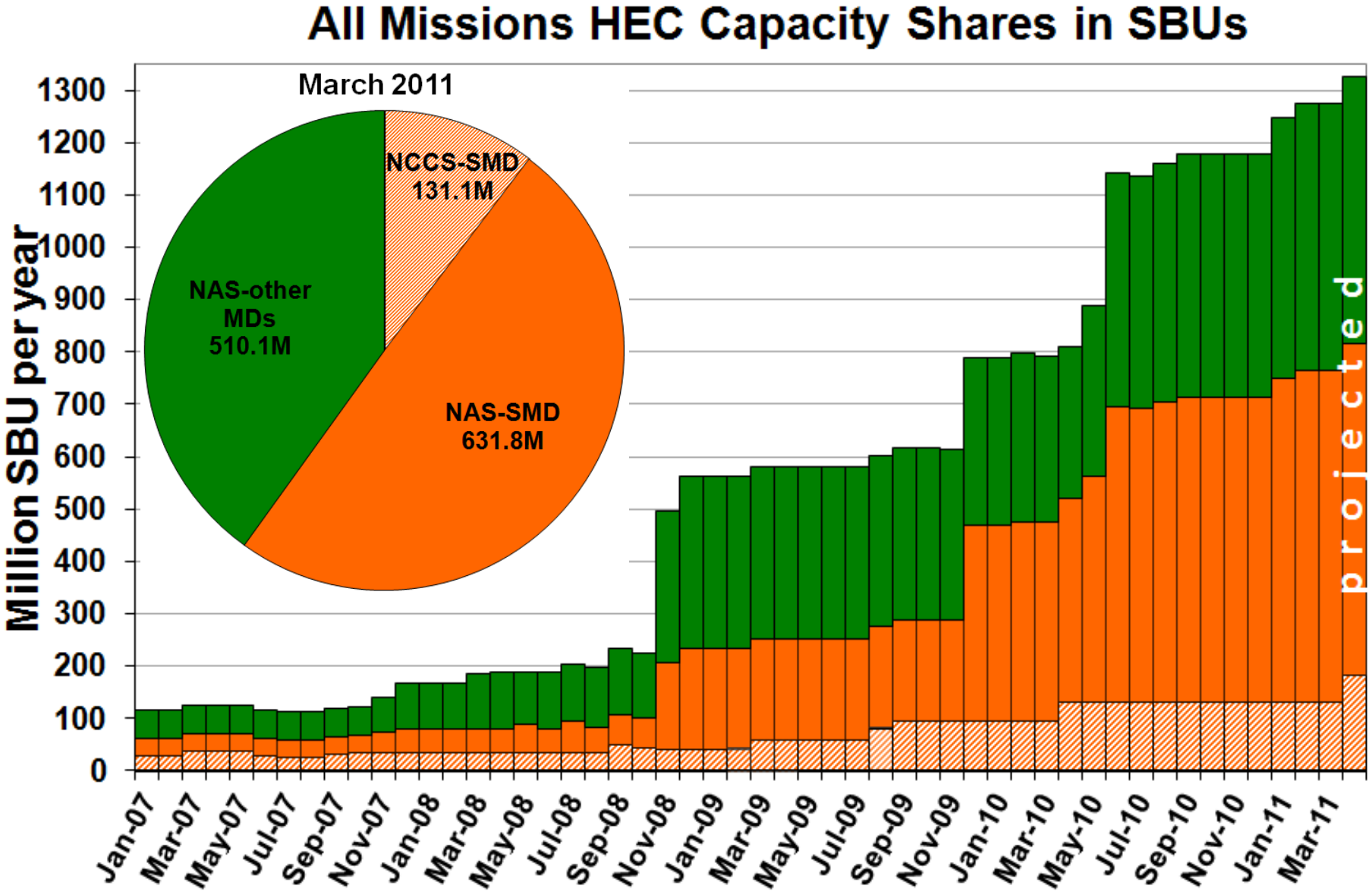
www.galaxyzoo.org

ITIC Finding: NASA is Falling Behind Federal and Non-Federal Institutions

	Big Data CI	10G→ 100G	GPU Clusters	Hybrid HPC
Non-Fed	Google/MS/ Amazon	GLIF/I2/ CENIC	Japan TSUBAME2 4224 GPUs 2.4 PF	China #2 Fastest 5 PF MC/GPU
NSF	Gordon	GENI Next Gen Internet	TAAC 512 GPUs	Blue Waters* MC/GPU 12 PF
DOE	Magellan	ANI ARRA 100Gb	ANL 256 GPUs	NG Jaguar* MC/GPU 20 PF
NASA	Nebula, Testbed	Goddard to Ames 10G	Ames 136 GPU 2 x 64 at Ames & GFSC	Pleiades MC 1PF

* Later in 2012

SMD is a Growing NASA HPC User Community



Source: Tsengdar Lee, Mike Little, NASA

Leading Edge is Moving to Hybrid Processors: Requiring Major Software Innovations

China takes HPC heavyweight title

GPUs, Arch interconnect knocks out Jaguar and Roadrunner

By **Timothy Prickett Morgan** • **Get more from this author**

Posted in HPC, 28th October 2010 14:07 GMT

Jaguar is Getting a GPU Upgrade, to Make it the World's Fastest Supercomputer Again

The new souped-up supercomputer will be renamed Titan

By Clay Dillow Posted 10.12.2011 at 2:02 pm 18 Comments

NVIDIA Tesla GPUs to Accelerate NCSA Blue Waters Supercomputer

Petascale Supercomputer to Be One of the World's Most Powerful Scientific Tools

SEATTLE, WA -- (Marketwire) -- 11/14/2011 -- **SC11 --**

ENERGY.GOV

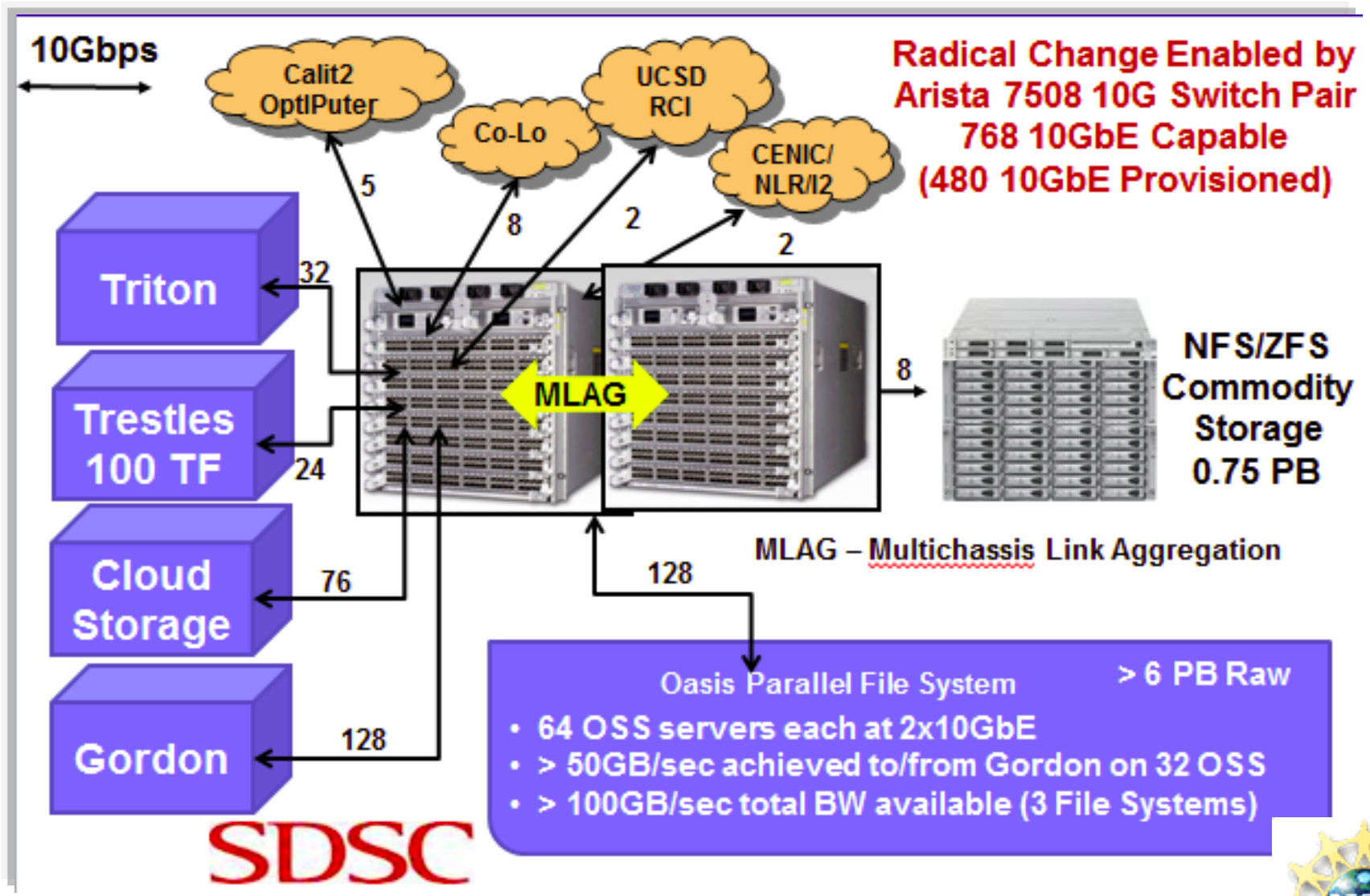
TECHNOLOGY - Written by OLCF Staff Writer on February 6, 2012

Computer Scientists Collect Computing Tools for Next-Generation Machines

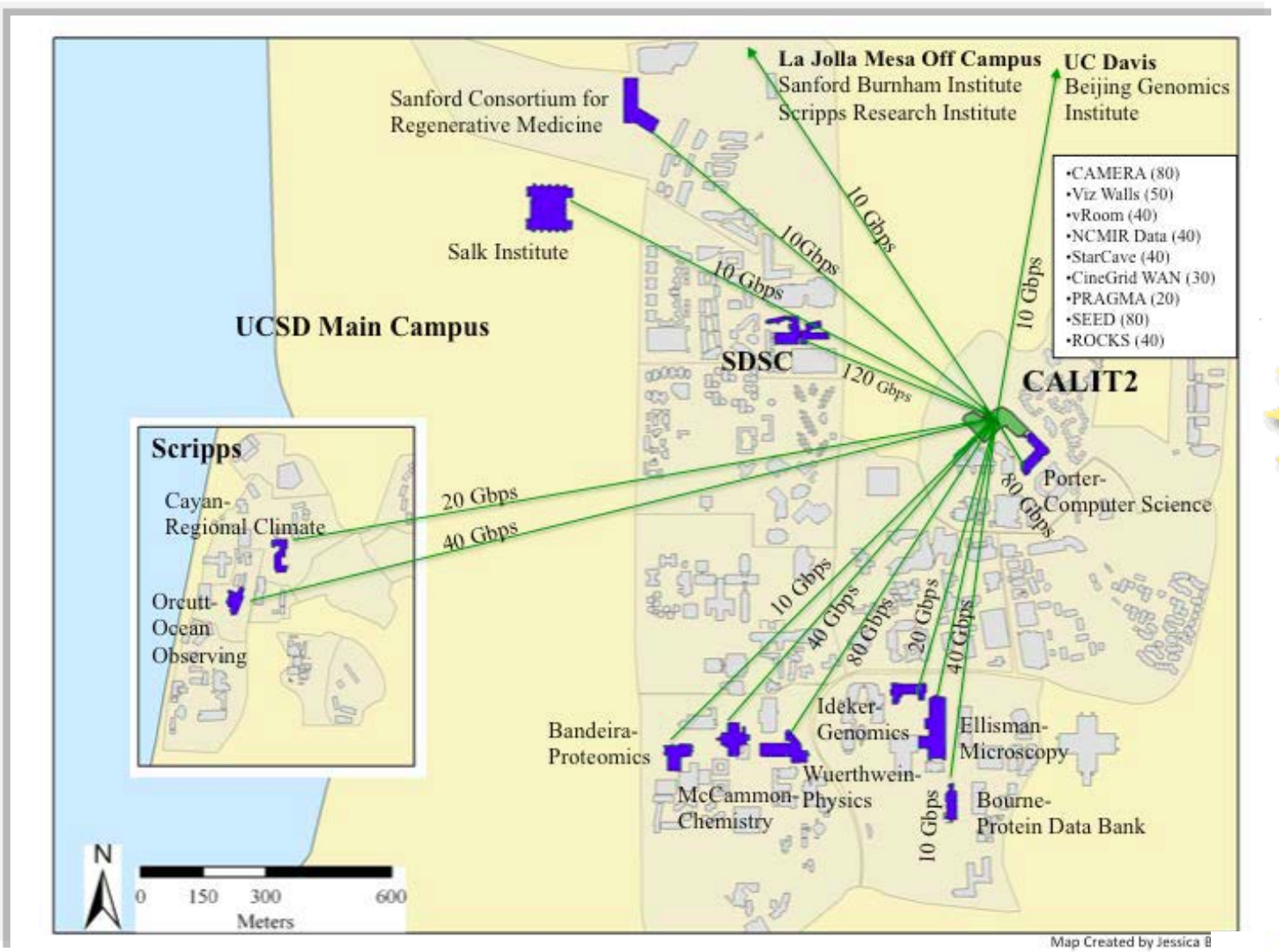
Tags: GPUs, Jaguar, Titan, Tools

"With Titan's arrival, fundamental changes to computer architectures will challenge researchers from every scientific discipline."

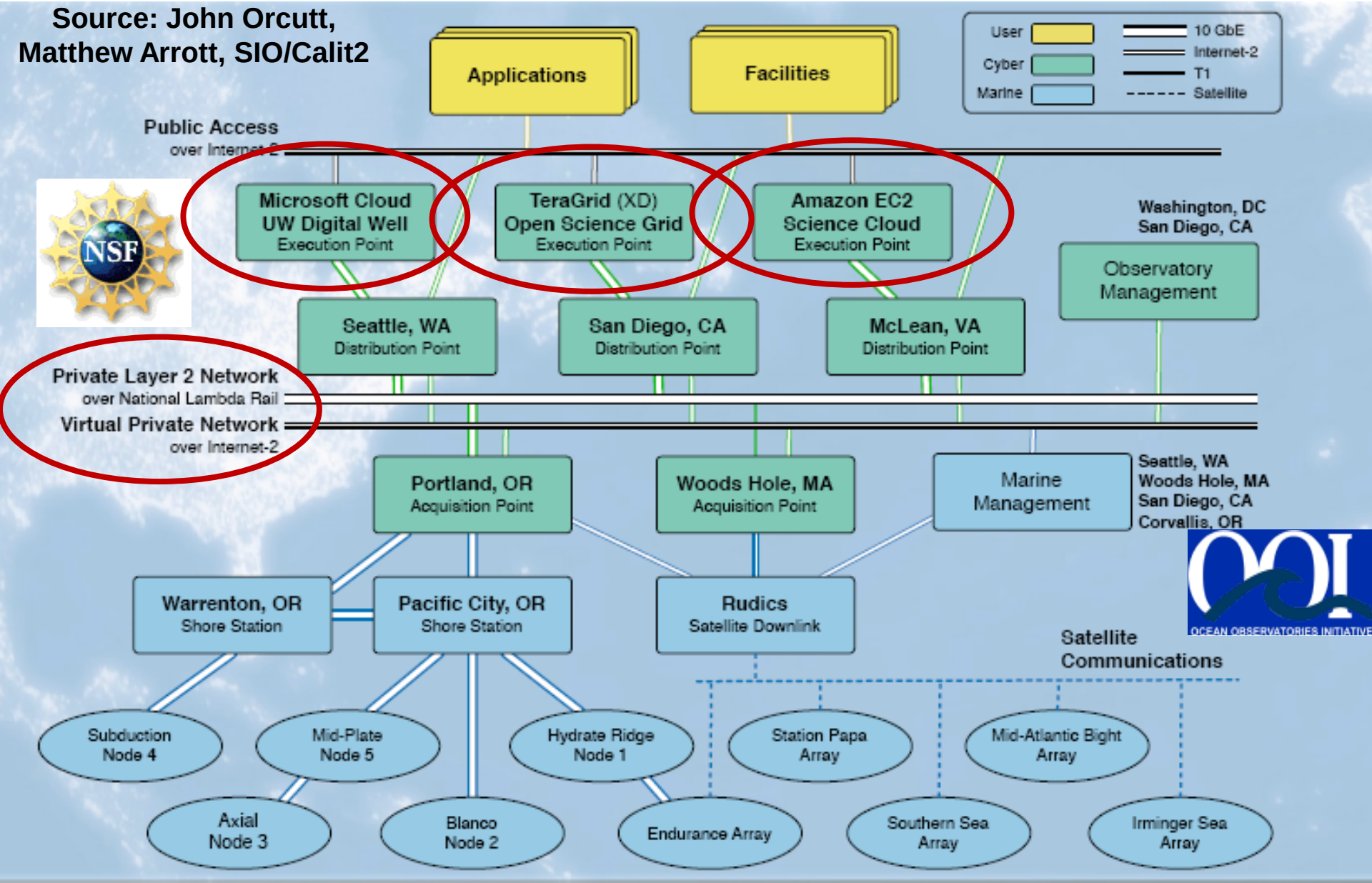
Arista Enables SDSC's Massive Parallel 10G Switched Data Analysis Resource



Prism@UCSD Optical Connections



OOI CI is Built on Dedicated Optical Networks and Federal Agency & Commercial Clouds



The Next Step for Data-Intensive Science: Pioneering the HPC Cloud

HPC Cloud Computing

Self Service and Dynamically Scalable

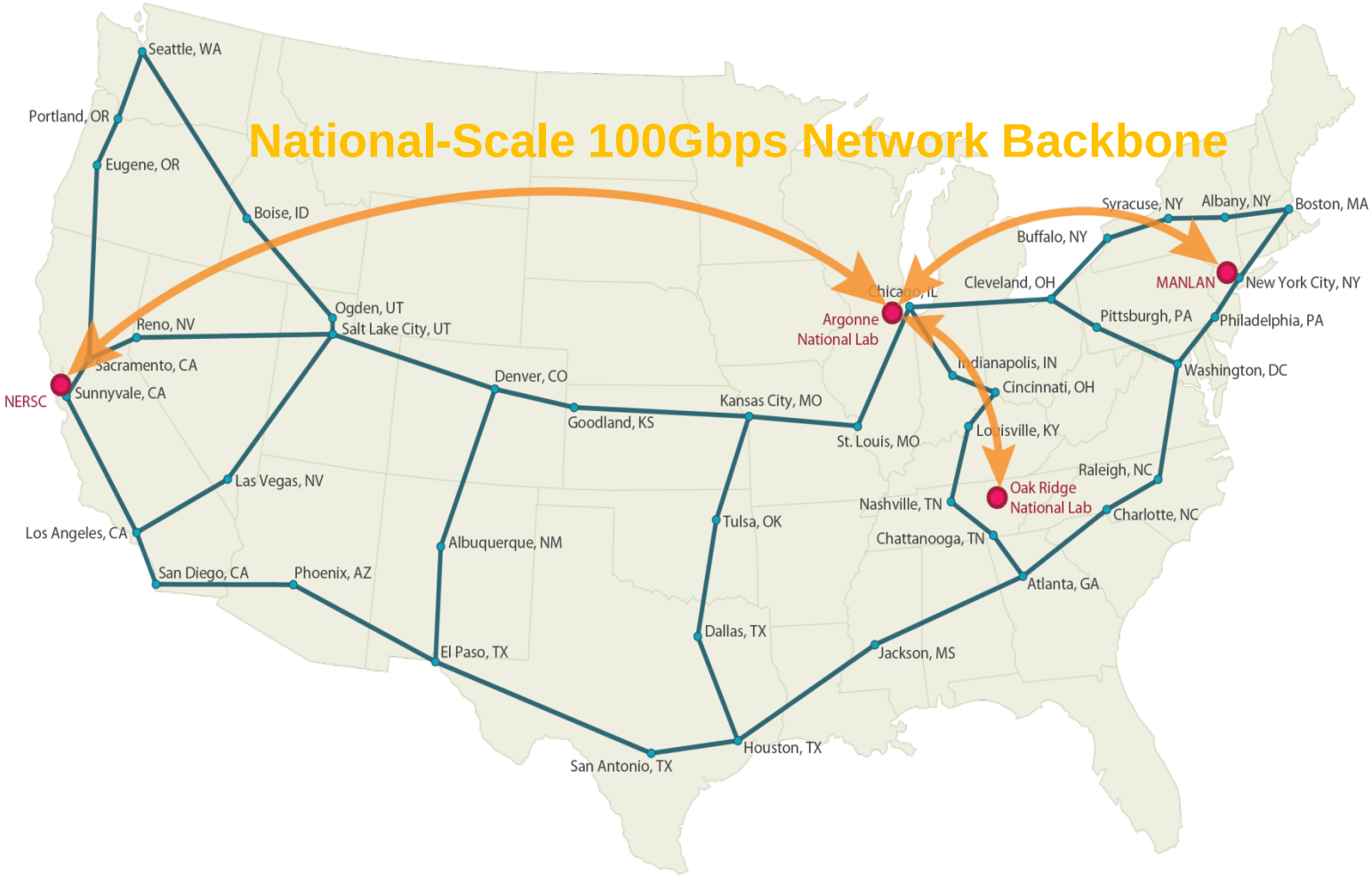
High Performance Computing (HPC) on

Partnering Opportunities with Universities
John Hopkins University DataScope

- **Private Science Cloud for Sustained Analysis of PB Data Sets**
  - Built for Under $1M
  - 6.5PB of Storage, 500 Gbytes/sec Sequential BW
  - Disk IO + SSDs Streaming Data into an Array of GPUs
  - Connected to Starlight at 100G (May 2012)
- **Some Form of a Scalable Cloud Solution Inevitable**
  - Who will Operate it, What Business Model, What Scale?
  - How does the On/Off Ramp Work?
- **Science has Different Tradeoffs than eCommerce:**
  - Astronomy,
  - Space Science,
  - Turbulence,
  - Earth Science,
  - Genomics,
  - Large HPC Simulations Analysis

Source: Alex Szalay, JHU

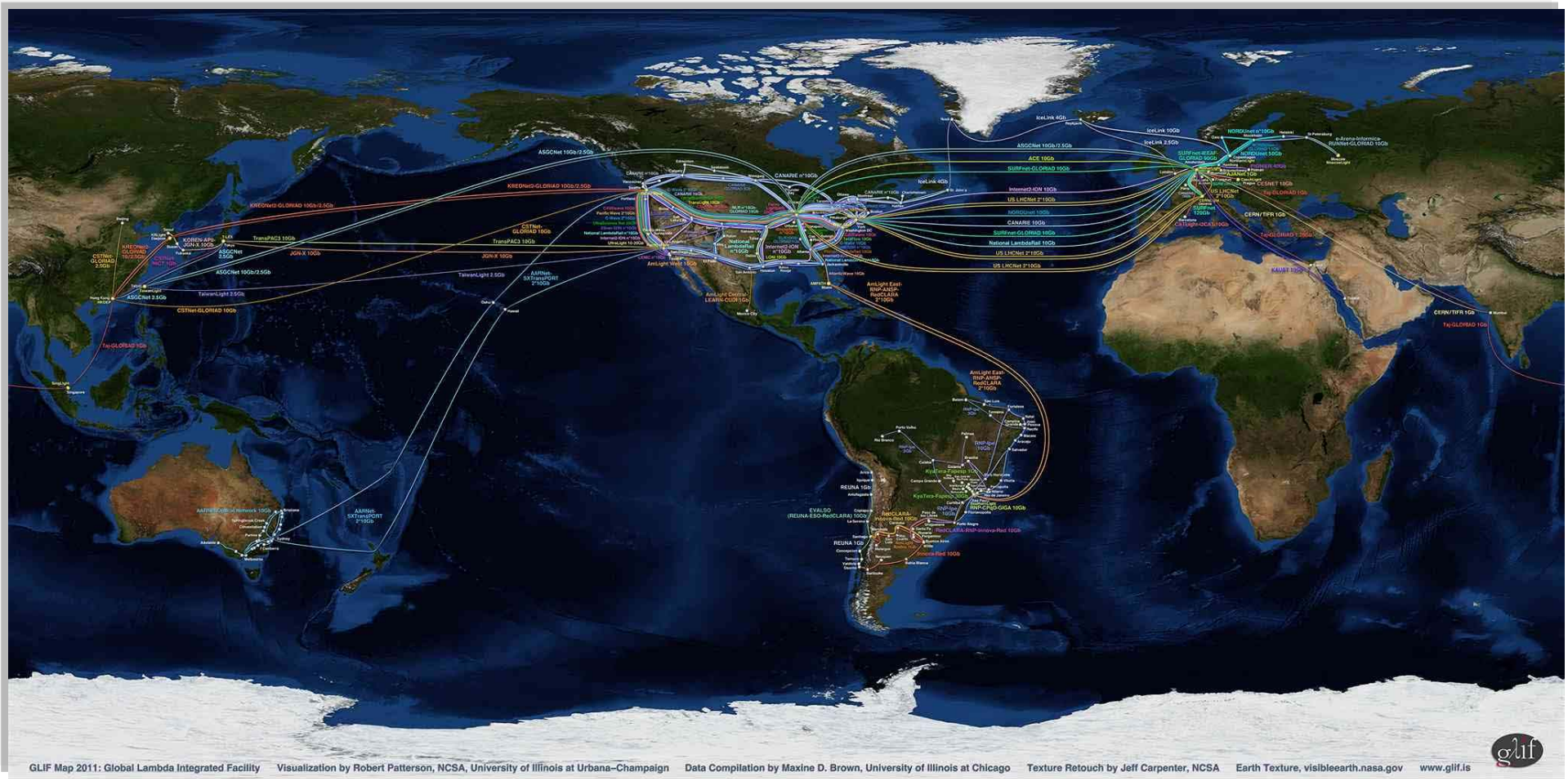
Partnering Opportunities with DOE: ARRA Stimulus Investment for DOE ESnet



Source: Presentation to ESnet Policy Board

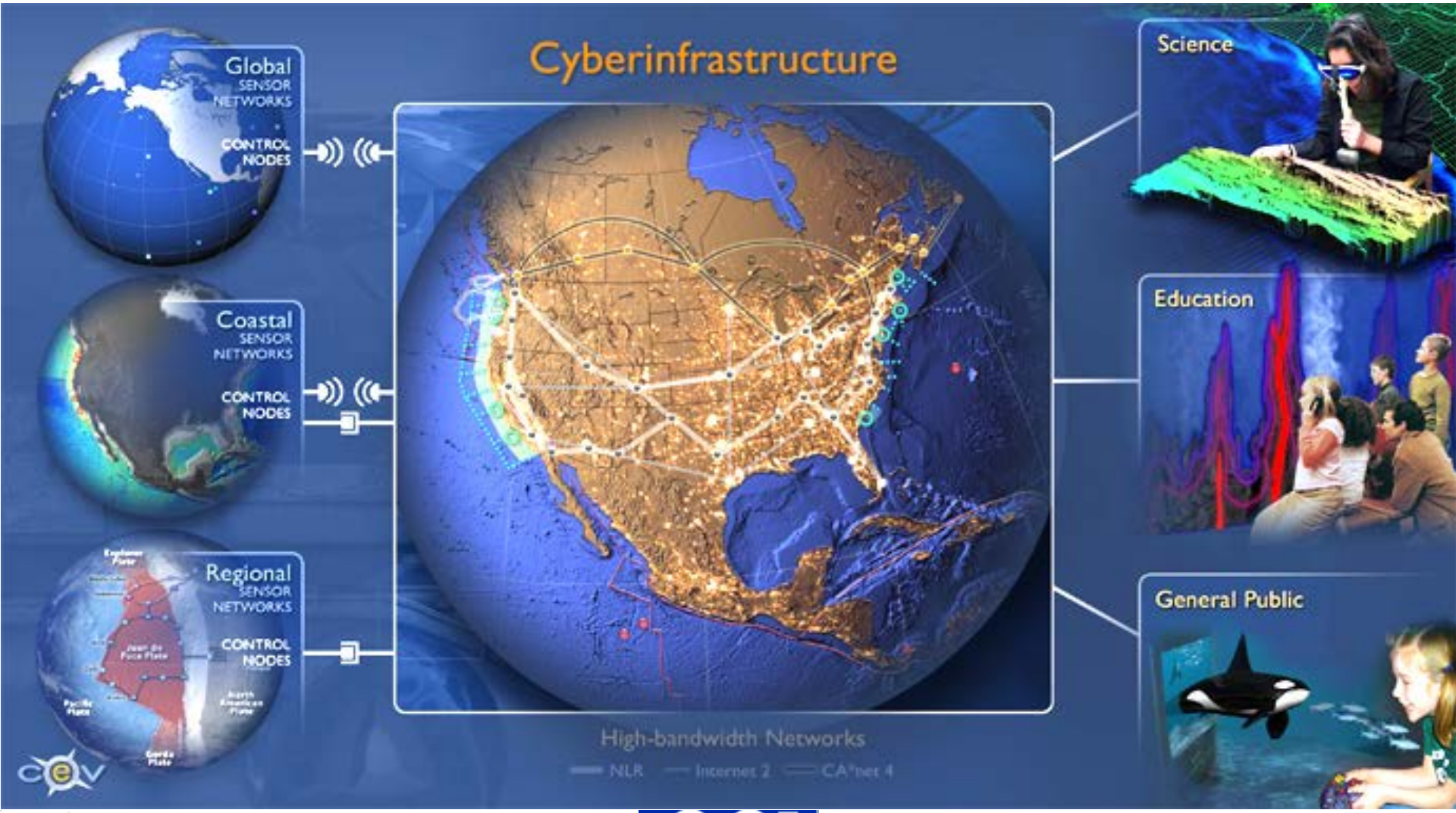
Global Partnering Opportunities:
The Global Lambda Integrated Facility

Research Innovation Labs Linked by 10Gps Dedicated Networks



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

NSF's Ocean Observatory Initiative Cyberinfrastructure Supports Science, Education, and Public Outreach



Source: Matthew Arrott, Calit2 Program Manager for OOI CI

Committee Concern: NASA and White House Big Data Initiative

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

Discussion