Oklahoma Supercomputing Symposium -2014

Supporting Collaborative Cyberinfrastructure for Ground-breaking Research

Irene M. Qualters
Division Director, Advanced Cyberinfrastructure
National Science foundation
September 24, 2014

NSF Core Mission: Fundamental Research

$7.1 billion FY 2014 research budget

94% funds research, education and related activities

50,000 proposals

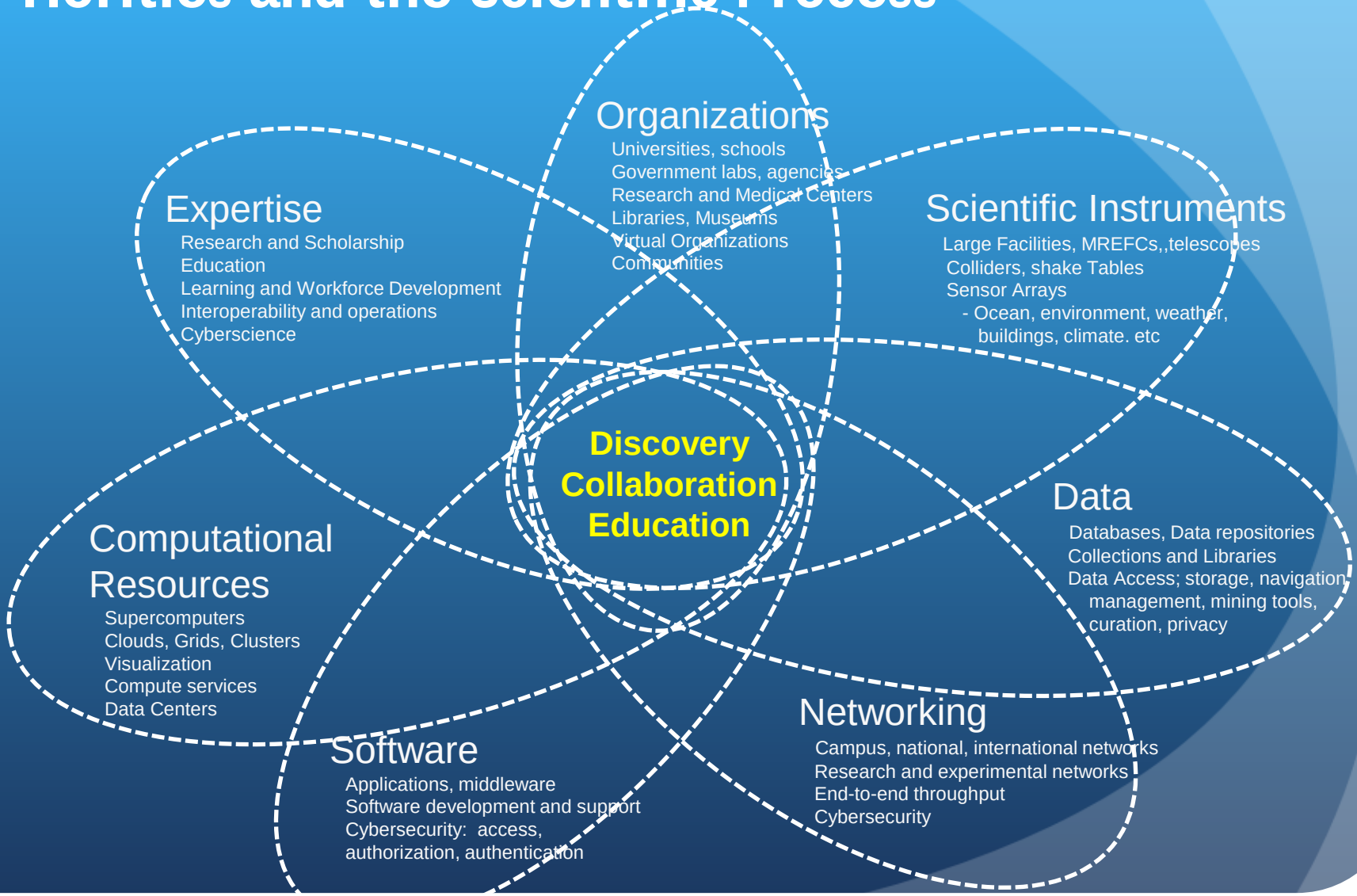
11,000 awards funded

2,000 NSF-funded institutions

300,000 NSF-supported researchers

Fundamental Research

NSF Embraces an Expansive View of Cyberinfrastructure Motivated by Research Priorities and the Scientific Process



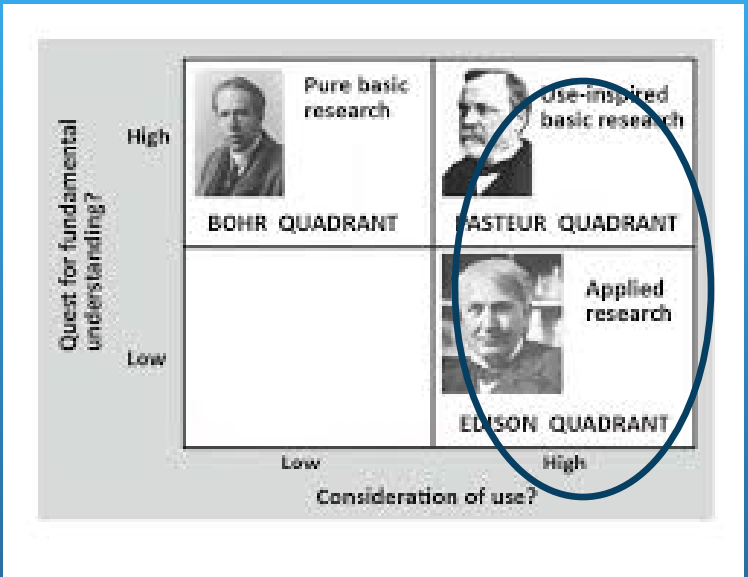
NSF Cyberinfrastructure (ACI) is part of the CISE Directorate and responsible for NSF-wide CI coordination and support

Computer and Information Science and Engineering (CISE)

Division of Advanced Cyberinfrastructure (ACI)	Division of Computing and Communications Foundations (CCF)	Division of Computer and Network Systems (CNS)	Division of Information and Intelligent Systems (IIS)
Data	Algorithmic Foundations	Computer Systems Research	Cyber-Human Systems
High Performance Computing	Communication and Information Foundations	Networking Technology and Systems	Information Integration and Informatics
Networking/ Cybersecurity	Software and Hardware Foundations		Robust Intelligence
Software			

Core Research Programs

ACI Mission: To support advanced cyberinfrastructure to accelerate discovery and innovation across all disciplines



- Coordination role across NSF
- Interagency & international partnerships
- Supports Use-inspired Cyberinfrastructure
  - Research and Education
  - Science and Engineering
- Inherently multidisciplinary with strong ties to all disciplines/directorates

Biological Sciences Directorate

Computer and Information Science and Engineering Directorate

Education & Human Resources Directorate

Engineering Directorate

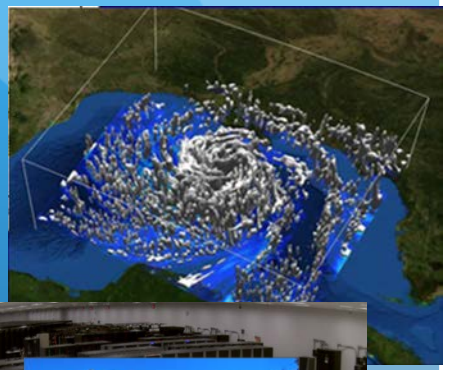
Mathematical & Physical Sciences Directorate

Social, Behavioral & Economic Sciences Directorate

Geosciences Directorate

Advanced Cyberinfrastructure (ACI)

Supports the research, development, acquisition and provision of state-of-the-art CI resources, tools, and services:



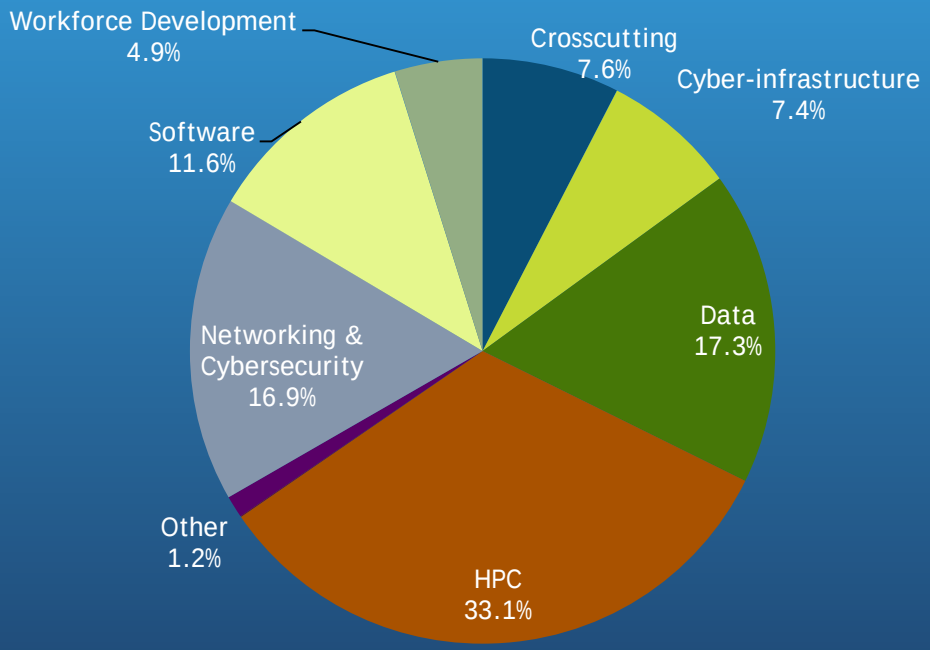
- ***Advanced Computing****:* Provide open-science community with state-of-the-art computational systems ranging from loosely coupled clusters to large scale HPC instruments; develop a collaborative and innovative scientific computational environment.

- ***Data:*** Support scientific communities in the use, sharing and archiving of data by creating building blocks to address community needs in data infrastructure.

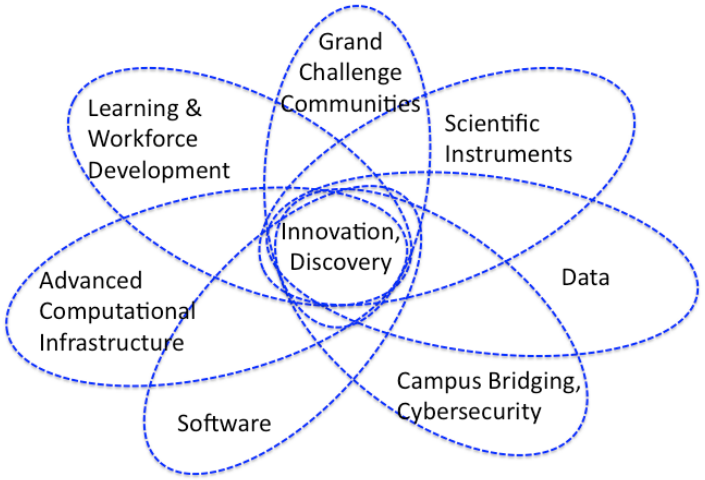
- ***Networking and Cybersecurity****:* Invest in campus network improvements and re-engineering to support computational and data science. Support transition of cybersecurity research to practice.

- ***Software:*** Transform innovations in research and education into sustained software resources (shared tools and services) that are an integral part of cyberinfrastructure.

ACI FY2013 investments reflect a balance across Cyberinfrastructure categories consistent with NSF's CI strategy (CIF21)

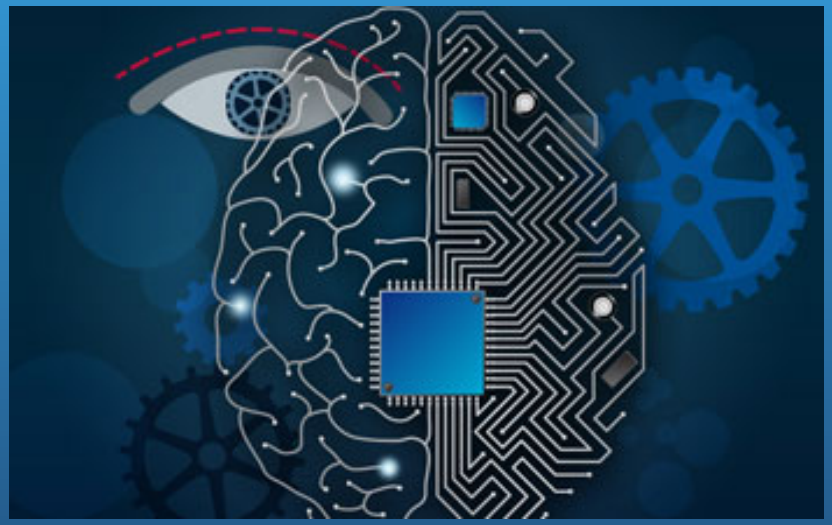


CIF21: cyberinfrastructure as an ecological system

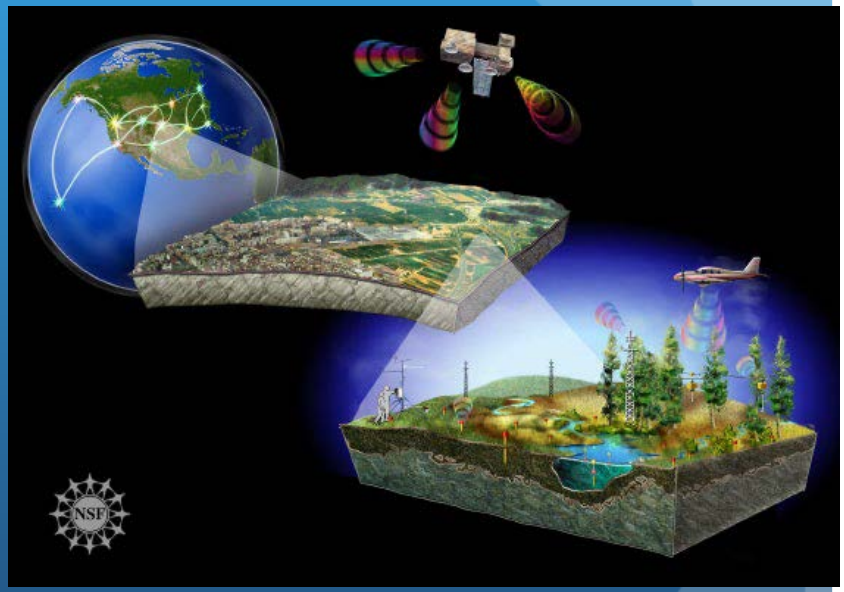


Total ACI FY 2013 funding = $210,772,572

Ubiquity in mobile devices, social networks, sensors, advanced computing and instruments have created a complex data-rich environment ripe for new scientific and engineering advances



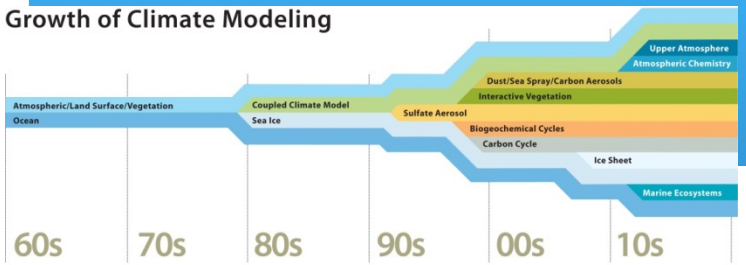
Credit: *Christine Daniloff/MIT*

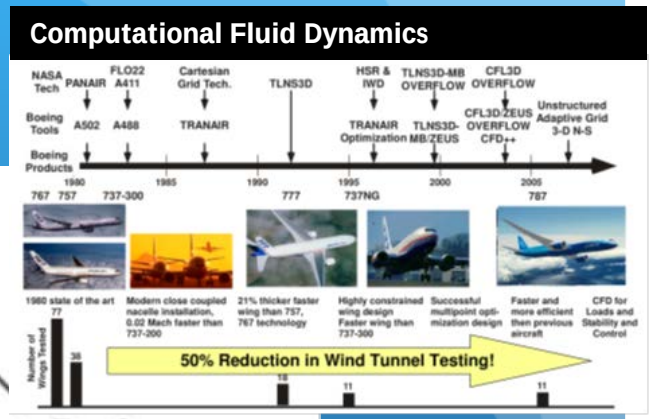


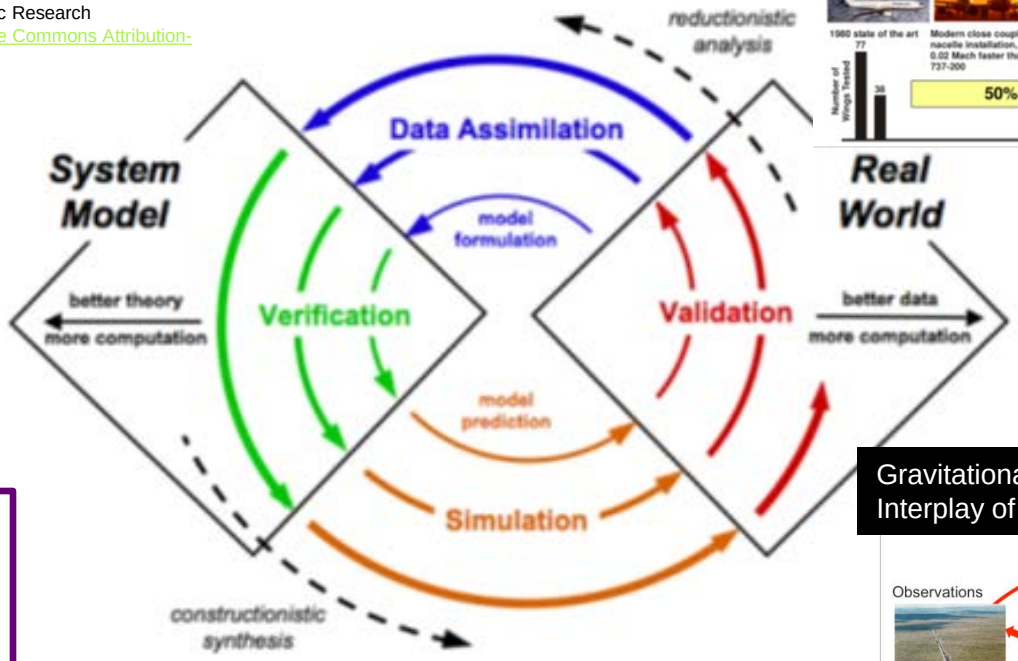
An artist's conception of the National Ecological Observatory Network (NEON) depicting its distributed sensor networks, experiments and aerial and satellite remote sensing capabilities, all linked via cyberinfrastructure into a single, scalable, integrated research platform for conducting continental-scale ecological research. NEON is one of several National Science Foundation Earth-observing systems.

Credit: Nicolle Rager Fuller, National Science Foundation

As models and workflows become more complex and data-rich, collaborative and capable CI is transformative to results and conduct of science





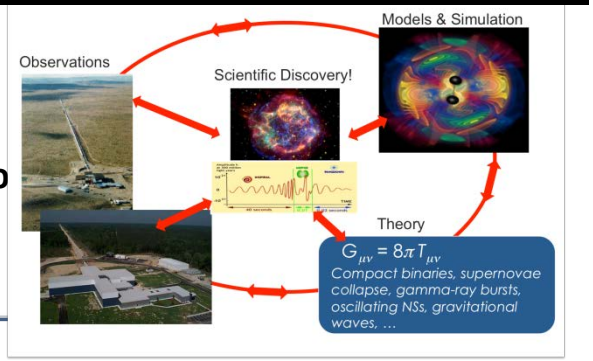


Iterative, computationally intense process of mo formulation and verification in earthquake forecasting

Across many disciplines:

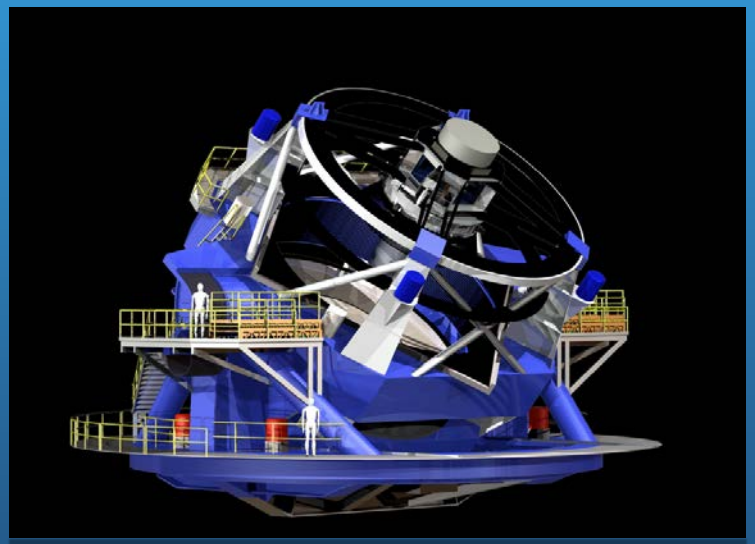
- **Molecular Dynamics**
- **Cosmology & Astronomy**
- **Plasma Physics**
- **Quantum Chromo Dynamics**
- **Seismic Hazards**
- **Gene Sequence Analysis**

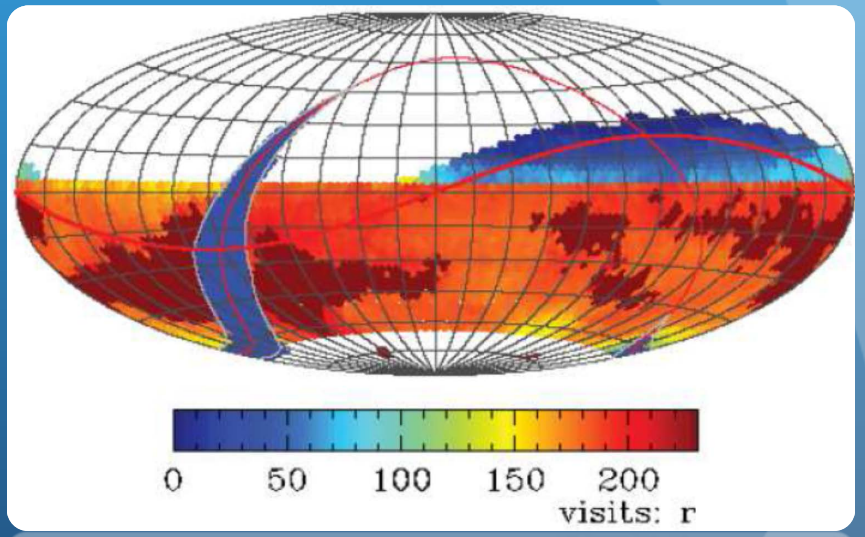
Gravitational Wave Astrophysics
Interplay of Big Data & Big Simulation



LSST: A Deep, Wide, Fast, Optical Sky Survey

8.4m telescope optical (ugrizy) 0.5-1% photometry (sys)
3.2Gpix camera 2 x 15sec exp / 2sec read





Location: Cerro Pachon, Chile
First Light: May 2019

Construction Start: July 2014
Operations: May 2022

LSST Cyberinfrastructure for Data Capture

HQ Site

HQ Facility

Observatory Management
Science Operations
Education and Public Outreach

Archive Site

Archive Center

Alert Production
Data Release Production
Calibration Products Production
EPO Infrastructure
Long-term Storage (copy 2)

Data Access Center

Data Access and User Services

French Site

Processing Center

Data Release Production

Base Site

Base Facility

Long-term storage (copy 1)

Data Access Center

Data Access and User Services

Summit Site

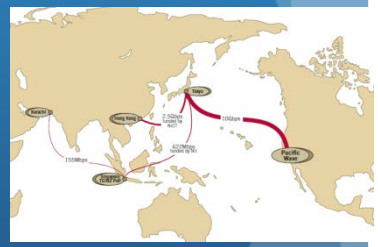
Summit Facility

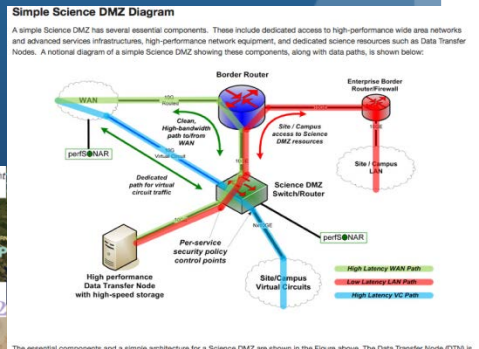
Telescope and Camera
Data Acquisition
Crosstalk Correction

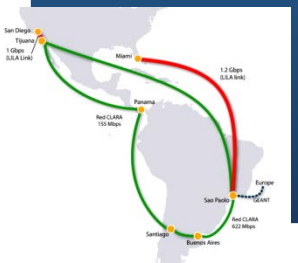
Document LDM-129

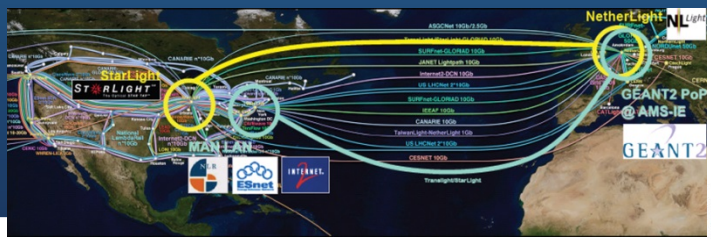
Networking Programs in CISE/ACI

- **Fundamental enabling layer and CI underpinning**
- **CC-NIE (Campus Cyberinfrastructure – Network Infrastructure and Engineering): joint with CNS**
  - Campus networking upgrades (10/100Gbps), re-architecting and innovation
  - Directly responsive to ACCI 2011 Task Force report
- IRNC – International R&E Network Connections: joint with OIA/ISE
  - Scientific discovery as a global collaborative endeavor
  - Provide network connections linking U.S. research with peer networks in other parts of the world
  - Stimulate the deployment and operational understanding of emerging network technology and standards in an international context







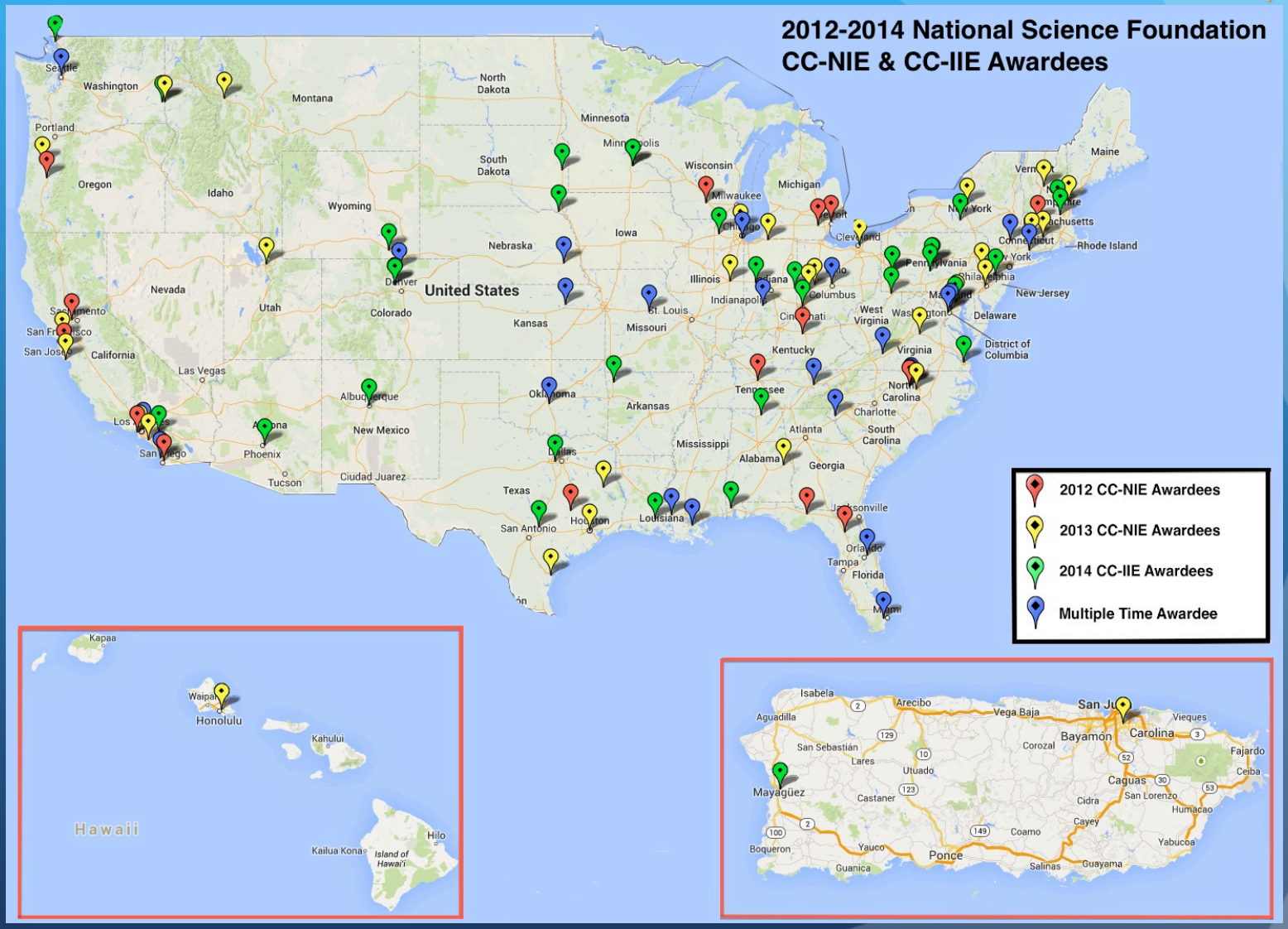


Networking/Cybersecurity -ACI

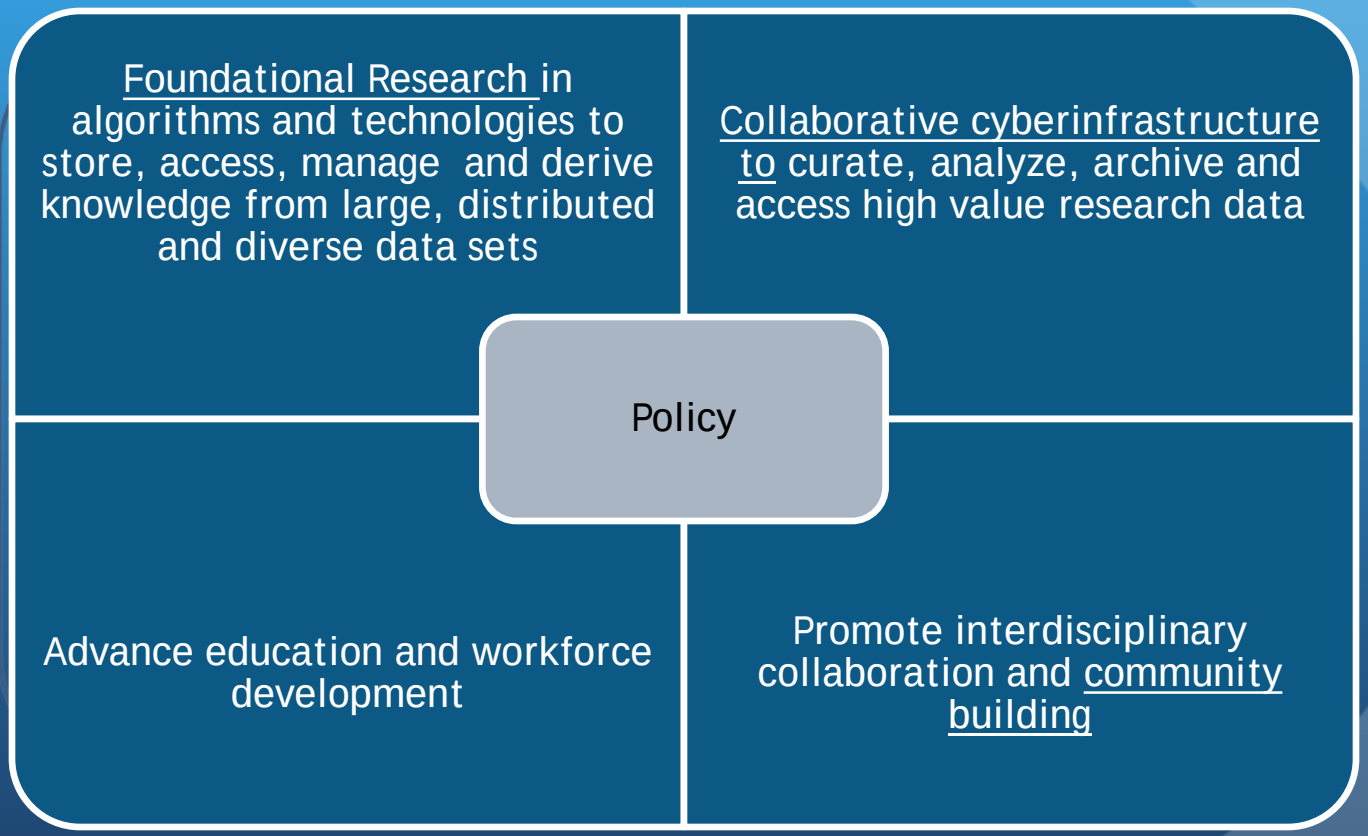
Program Directors: Kevin Thompson, Anita Nikolich

- **Campus Networking (CC-NIE/CC*IIE programs)**
  - Campus Cyberinfrastructure – Infrastructure, Innnovation and Engineering (CC*IIE) – campus network upgrades (10/100Gbps), re-architecting
  - Continue re-forming campus (inter/intra) networking (entering year 4)
  - Address needs of smaller schools, regional coordination, science VO's
- **International networking (IRNC program) NSF 14-544**
  - Scientific discovery as a global endeavor
  - Stimulate deployment of emerging network technology and standards in an international context
- **Secure and Trustworthy Cyberspace (SaTC) NSF 14-599**
  - .ACI provides funding for transition to practice (TTP) projects (7 awards in FY2014)
- **Emerging Cybersecurity Topics**
  - **Secure and Trustworthy Software**
  - **Data Provenance and Privacy**
  - **Coherent Cybersecurity for national CI**
    - Center for Trustworthy Scientific Infrastructure: PI, Von Welch; Indiana University

Transforming the national CI foundation to increase capacity and innovative capability



Overall Data Investment Framework



ACI data focused CI - “A view towards the horizon”

Program Directors: Bob Chadduck, Amy Walton; Irene Lombardo

- **Catalyze transformative** , interdisciplinary **science, & engineering**, research & education through robust, shared, resources, services, & collaborations across diverse communities;

- **Advance capabilities** for robust shared resources to capture, manage, curate, analyze, interpret, archive & share data at unprecedented scales & complexities enabling discoveries in all areas of inquiry and from all facilities, ranging from the campus to the national level;

- Ensure data focused research cyberinfrastructure remains **aligned with research priorities and at the forefront of technologie**s necessary to advance fundamental science and engineering;

- Ensure that the **future workforce of scientists, engineers & educators** are equipped with skills to make use of, and build the next generation of data focused cyberinfrastructure

- **Contribute to** advances in comprehensive **policies** for data, software, publications & other digital outputs

FY13 data awards - towards an integrated portfolio for science, engineering, research & education

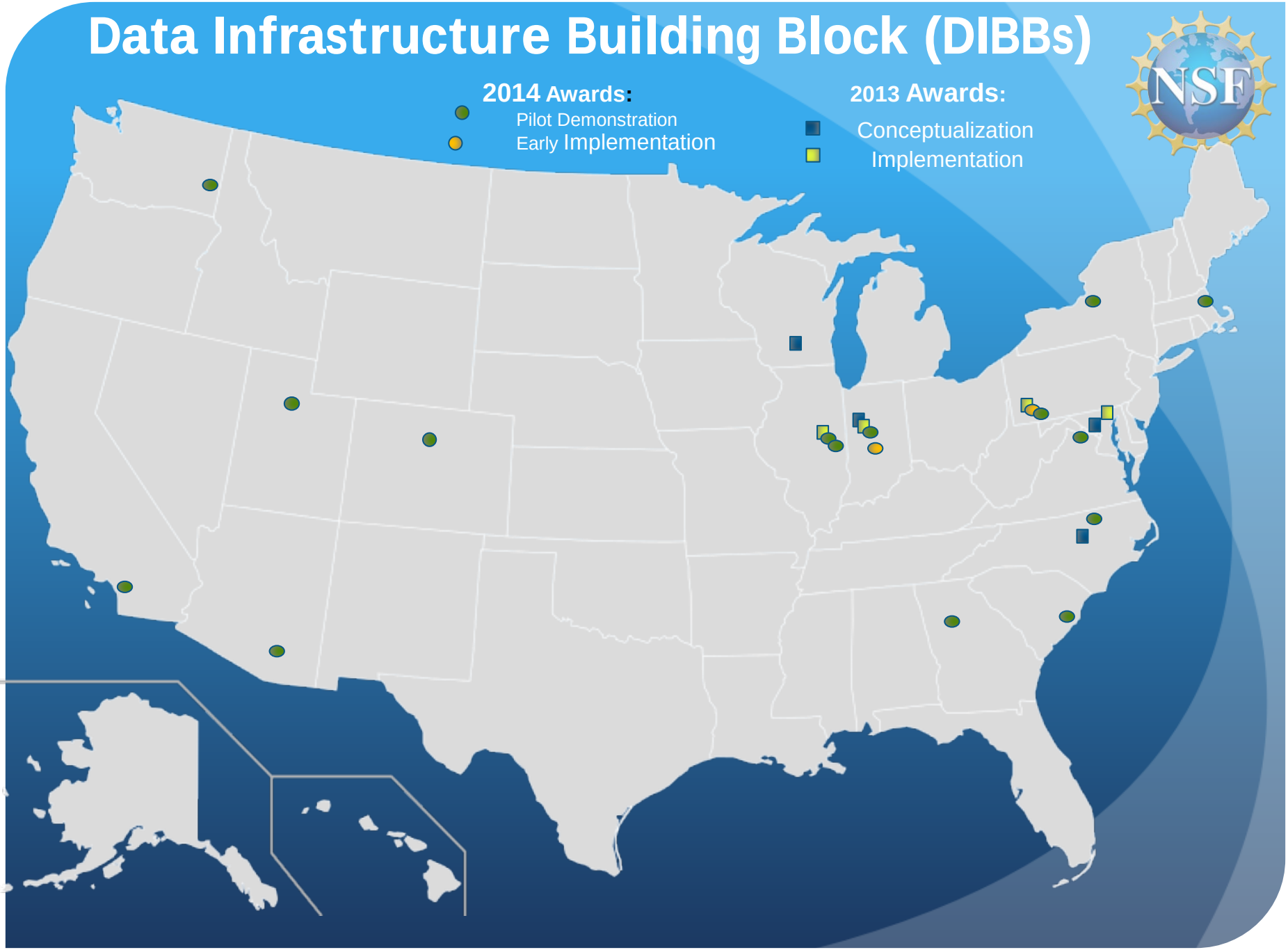
- HPC Acquisition Data Resource -**TACC "Wrangler" –** PI: Dan Stanzione; Novel data intensive CI resource advancing open science communities;

- DIBBS – Data Information Building Blocks
  - **NCSA "Brown Dog"** – PI: Kenton McHenry; CI enabled tools & capabilities advancing integration of large unstructured, heterogeneous data collections as resources in science;
  - **Purdue "Geospatial data Analysis Building Blocks" (**GABBs)– PI: Carol Song; Novel HUBzero CI capabilities enabling creation & sharing of geospatial data & modeling tools;
  - **JHU "SciServer**" (formerly "SkyServer") – PI: Alex Szalay; Novel data focused CI responsive to interests in multiple large scientific data sets serving broad science communities;
  - **PSC "Exacell**" - PI: Mike Levine – Novel data focused CI building blocks & hardware supporting large scale analytics enabling data intensive scientific research;

- Research Collaboration Network – **RPI "Research Data Alliance**" – PI: Fran Berman; data focused CI pilots, tools and capabilities developments responsive to global communities' interests to share & access data catalyzing science advances;

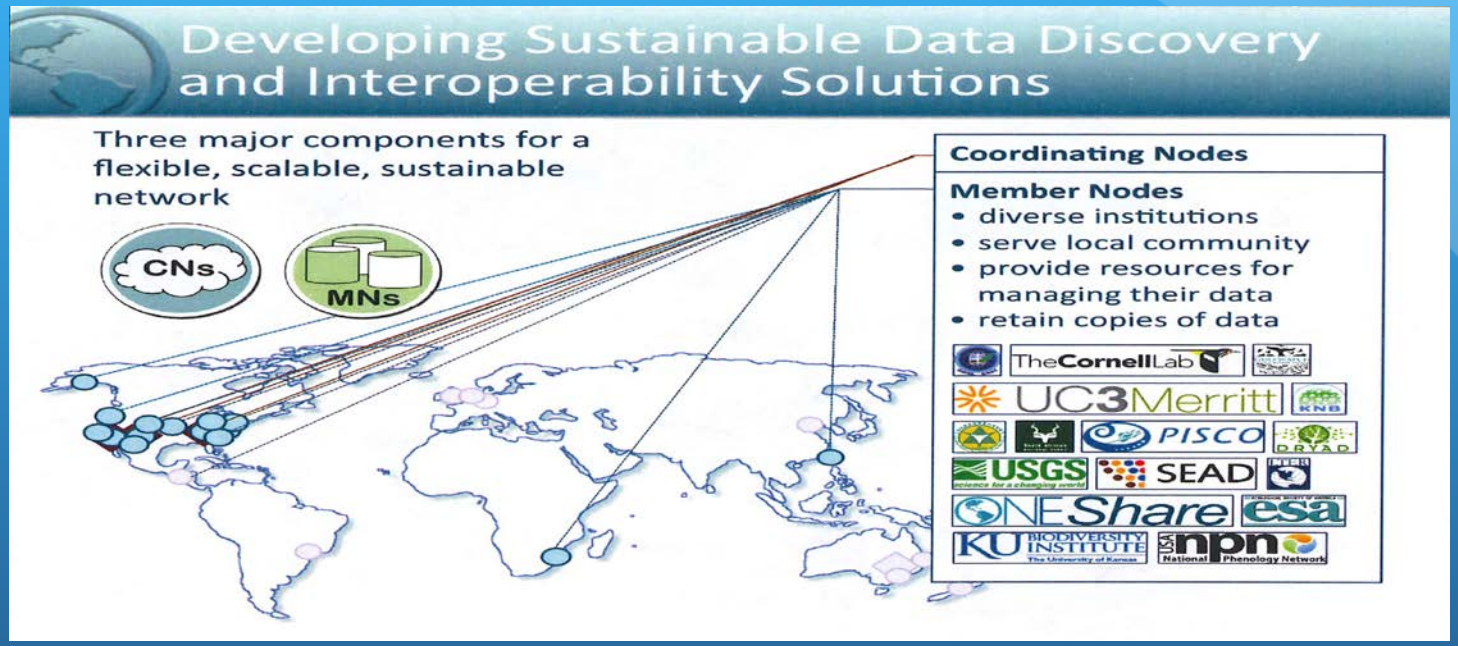
- Big Data **- UChicago "Open Flow Enabled Hadoop over Local & Wide Area Clusters"** PI: Bob Grossman – integration of OpenFlow enabled switches with Hadoop to handle more efficiently large data flows arising in data intensive computing

Data Infrastructure Building Block (DIBBs)



DataONE (Observation Network for Earth) Award

William Michener, University of New Mexico



- DataONE has interacted with almost 20,000 users; has 462,00 objects, 150,000 datasets and about 200,000 metadata records. It is a large and diverse community with 55 partnering projects and 300 plus collaborators with 170 plus persons that are active in the DataONE Users Groups as well as many and varied training workshops.

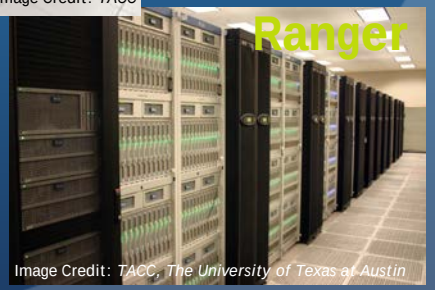
- DataONE has been actively involved with federal, local government agencies, as well as industry (such as Microsoft), contributed to OSTP lead safety data challenge to improve public preparedness & emergency response, advised OSTP on Big Data challenges, advised the Federal Reserve Bank regarding data focused cyberinfrastructure topics.

Advanced Computational Infrastructure

Program Directors: Rudi Eigenmann, Bob Chadduck; (Tom Russell, Irene Lombardo)

- Anticipate and invest in diverse and innovative national scale shared resources, outreach and education complementing campus and other national investments
- Leverage and invest in collaborative flexible “fabrics” dynamically connecting scientific communities with computational resources and services at all scales (campus, regional, national, international)

XSEDE
Extreme Science and Engineering Discovery Environment



In 2012 NSF published its Advanced Computing Infrastructure Vision and Strategy (12-051)

- Vision: Position and support the entire spectrum of NSF-funded communities at the cutting edge of advanced computing technologies, hard, and software

- Strategies
  - Foundational research to full exploit parallelism and concurrency
  - Application research and development in use of high-end computing resources
  - Building, testing and deploying resources into a collaborative ecosystem
  - Development of comprehensive education and workforce programs
  - Development of transformational and grand challenge communities

In 2013 Major Computing NSF Computing Infrastructure deployed

Blue Waters, UIUC

Stampede, UT Austin

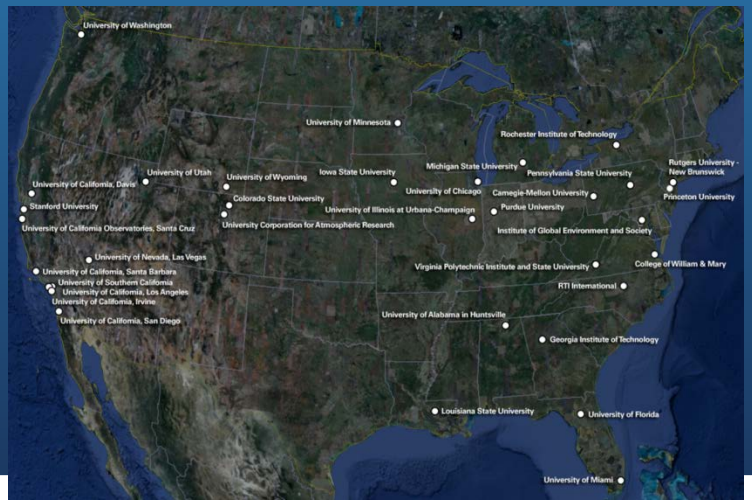
NCAR/ Wyoming Supercomputing Center

Blue Waters: Grand Challenge Computational Science/ Engineering through Sustained Petascale Performance

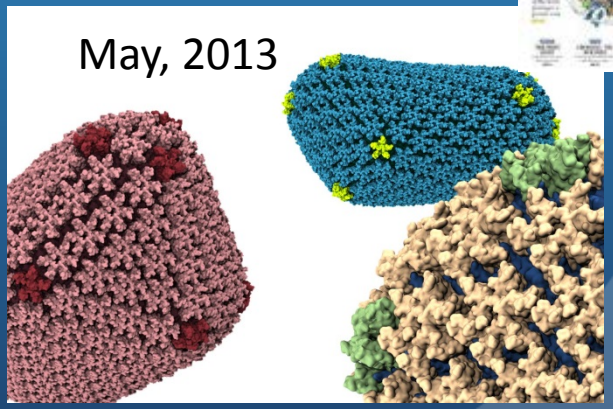
UIUC Data Center

Cray XE6/XK7 accepted December, 2012

Petascale Application Projects



May, 2013



Credit: *Theoretical and Computational Biophysics Group (www.ks.uiuc.edu), Beckman Institute for Advanced Science and Technology, UIUC*

Stampede is both innovative and highly capable, doubling the resource pool for XRAC/XSEDE allocations

A YEAR WITH STAMPEDE

Stampede, one of the most powerful supercomputers in the world for open science research, celebrated its first birthday on January 7, 2014, by completing more than 75,000 years of scientific computations – not bad for a one-year-old. Here are some facts, figures & science highlights that capture the comprehensive impact of the system.

Funded by the National Science Foundation Grant ACI-1134872 and built in partnership with Intel, Dell and Mellanox, Stampede and its academic partners will continue to enable promising computational research in 2014 and beyond.

WORLD RANKING 7th MOST POWERFUL
www.top500.org

1247 PROJECTS

2,196,848 COMPLETED JOBS

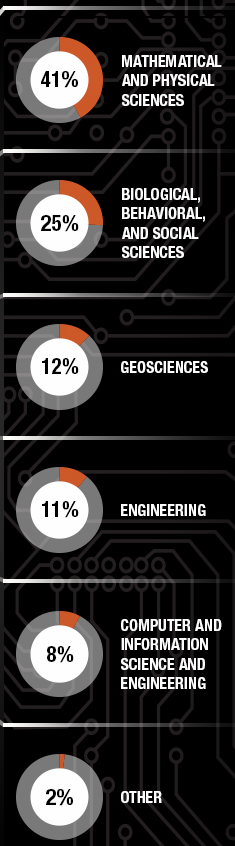
3400 RESEARCHERS

9.6 QUADRILLION Floating Point Operations per second

TACC Texas Advanced Computing Center
www.tacc.utexas.edu
www.tacc.utexas.edu/stampede
#useTACC #UltimateScienceEnviroment

NSF XSEDE THE UNIVERSITY OF TEXAS AT AUSTIN

SCIENCE FIELDS

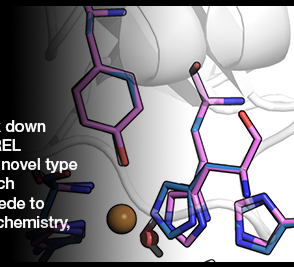


SCIENCE HIGHLIGHTS

Stampede supports the largest number of open science projects in the world across science and engineering domains. Below are three recent highlights:

BIOFUEL PRODUCTION

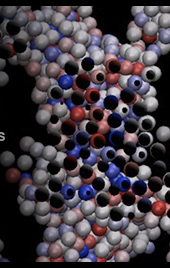
Researcher: Gregg Beckham
National Renewable Energy Laboratory (NREL)



Stampede is helping to determine how enzymes break down cellulose to improve biofuel production. A group of NREL researchers used the supercomputer to predict how a novel type of oxidative enzyme can speed up the process by which cellulose breaks down. The group is also using Stampede to design catalysts for high-temperature deoxygenation chemistry, which is important to convert biomass to fuel.

BIOMEDICINE AND SMART MATERIALS

Researcher: Roseanna Zia
Cornell University



Colloidal gels have huge promise in biomedicine. Comprised of microscopic particles suspended in a solvent, these gels form networks of particles that support their weight under gravity. For this reason, these soft solids can be used as injectable pharmaceuticals and artificial tissue scaffolds; however, they are beset by stability problems. Stampede enabled Cornell researchers to conduct the largest, longest simulation of a colloidal gel, which is answering important questions about the structure, dynamics, and stability of the particle network.

SUPERNOVA EXPLOSION

Researcher: Philipp Mösta and Christian D. Ott
California Institute of Technology (CalTech)

Using Stampede, astrophysicists succeeded in performing the first fully general-relativistic 3D MHD simulations of progenitor stars that are believed to lead to very energetic, jet-driven supernova explosions. The researchers found that the simulations behave very differently in full unconstrained 3D compared to the same model simulated with imposed symmetries. Stampede's per-core performance pushed these simulations to the limit.

NSF

National Academies Study launched in 2013: Future Directions for NSF Advanced Computing Infrastructure to support US Science in 2017-2022

- **Bill Gropp/University of Illionois at Urbana-Champaign**
- **Robert Harrison/Stony Brook University**
- Mark Abbott/Oregon State University
- David Arnett/University of Arizona
- Robert Grossman/University of Chicago
- Peter Kogge/University of Notre Dame
- Padma Raghavan/Penn State University
- Dan Reed/University of Iowa
- Valerie Taylor/Texas A&M
- Kathy Yelick/UC Berkeley

http://sites.nationalacademies.org/CSTB/CurrentProjects/CSTB_087924

Goals for this Study- Final Report

By Summer, 2015, the final report will yield insights such as:

- **The contribution of high end computing to U.S. leadership and competitiveness in basic science and engineering**, **the roles that NSF can play in sustaining this leadership**, and how leadership can be measured, benchmarked, and monitored.
- **Expected future national-scale computing needs across the full range** of basic science and engineering research supported by NSF.
- Tradeoffs among investments in computing, software, data, and networking infrastructure and between those that advance the computational frontiers vs. those that focus on delivering more aggregate computing capacity.
- The **roles of different models for advanced computing infrastructure,** including NSF centers and consortia, campus-based infrastructure, and the commercial marketplace.
- **Technical challenges** to delivering needed computing capabilities, and what research and development may be needed to deliver expected future capabilities.

Goals for the Study – Interim Report

By Summer, 2014, the **interim** report will focus on the 2017-2020 timeframe and yield insights on key issues such as:

- Characterization of the **trajectory and relevance of large scale simulation's impact on foundational advances** in science and engineering.

- Identification of **scientific research grand challenges that will be substantially advanced by large scale data analytics and data mining** not currently possible in research infrastructures.

- Based on the response to the previous two bullets, identify categories for research cyberinfrastructure investments (e.g. emergent technologies and algorithms, balance between experimental and "production", education and workforce development, community software) required to support sustained advances in U.S. science

- **Challenges and responses by research infrastructures at all scales** (e.g. campus, regional, national; problem focused or multipurpose) to the items above, identifying those which can be most positively impacted by NSF. These should encompass economic, cross agency and international considerations.

NSF Supported National Resources are increasingly diverse and collaborative

Trestles
IO-intensive
10k cores
160 GB SSD/Flash

Gordon
Data intensive
64 TB memory
300 TB Flash Mem

Comet
"Long Tail Science"
47k cores/2 PF
High throughput

Yellowstone
Geosciences

Blue Waters
Leadership Class

Wrangler
Date Analytics

Stampede
460K cores
w. Xeon Phi
>1000 users
Upgrade in 2015

Maverick
Visualization
Data Analytics

Lonestar*
Large memory

Open Science Grid
High throughput

FutureGrid*
CS testbed

Blacklight
Shared Memory
4k Xeon cores

SuperMIC
380 nodes – 1PF
(Ivy bridge, Xeon Phi, GPU)

Darter
24k cores

Nautilus
Visualization
Data Analytics

Keeneland
CPU/GPGPU

To be deployed in FY2015 with XSEDE integration and user services

XSEDE integration

** To be retired in 2014*

Recent ACI supported acquisitions & deployments are complementary

Xeon Upgrade

Renewal?

Long-tail/high throughput

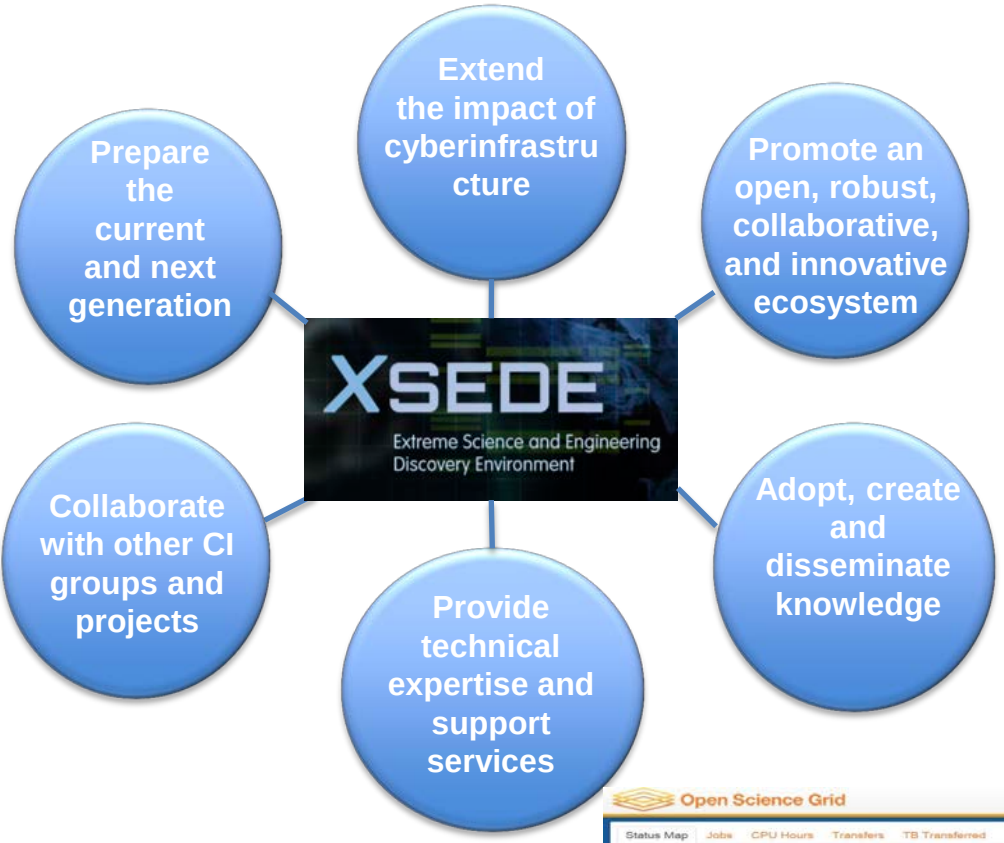
Data Analytics

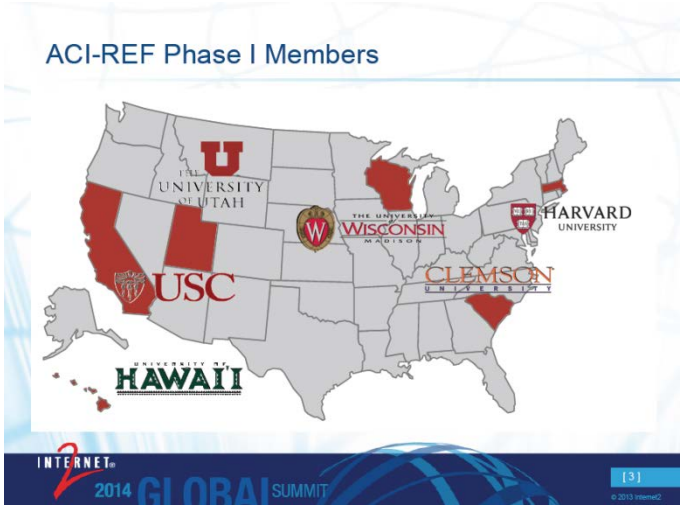
NSF 14-536

TBA

2013 2014 2015 2016 2017

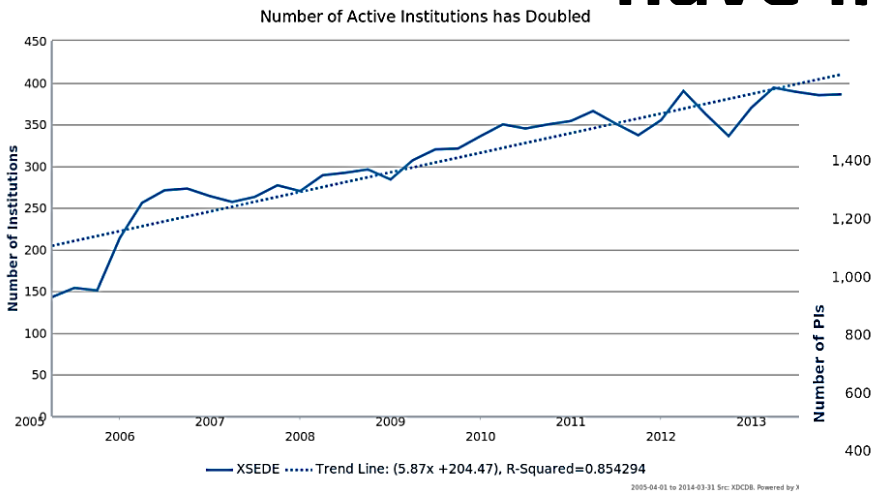
Continued investment in evolving and new models for collaborative CI to advances science frontiers

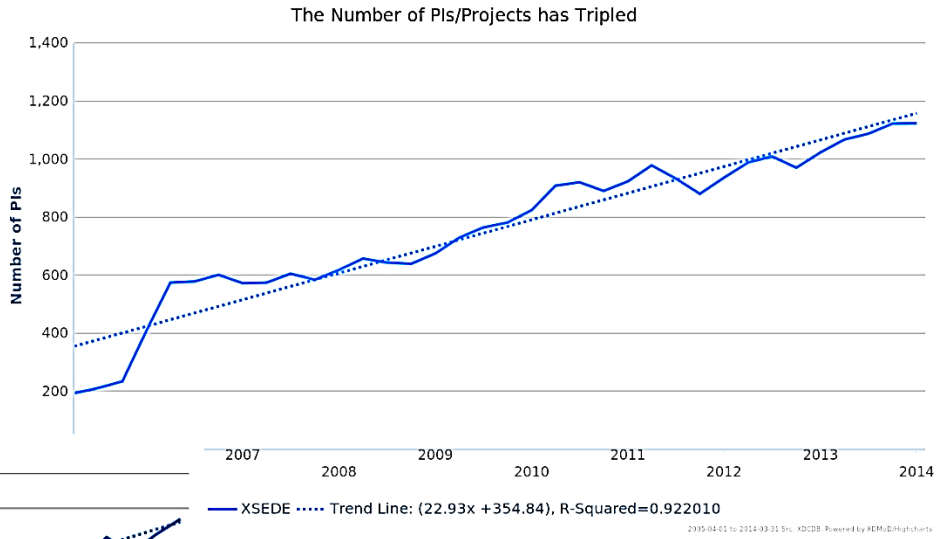


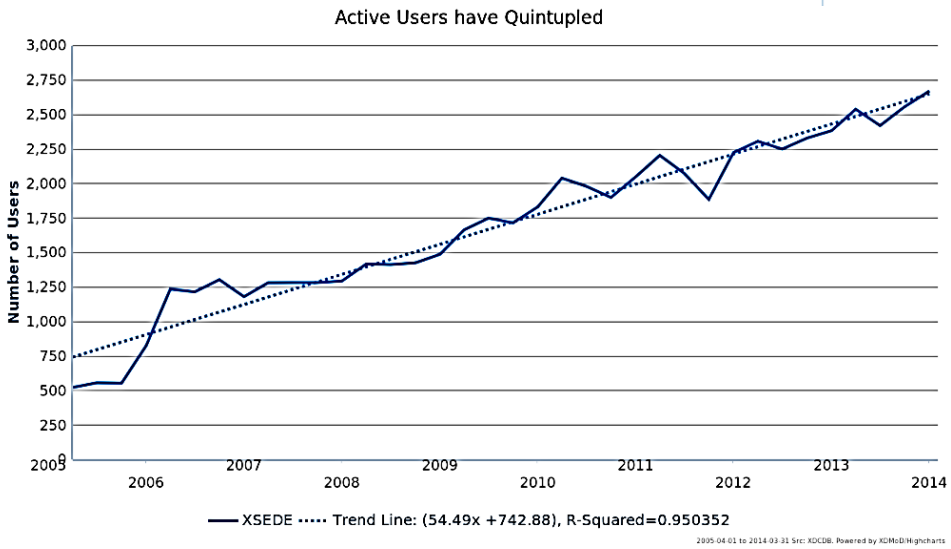


XDMoD
METRICS ON DEMAND

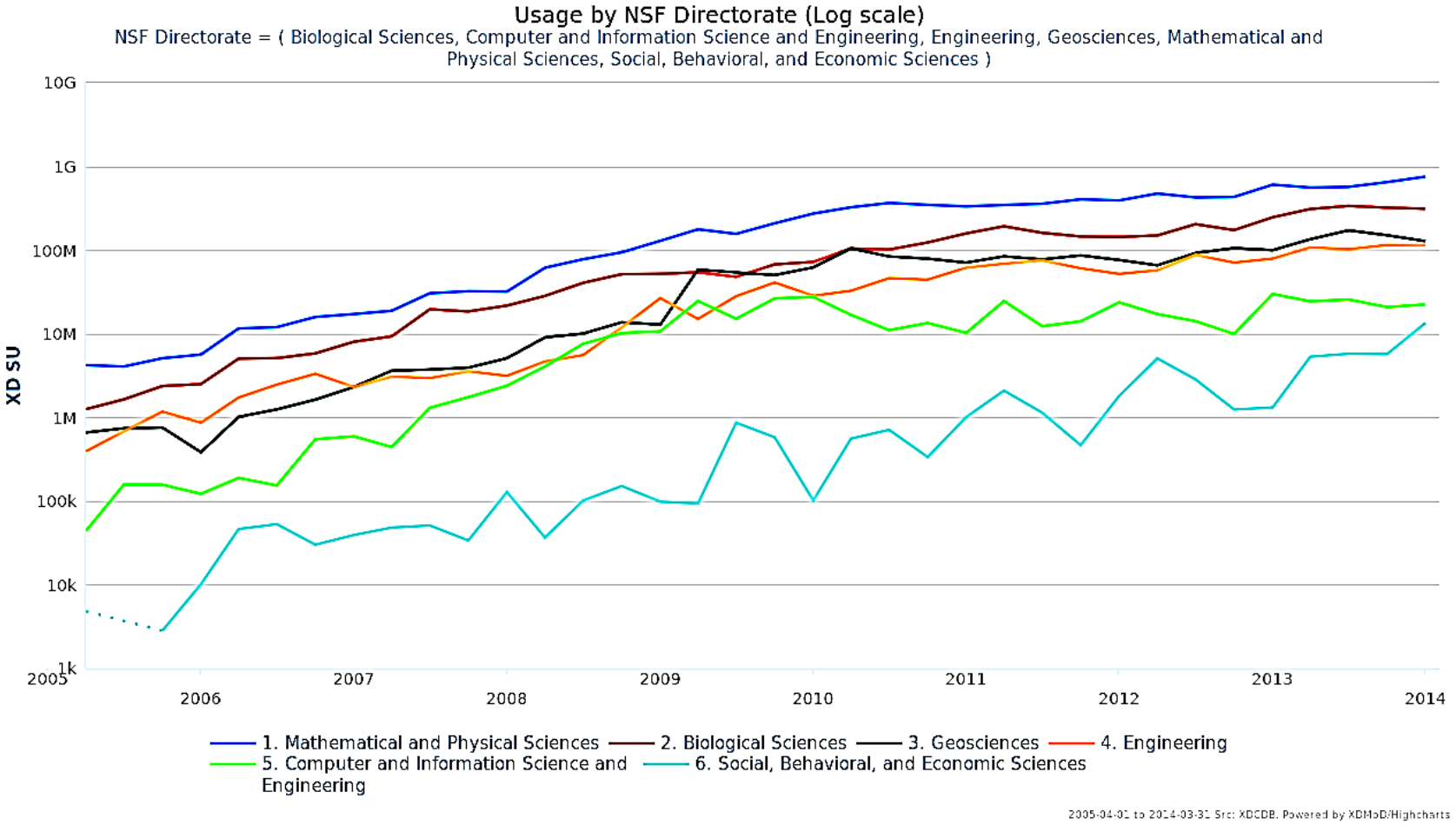
Quantitative Trends: the Number of Research Projects, Institutions and Users have increased



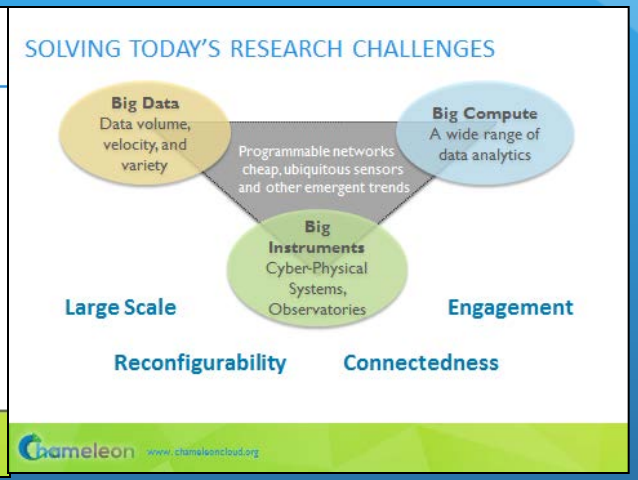


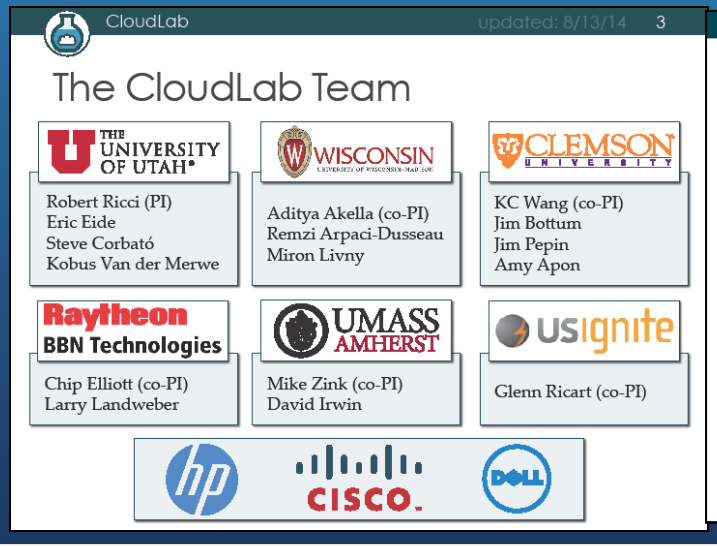


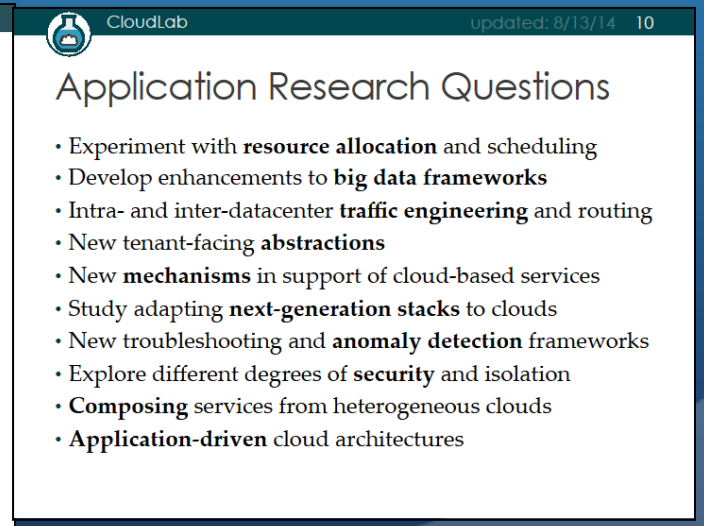
All Scientific Disciplines have increased use of national resources



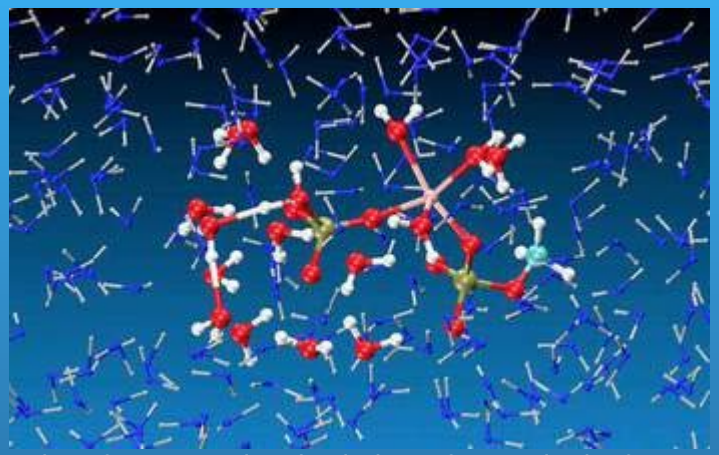
2014 NSF Cloud Awards: Multi-institutional CI Research







Computational science takes the Nobel stage



Snapshots from extensive simulations of a step in the free energy perturbation study of phosphate hydrolysis in water. The simulations involve a QM/MM surface with a description of the phosphate (red and white) from first principle and an MM description of the water molecules (blue).

Credit: *Arieh Warshel*

The molecular dynamics methods that Karplus, Warshel, and Levitt pioneered using NSF supported HPC systems have evolved into the standard approach to investigate complex chemical and biochemical processes and the behavior of materials.

Michael Levitt, 2013 Nobel Laureate in Chemistry

Credit: *Alexander Mahmoud. Copyright*

Arieh Warshel, 2013 Nobel Laureate in Chemistry

Credit: *Alexander Mahmoud. Copyright Nobel Foundation 2013*

Martin Karplus, 2013 Nobel Laureate in Chemistry

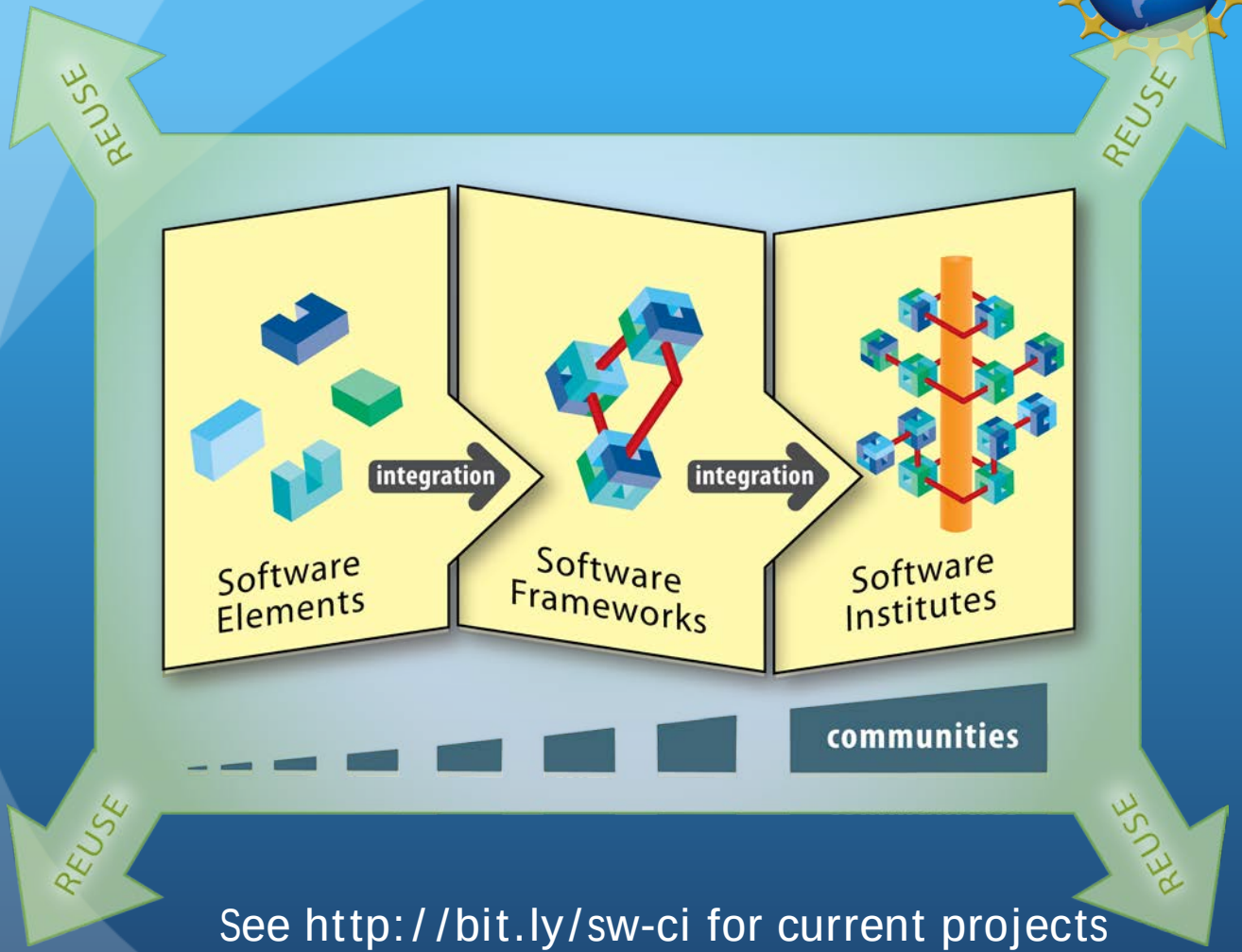
Credit: *Alexander Mahmoud. Copyright Nobel Foundation 2013*

NSF Software Infrastructure Projects

5 rounds of funding, 65 SSEs

4 rounds of funding, 35 SSIs

2 rounds of funding, 14 S2I2 conceptualizations



See http://bit.ly/sw-ci for current projects

Software as Infrastructure- Role & Lifecycle

Create and maintain a software ecosystem providing new **capabilities** that advance and accelerate scientific inquiry at unprecedented complexity and scale

Support the foundational **research** necessary to continue to efficiently advance scientific software

Enable transformative, interdisciplinary, collaborative, **science and engineering** research and education through the use of advanced software and services

Transform practice through new policies for software, addressing challenges of academic culture, open dissemination and use, reproducibility and trust, curation, sustainability, governance, citation, stewardship, and attribution of software authorship

Develop a **next generation diverse workforce** of scientists and engineers equipped with essential skills to use and develop software, with software and services used in both the research and **education** process

Working Towards Sustainable Software for Science: Practice and Experiences

- https://sites.google.com/site/si2pimeeting2014/agenda
- http://wssspe.researchcomputing.org.uk
- First Workshop on Sustainable Software for Science: Practice and Experiences (WSSSPE1), @ SC13, 17 November 2013, Denver
  - 2 keynotes, 54 accepted papers
  - Discussion sessions: Developing software; Policy; Communities
  - Cross-cutting (emergent) topics: Defining sustainability; Career paths
  - Post-workshop paper: http://arxiv.org/abs/1404.7414
- WSSSPE1.1, @ SciPy2014, Austin
- WSSSPE2, @ SC14, 16 November 2014, New Orleans
  - 29 accepted papers
  - Keynotes/sessions being planned

Advanced Cyberinfrastructure

Supports the research, development, acquisition, and provision of state-of-the-art CI resources, tools, and services:

- ***High Performance*** computing; provide the-art HPC assets large scale instrum environment.
- ***Data:*** Support scie archiving of data by community needs i
- ***Networking and C*** improvements and activities in modern of cybersecurity res
- ***Software:*** Transform innovations in research and education into sustained software resources that are an integral part of cyberinfrastructure.

PEOPLE

Learning and Workforce Development

Workforce as Cyberinfrastructure

CI-focused
Cyber Scientists
to develop, pilot and deliver
new capabilities

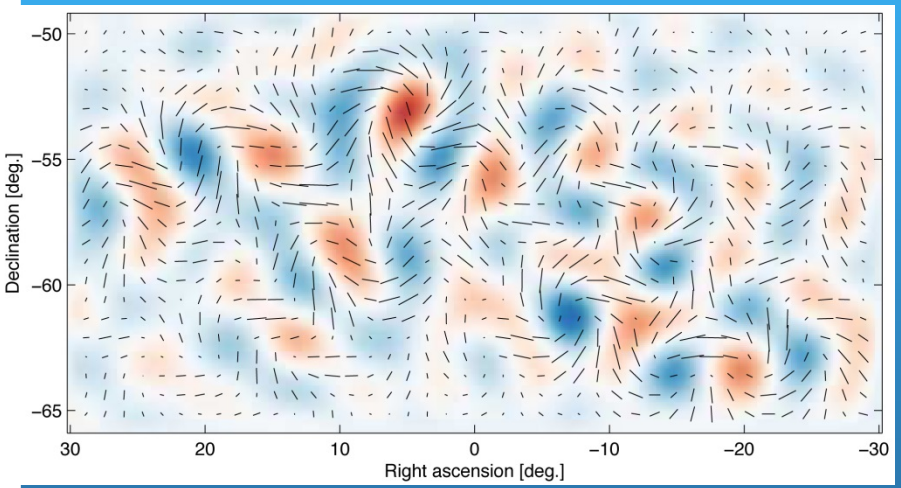
Computational Scientists
Data Scientists
Architects/Designers
Software Engineers
System Management

CI-enabled
Domain Scientists
To explore and exploit
new capabilities

CAREER: Sharing Deep Cosmic Microwave Background (CMB) Maps for Cosmological Discovery

John M. Kovac, Harvard University

supported by MPS/AST, MPS/PHY, CISE/ACI, GEO/PLR



The actual B-mode pattern observed with the BICEP2 telescope, with the line segments showing the polarization from different spots on the sky. The red and blue shading shows the degree of clockwise and anti-clockwise twisting of this B-mode pattern. Credit: *BICEP2 Collaboration*

- Derive needed portable data products and software tools, test them in key joint analyses of overlapping maps with external collaborators
- Share data products and software tools with the full astrophysics community
- Encourage independent reanalysis of findings

Project aims supported by CISE/ACI:

- Share ultra-deep CMB polarization maps from SPUD, BICEP2, BICEP1, and DASI
- Develop and disseminate complete data products and software tools that broaden the use and impact of path finding surveys while setting what are expected to be useful standards for others to do the same

The sun sets behind BICEP2 and the South Pole Telescope at NSF's Amundsen-Scott South Pole Station. Credit: *Steffen Richter, BICEP*

Vetria L. Byrd

REU Site: Undergraduate Research in Collaborative Data Visualization Applications

Clemson University
Clemson, South Carolina
vlbyrd@clemson.edu

Research areas of this site: Visualization, Computer Science, Genetics and Biochemistry, Geophysics, Sociology, Molecular Modeling and Simulation, Inorganic Chemistry, Social Media, Parks Recreation Tourism Management, Biological Sciences and Digital Humanities

Unique Feature in 2014

The program identifies research collaborators with visualization needs and assigns each student to a REU team. Each REU Team consists of a research/science collaborator, a visualization mentor and a visualization REU student. This arrangement fosters further collaboration among team members, an appreciation of the visualization process and an understanding of the role visualization plays in discovery and analysis. REU students participate in activities as a member of their REU Team, their Research Lab, and their REU cohort thereby creating a rich, multidisciplinary research experience.

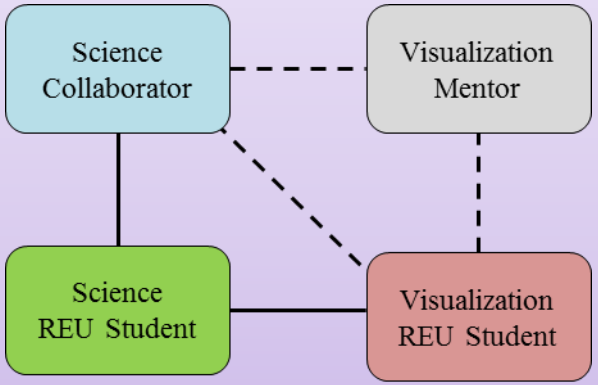


Figure 1: REU Team.

ACI LWD

Name of Program	Solicitation Number
CAREER	*NSF 11-960*
REU Sites	NSF 13-542
EXTREEMS-QED	NSF 12-606
STEM-C Partnerships	NSF 14-522
Cyberlearning	NSF 14-526
NRT (formerly CIF21 IGERT)	NSF 14-548

Current Research Context

- Changing practice of science: interdisciplinary, team-oriented, global, data intensive, complex work and data flows increasingly integrated with technology
- The power and challenges of technology: instrumenting everything; computational and data learning and workforce development; ubiquitous connectivity
- Shifting funding landscape and role for foundational research in the face of escalating global challenges
- Changing demographics: diversity, increased need for more computational and data scientists

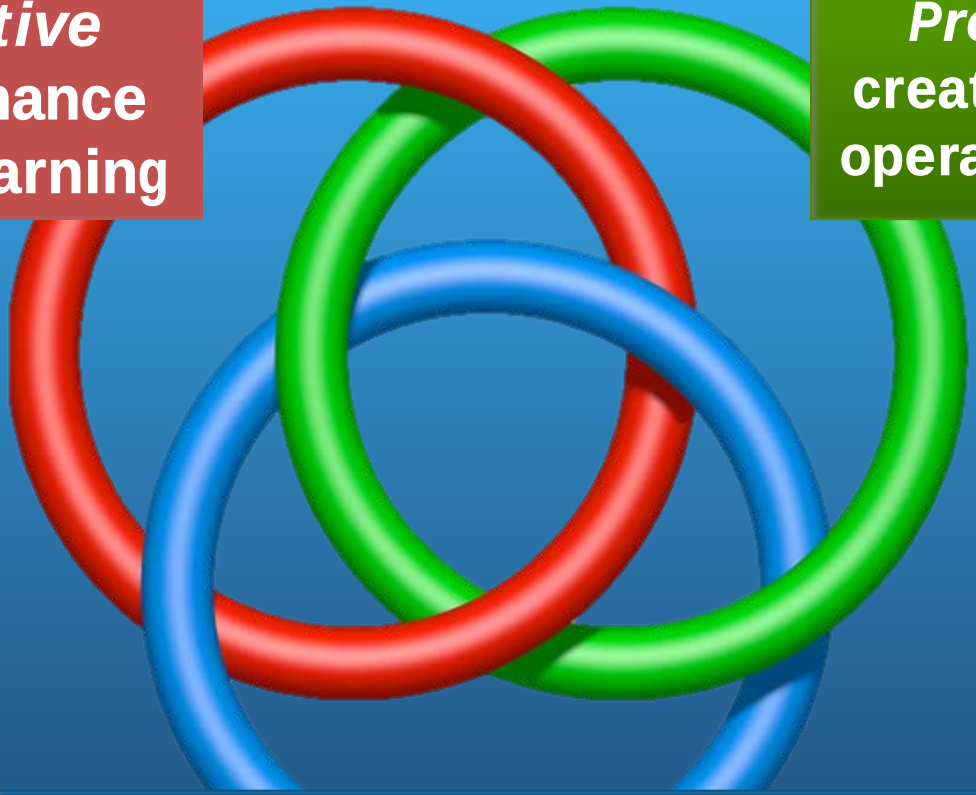
Research Cyberinfrastructure Response

- Cooperating campus, national and global cyberinfrastructure at all scales
- Sustainability Model Development, especially for burgeoning CI areas of software and data
- Ubiquitous, capable, secure and facile CI access for more researchers, educators, institutions and communities
- Integrated Learning and Workforce Development (LWD) for both CI creators and users

Advanced Cyberinfrastructure

Transformative CDS&E **to enhance discovery & learning**

Provisioning **to create, deploy, and operate advanced CI**



Research and Exploration of future CI environments for science and engineering

Thank you!

Contact Information:
iqualter@nsf.gov
+1 703 292 2339