Advancing the Frontiers of Science through Cyberinfrastructure

Bill Miller
Science Advisor
CISE Division of Advanced Cyberinfrastructure

NSF Cybersecurity Summit, August 19, 2015
Other than the FY 2015 appropriation, numbers shown are based on FY 2014 activities.

$7.3 billion FY 2015 appropriation

94% funds research, education and related activities

48,100 proposals

11,000 awards funded

1,826 NSF-funded institutions

320,900 NSF-supported researchers

All S&E disciplines funded

Funds research into STEM education

214 Nobel Prize winners

NSF: Advancing Fundamental Science & Engineering (S&E) Research & Education
Cyberinfrastructure in NSF-Wide FY 2016 Budget Priorities

- Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21)
- Innovations at the Nexus of Food, Energy, and Water Systems (INFEWS)
- NSF Research Traineeship (NRT)
- Risk and Resilience
- Secure and Trustworthy Cyberspace (SaTC)
- Understanding the Brain (UtB)
- Urban Science

www.NSF.gov/about/budget/fy2016
Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21)

• Major effort across NSF to coordinate on CI.
• Embraces an expansive view of CI driven by research priorities and the scientific process.
• Includes investments in:
  – BIGDATA foundational research program
  – Data Infrastructure Building Blocks (DIBBs)
  – Software Infrastructure for Sustained Innovation (SI²)
  – Computational and Data-enabled Science and Engineering (CDS&E)
  – Data Science Pilots
NSF’s vision for cyberinfrastructure is informed by community input, development, and experience.
Mission: Support advanced cyberinfrastructure to accelerate discovery and innovation across all disciplines

Division Director: Irene Qualters
Division Assistant Director (Acting): Amy Friedlander

Science Advisor, Integrative Activities: Bill Miller

Data: Bob Chadduck, Amy Walton
High Performance Computing: Bob Chadduck, Rudi Eigenmann, Ed Walker
Networking & Cybersecurity: Anita Nikolich, Kevin Thompson
Software: Dan Katz, Rajiv Ramnath

Learning and Workforce Development: Sushil Prasad
ACI: Operational View

Supporting advance CI to accelerate discovery and innovation

Science Drivers
Constant exchange with NSF Directorates, Divisions and Programs

ACI investments
Convergent investments in technologies and communities to maximize impact

Leadership, Coordination, Partnership

Materials Genome
Understanding the Brain
INFEWS
Facilities

High Performance Computing
Networking and Cybersecurity

Data
Software

People (LWD)

NSF Directorates, NSF-Wide Initiatives, Industry, Agencies, International
Long Term Access to Large Scientific Data Sets: From SkyServer to SciServer

SciCloud

DATA
Sci 1 Sci 2 Sci 3 → Sci New

Large TB Scratch

Personal Store (MyDB)

100 Gb

XSEDE
Sci External

Sci 1 Sci 2 Sci 3 → Sci New

WEB Service API

Scheduler Registry Identity

WEB

User App User App School App Future App

User Citizen Educator Student Future

Scientist Citizen Scientist

Future Researcher

SCISERVER

CIF21 DIBBs: ACI-1261715
**NSF-supported Network of National HPC Resources & Services**

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

- **Gordon**
  - Data intensive
  - 64 TB memory
  - 300 TB Flash Mem

- **Comet (2015)**
  - “Long Tail Science”
  - 47k cores/2 PF
  - High throughput

- **Blue Waters**
  - Leadership Class

- **Yellowstone**
  - Geosciences

- **Open Science Grid**
  - High throughput

- **Blacklight**
  - Shared Memory
  - 4k Xeon cores

- **Bridges (2016)**
  - Large, coherent shared-memory

- **Darter**
  - 24k cores

- **Nautilus**
  - Visualization
  - Data Analytics

- **Jetstream (2016)**
  - Cloud-based

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

- **Maverick**
  - Visualization
  - Data Analytics

- **SuperMIC (2015)**
  - 380 nodes – 1PF (Ivy bridge, Xeon Phi, GPU)

**Coordination through XSEDE**
- Resource Allocation
- Advanced User Support
- Digital Services Architecture
ACI Networking Programs
A fundamental layer underpinning CI

- **Campus Cyberinfrastructure – Network Infrastructure and Engineering (CC-NIE/CC-IIE)**
  - Campus networking upgrade (re-design to science DMZ at campus border and 10/100Gbps) and innovation program. Joint with CISE/CNS

- **International R&E Network Connections (IRNC)**
  - Enable global scientific collaboration. Joint with NSF International Ofc.
  - Provide network to link U.S. research with peer networks globally
  - Stimulate the deployment and operational understanding of emerging network technology and standards in an international context
ACI Cyber Security Programs

Secure and Trustworthy Cyberspace (SaTC)

- Aligns with President’s *Strategic Plan for the Federal Cybersecurity R&D Program* (2011)
- Partners: CISE, SBE, EHR, ENG, and MPS
- Investments:
  - SaTC solicitation: Transition to Practice (TTP). Supports development, implementation, and deployment of applied security research into an operational environment.
  - NSF/Intel Partnership on Cyber-Physical Systems Security and Privacy (CPS-Security)
  - Education and training in cybersecurity

Cybersecurity Innovation for Cyberinfrastructure (CICI)

- Supports development/deployment of hardware and software technologies and techniques to protect research CI across every stage of the scientific workflow.
NSF Software Infrastructure Projects

5 rounds of funding, 65 SSEs

4 rounds of funding, 35 SSIs

2 rounds of funding, 14 S2I2 conceptualizations


SSE & SSI – NSF 14-520: Cross-NSF, all Directorates participating
Next SSEs due Feb 2015; Next SSIs due June 2015
Research Data Alliance

Building a Global Research Data Community

>2700 Members from 95 countries

Initial Delivery of Products

- A basic vocabulary of foundational terminology and query tools.

- A data type model and registry (“MIME-types” for data) to help tools interpret, display, and process data.

- A persistent identifier type registry to help search engines understand what they are pointing to and retrieving.

- A basic set of machine actionable rules to enhance trust
Create a coordinated Federal strategy in High Performance Computing research, development, and deployment to maximize the benefits of HPC for economic competitiveness and scientific discovery.

**Strategic Objectives**

1. Accelerate delivery of a capable exascale computing system to deliver approximately 100X performance of current 10PF systems.

2. Increase coherence between technology base used for modeling and simulation and that used for data analytic computing.

3. Establish, over the next 15 years, a viable path forward for future HPC systems in the post Moore’s Law ...

4. Increase capacity and capability of an enduring national HPC ecosystem. Use a holistic approach ... networking, workflow, downward scaling, foundational algorithms and software, workforce development.

5. Develop enduring public-private partnerships
CI trends and challenges

- Very dynamic environment – sensors, software, & data management/sharing tools are becoming ubiquitous.
- Researchers are more CI-aware: engaging many CI resources to *integrate data* and make discoveries.
- Federal policies encouraging open access to publications and research data, collaboration, sharing.
- Large scale efforts to develop shared CI resources and standards across fields – e.g. EarthCube, iPlant, ...

- What is “Data Science?”
- Demand for HPC resources for Big Data & Big Models
- Sustainability: workforce, software, hardware
- Who are the data users? Identity or identities?
Increasing Public Access to Research Results
NSF Plan released March 18, 2015

• **Requires** deposit of journal articles and juried conference papers in the NSF Public Access Repository (NSF-PAR), hosted by DOE/OSTI, within 12 months following initial publication, effective January 2016.
  
  – Allows for a **waiver** to the 12-month embargo for publications.

• **Retains** current Data Management Plan requirements and calls for community engagement to create more consistent management of research data.

• **Retains** current policies permitting costs of publication and the sharing of research results as a direct cost in the proposal budget.

CI Challenge: User-Centric Viewpoint

Revolution in the scientific workflow: many interfaces to shared services

Large Facilities

Shared Data/Software Gateway Resources

Collaboration Networks

Researcher

National Computing Resources

Cloud Services

Identities? Resources? Persistence?
NSF Sponsored Community Workshops on Facilities and Cyberinfrastructure *

First workshop is being planned for early December 2015

Stay Tuned!

*These workshops will not cover Cybersecurity, which is addressed via the Cybersecurity Summit
Thanks!
Innovations at the Nexus of Food, Energy, and Water Systems (INFEWS)

Securing and protecting food, energy and water resources

- Includes investments in:
  - New resource management algorithms, architectures
  - Real-time coordination, communications
  - Robust observation, sensing, inference
  - Large-scale data analysis/management, including modeling, simulation
  - Optimization of complex systems
  - Advancing computational infrastructure
- NSF-wide participation