# NSF Transforming Science: Computational Infrastructure for Research


BDEC Presentation

January 29 2015

Irene Qualters

Division Director/ACI

# NSF Core Mission: Support Fundamental Open Research across all Disciplines

$7.1 billion FY 2014 research budget

94% funds research, education and related activities

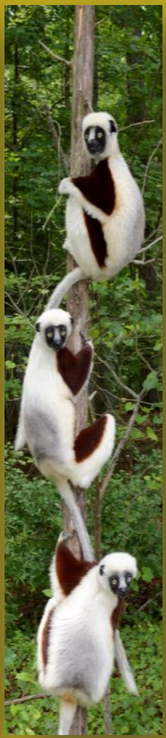
50,000 proposals

11,000 awards funded

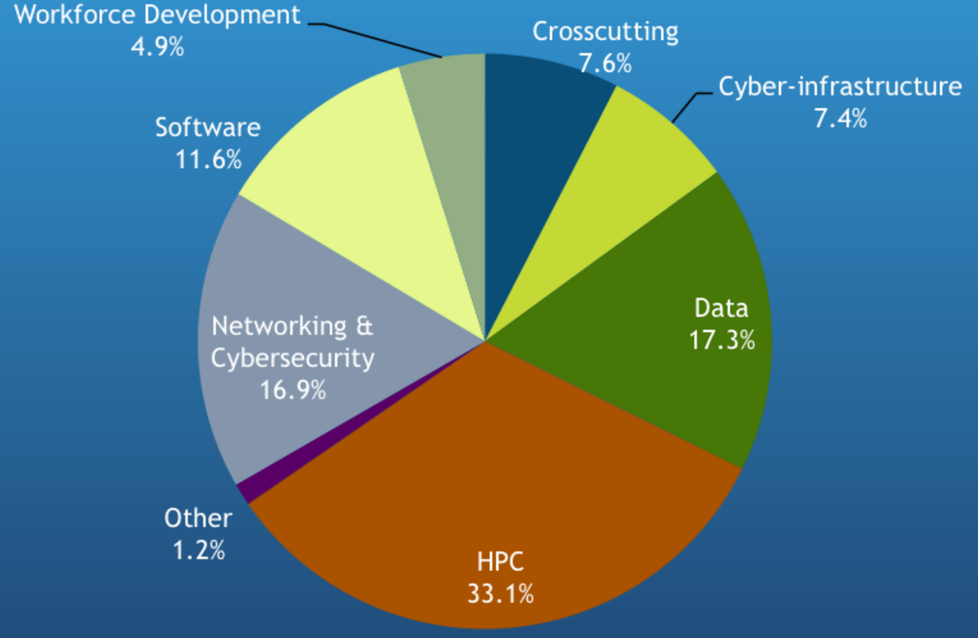
2,000 NSF-funded institutions

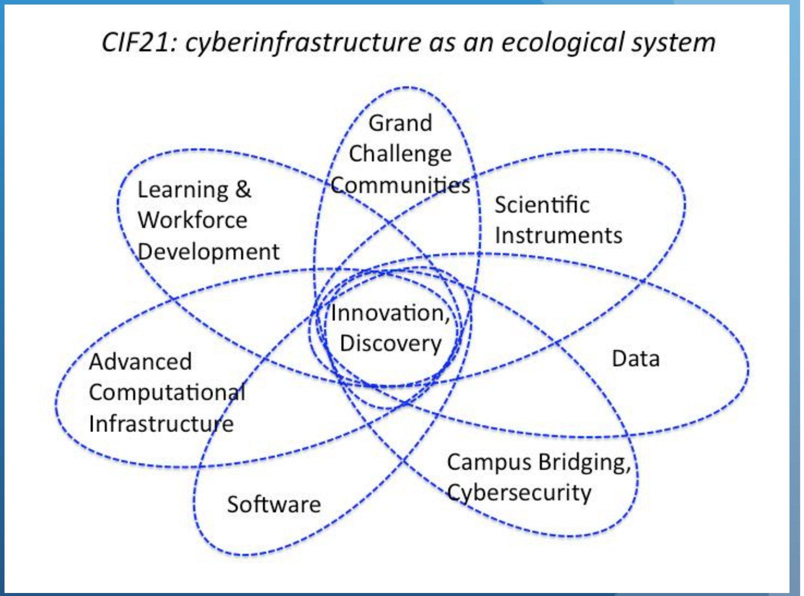
300,000 NSF-supported researchers

## Fundamental Research



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




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

# In 2012 NSF published its Advanced Computing Infrastructure Vision and Strategy (12-051) following several years of community input

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

- Strategies
  - Foundational research to fully 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

# National Academies Study launched in 2013: Future Directions for NSF Advanced Computing Infrastructure to support US Science in 2017-2020

- **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

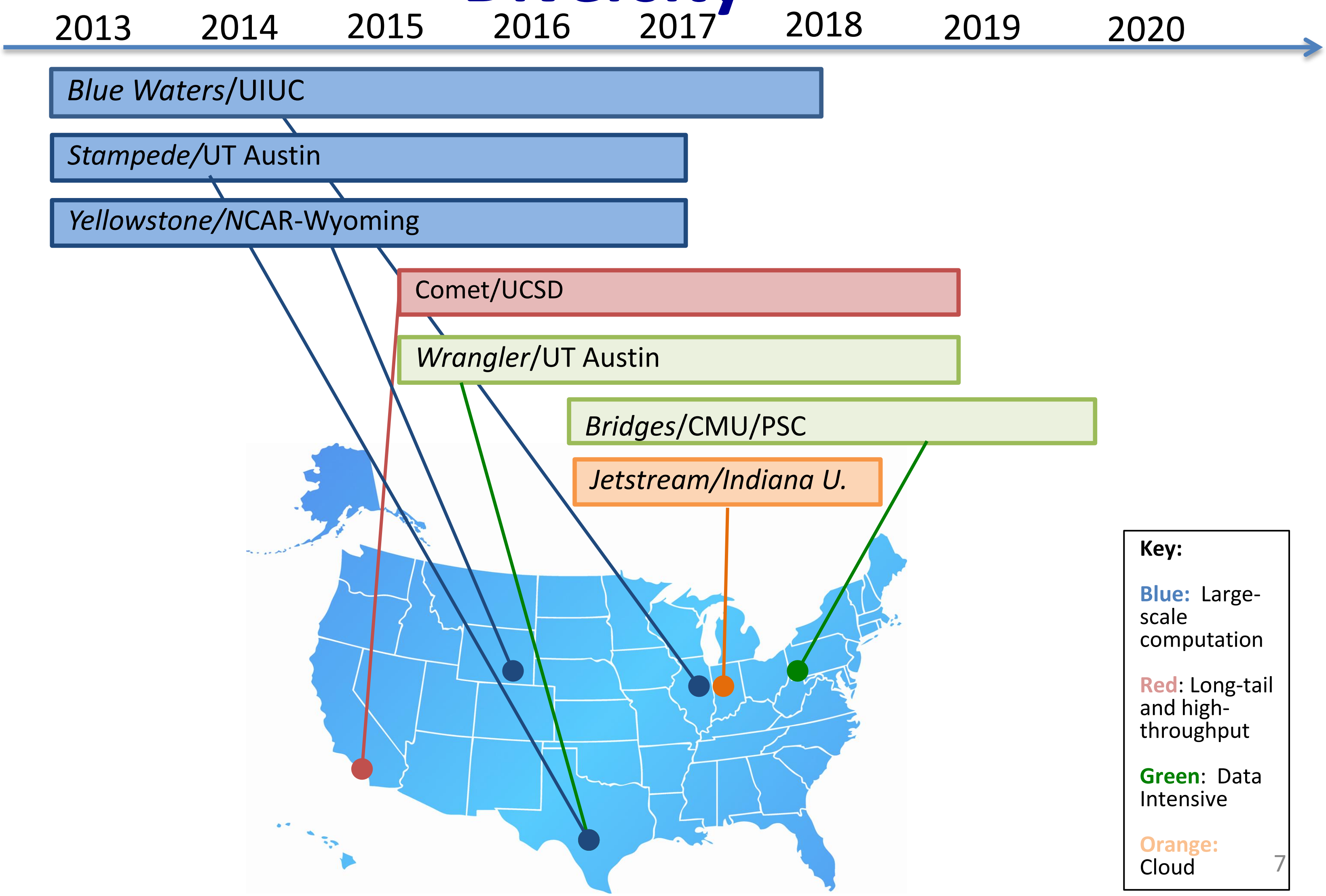
http://sites.nationalacademies.org/CSTB/CurrentProjects/CSTB_087924

# Goals for this Study- Final Report

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

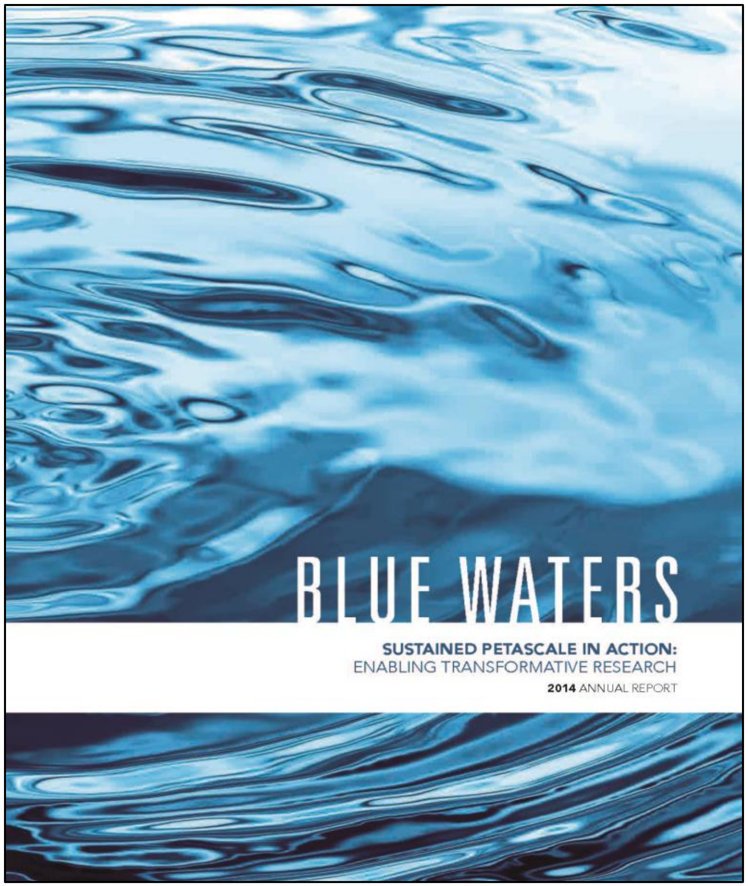
- The contribution of high end computing to U.S. leadership and competitiveness in basic science and engineering, the roles that NSF can play in sustaining this leadership, and how leadership can be measured, benchmarked, and monitored.
- Expected future national-scale computing needs across the full 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.

# Recent NSF-Supported Computational Investments Reflect Increasing National Diversity



# Blue Waters: Gaining New Insights via Sustained Petascale Performance

***Enabling science in many disciplines and researchers from many institutions***




- **DiMatteo (CMU)** Modeling formation of the first quasars
- **Jordan (USC)** 3-D Physics-based earthquake forecasting models
- **Stein (MSU)** *Ab Initio* models of solar activity
- **Reed (Cornell) & Wood (Princeton)** Advancing space-based earth science using petascale design and management of satellite assets
- **Wuebbles (UIUC)** High-resolution climate simulations

# Blue Waters: > 120 projects and 700 users, at 50 institutions, in 35 states

**Distribution of annual usage per general discipline area across all science teams**

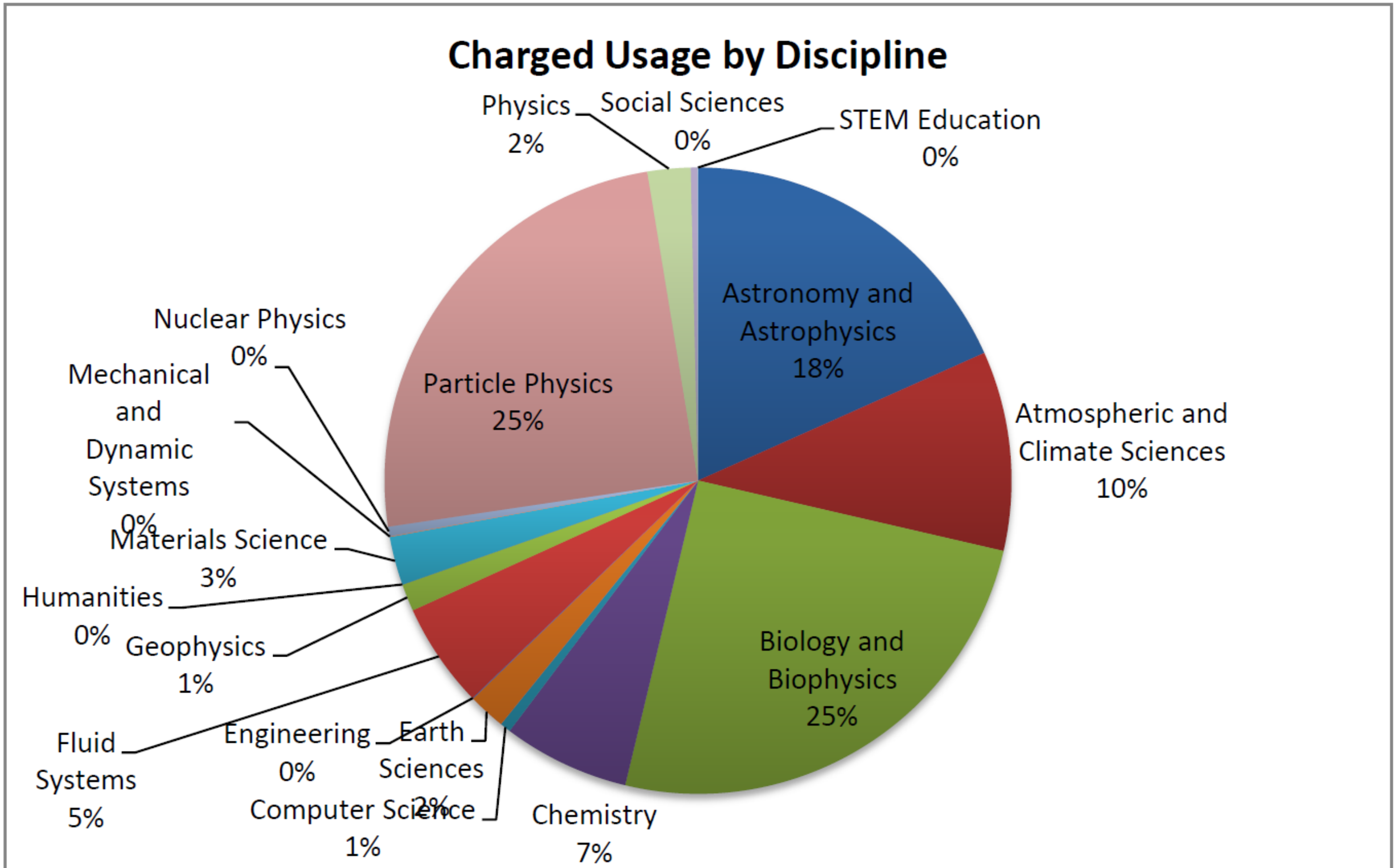



*Figure 15 Distribution of charged usage per general discipline area across all allocated projects*

# ACI supported resources are increasingly diverse yet collaborative


**UIUC/Blue Waters**
Leadership Class

**Open Science Grid** (OSG)
High throughput

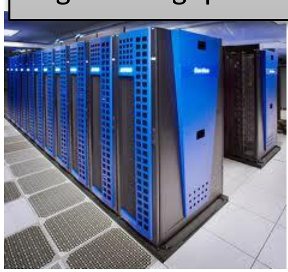
**Indiana U.**
❑ **Jetstream**
Cloud-based

**PSC**
Blacklight/Supercell
Large Shared Memory

❑ **Bridges**
Data intensive

Gordon
Data intensive
300 TB Flash Mem

❖ Comet
Long Tail Science
47k cores/2 PF
High Throughput



**UT- Austin**

Stampede
*102K cores*

*Maverick*
*Visualization, Data Analytics*

❖ *Wrangler*
*Novel Data Intensive*

**LSU SuperMIC**
380 nodes – 1PF

**UT-Knoxville/ORNL**
Darter
250 TFlops

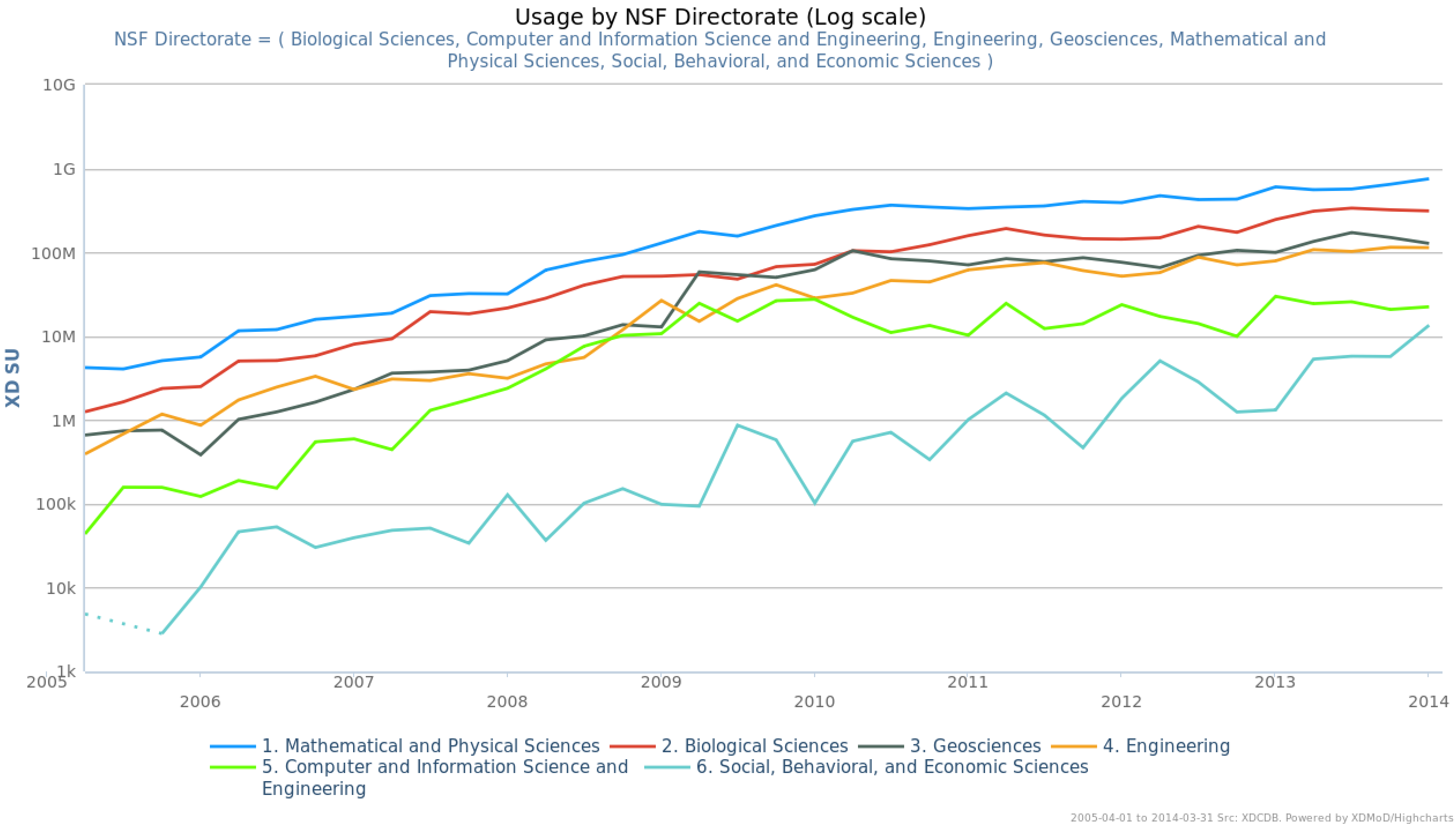
Nautilus
Visualization

Keeneland
CPU/GPGPU

❖ *To be deployed in 2015*
❑ *To be deployed in 2016*

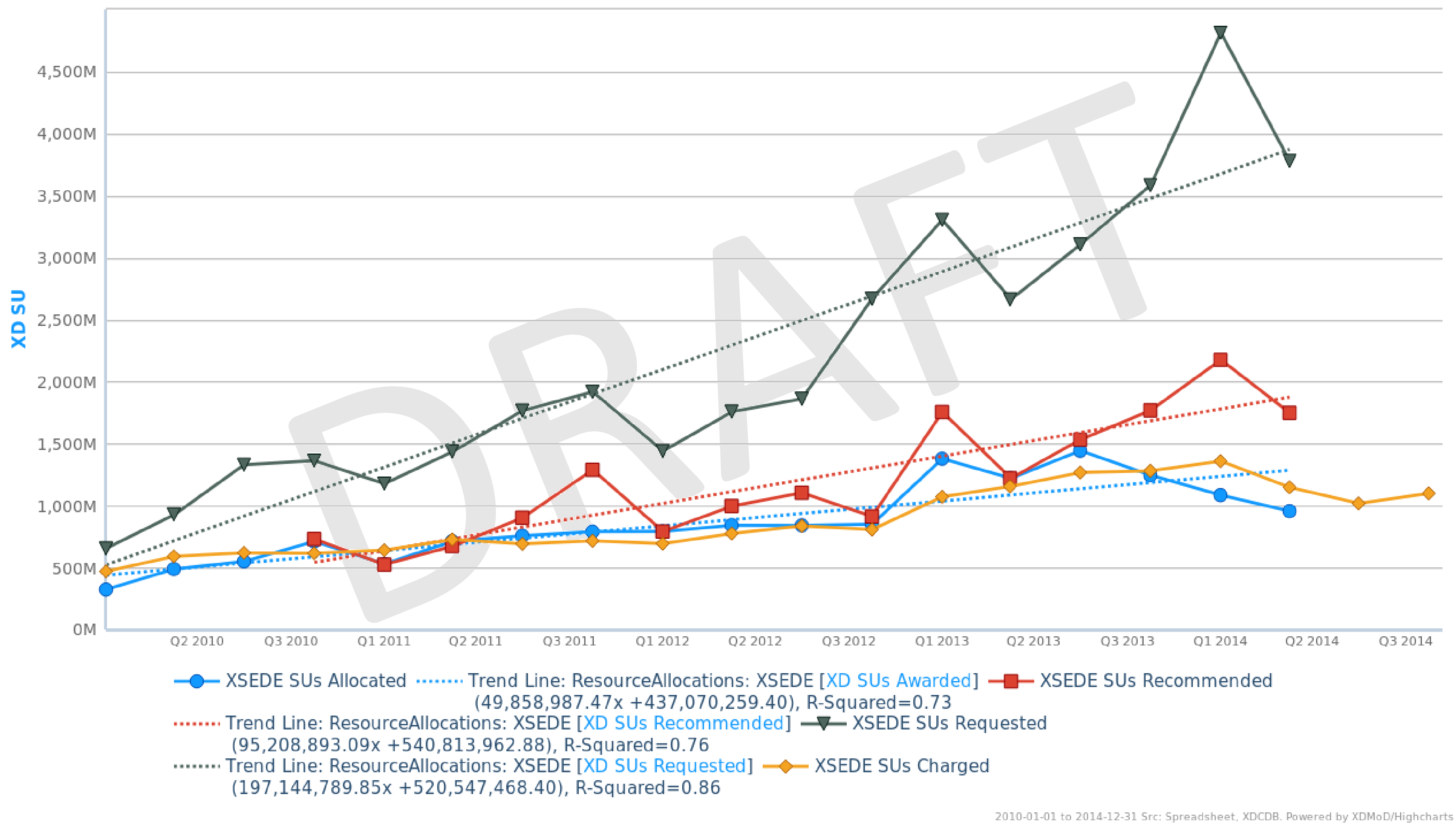
# TG-> XD: Growing Demand

***All disciplines are increasing use of national computational resources***



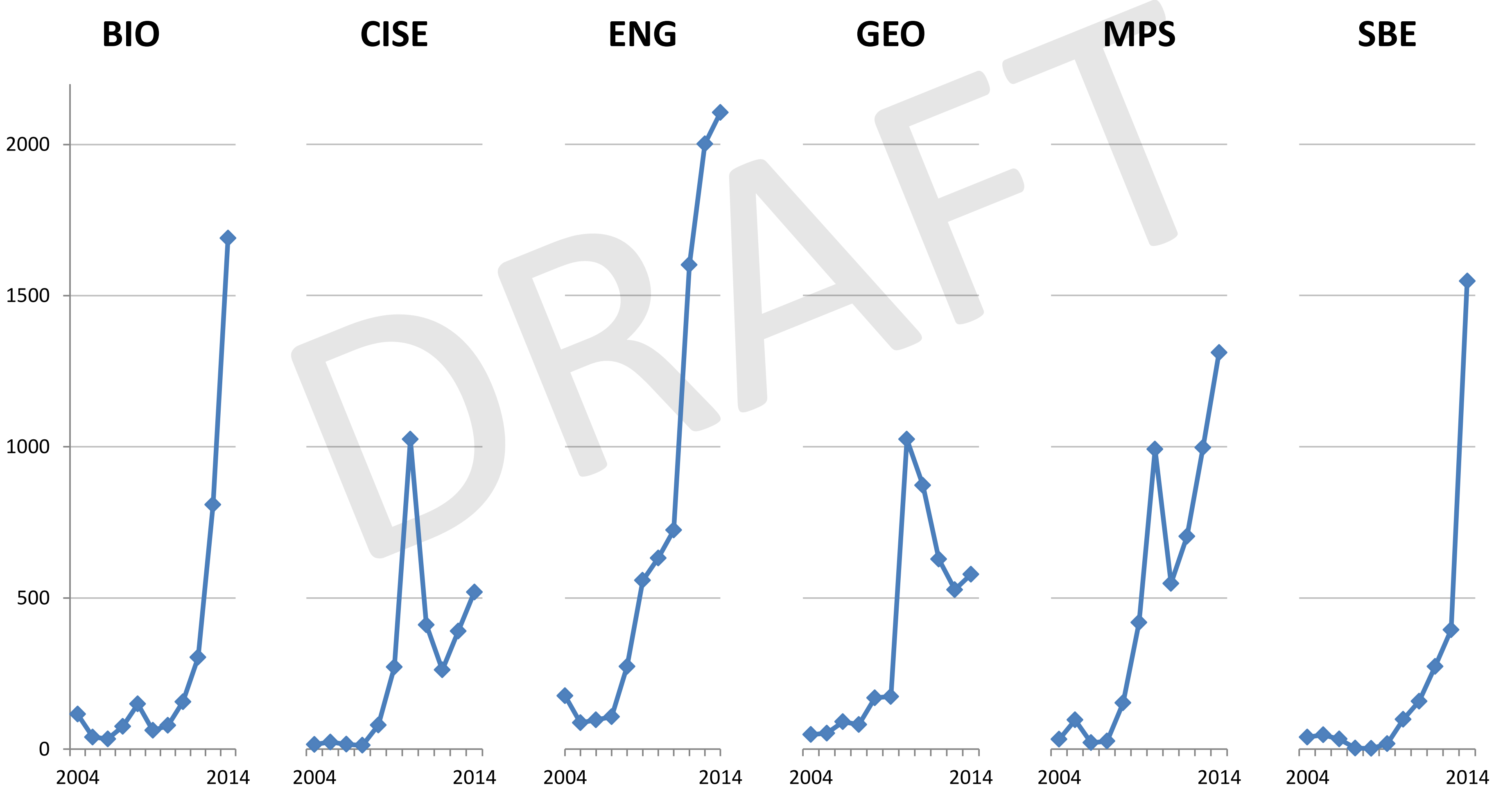
# XSEDE: Characterizing Trends

Consistent trend in requests vs. reviewer recommended allocations but recommendation are outpacing availability .



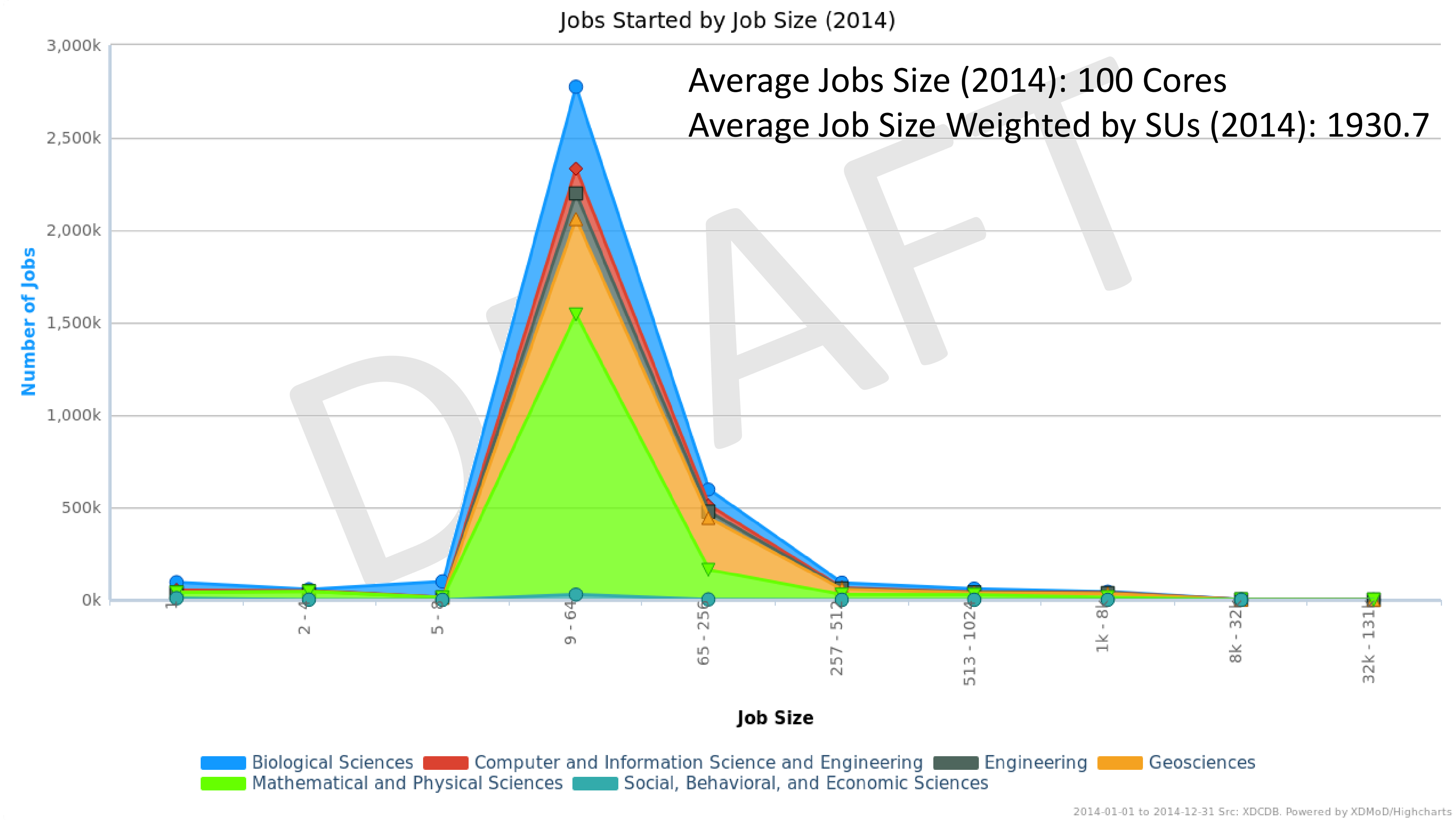
# Characterizing Workload

**XDMOD shows job sizes within XSEDE are increasing across all disciplines**

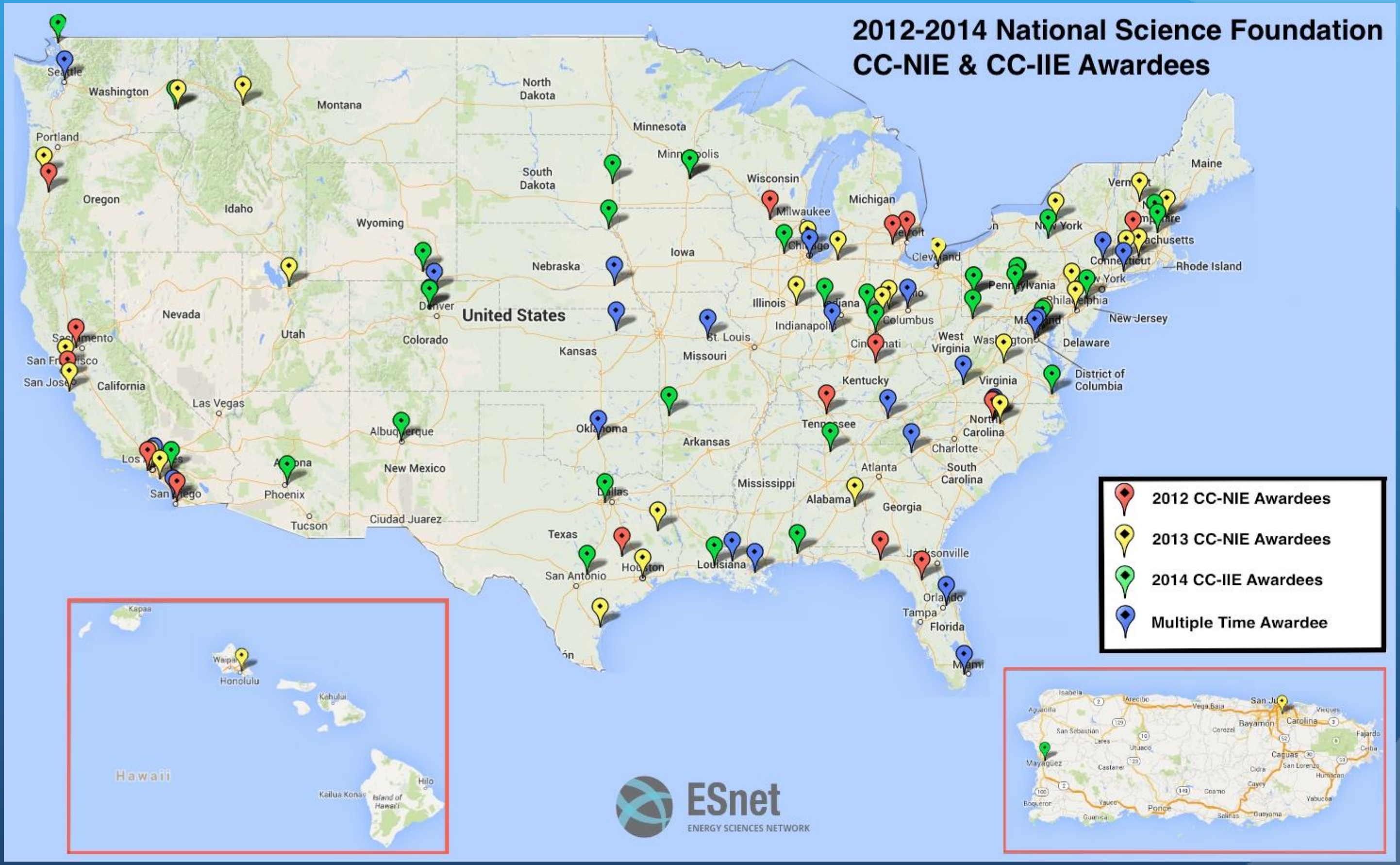


# XSEDE: Characterizing Workloads

2014 Distinctly different from Blue Waters workload , further examination of trends in process



# Stimulating Innovation on Campus
# CC*IIE Award Map 2012-2014



# Utah CC-NIE Integration project: *Science Slices* (NSF #ACI-1341034)

PI: S. Corbató; co-PIs: A. Bolton, T. Cheatham. R. Ricci, K. Van der Merwe;
SP: J. Breen, S. Torti

**Premise (Rob Ricci): What if we flipped the concept and built our Science DMZ on top of SDN infrastructure, rather than just plugging our SDN testbed into the DMZ?**

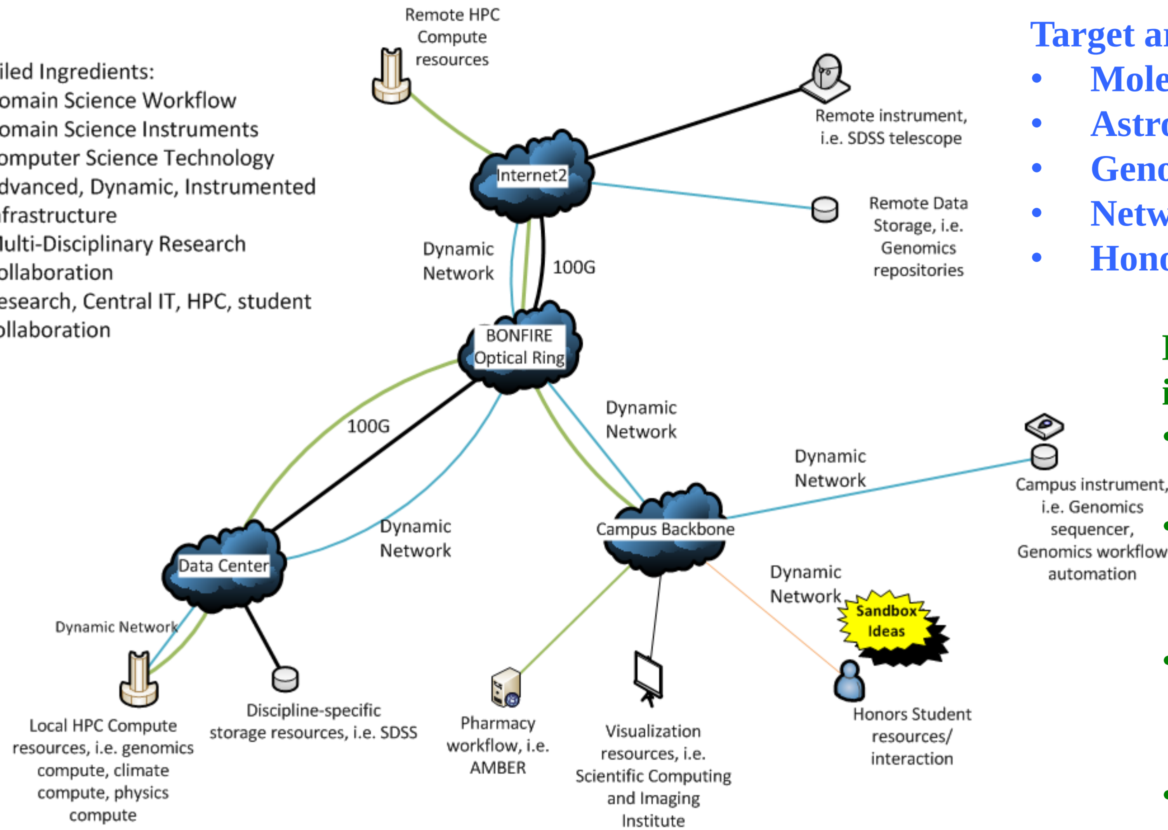
**1) Building a *dynamic* Science DMZ on top of an SDN-based framework (GENI)** [1]

**2) Extending slices to key campus labs, HPC center, and the Honors residential community**

**3) Working closely with central IT, campus IT Governance, and Utah Education Network**

Detailed Ingredients:

- Domain Science Workflow
- Domain Science Instruments
- Computer Science Technology
- Advanced, Dynamic, Instrumented Infrastructure
- Multi-Disciplinary Research Collaboration
- Research, Central IT, HPC, student collaboration




**Target areas:**

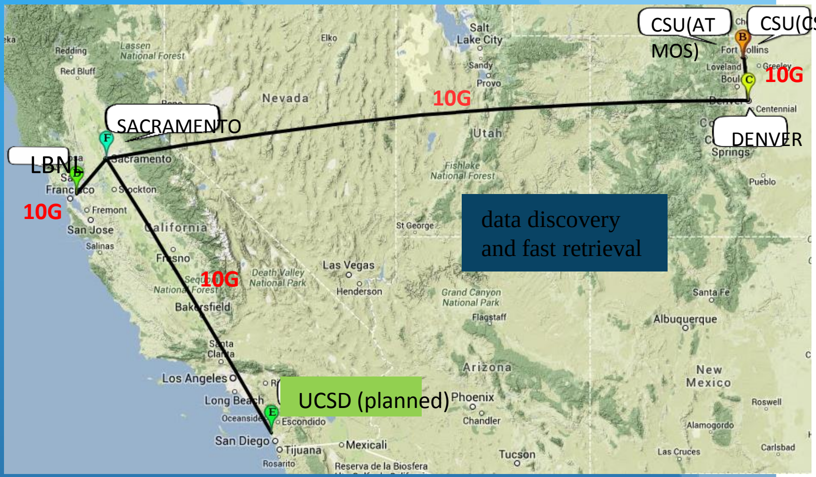
- **Molecular dynamics**
- **Astrophysics data**
- **Genomics**
- **Network/systems research**
- **Honors students**

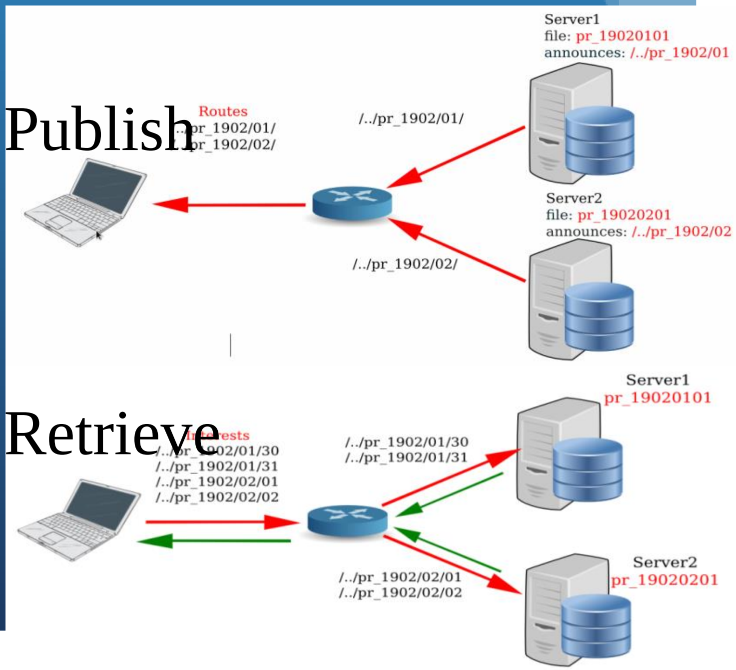
**Leverages new infrastructure:**

- **Downtown Data Center**
- **Utah Optical Network (BONFIRE)**
- **NSF MRI for novel cluster (Apt)**
- **Campus Net Upgrade**

# CC*IIE award #1340999 - Supporting Climate Applications over NDN (PI: Papadopoulos, CSU)

- **Need**: climate and other big data applications have overwhelmed existing networking and data management solutions
  - Data size and diversity
  - Naming, discovery, retrieval, sharing, etc.
- **Approach**: migrate workflows to NDN
  - Name based rather than host based paradigm
  - Easy migration: automatically translate existing ad-hoc names to structured NDN names
  - Evaluate over state-of-the-art NDN testbed deployed in partnership with ESnet
- **Benefit**: vastly simplified application and networking environment
  - Robustness and speed: in-network caching, efficient content distribution, automatic failover, security, etc.
  - Simplified management: highly structured, standardized naming across applicati... domai...
  - ... is...




Center for Multiscale Modeling of Atmospheric Processes CMMAP Reach for the sky.

Colorado State University

*#13410999: NSF CC-NIE Integration award*

# Optimizing NSF's Role in an Evolving Research Infrastructure Ecosystem

## Science Frontiers

data-intensive science;

software sustainability;

diverse priorities

## Technology Advances

end of Moore's law;

commoditization opportunities;

cohesive platform for simulation and data analytics

## Operating Models

efficient shared services;

local, regional, national models;

partnerships

## Human Considerations

workforce diversity;

career paths;

education;

community development

# Continuing Community Engagement

***Accelerating Science into the Future during a Period of Transition***

- **NSF Advanced Computing Infrastructure for 21st Century Science and Engineering: Vision and Strategic Plan (Feb 2012)**
  – Position, support spectrum of NSF-funded communities at cutting edge of advanced computing technologies, hardware, software, services
- **Future Directions of NSF Advanced Computational Infrastructure to Support US Science in 2017 – 2022**
  – National Academy of Sciences (NAS)
  – Interim Report (Oct 2014), Final Report (Summer 2015)
- **OSTP-led interagency strategic initiative**
  – Motivation: HPC essential to U.S. security, economic competitiveness, and scientific discovery


Interim report Co-chairs:
W. Gropp/UIUC
R. Harrison/Stony Brook