Advancing the Frontiers of Science through Cyberinfrastructure

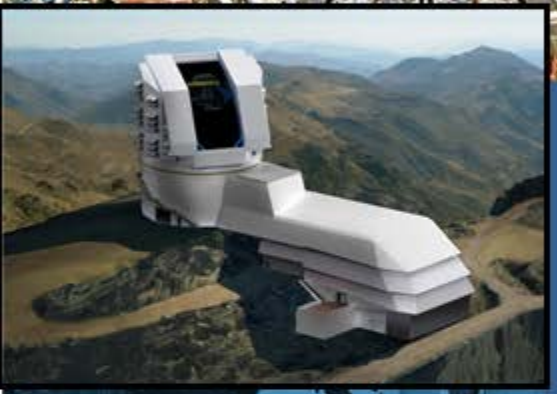
Bill Miller

Science Advisor
CISE Division of Advanced Cyberinfrastructure

NSF Cybersecurity Summit, August 19, 2015

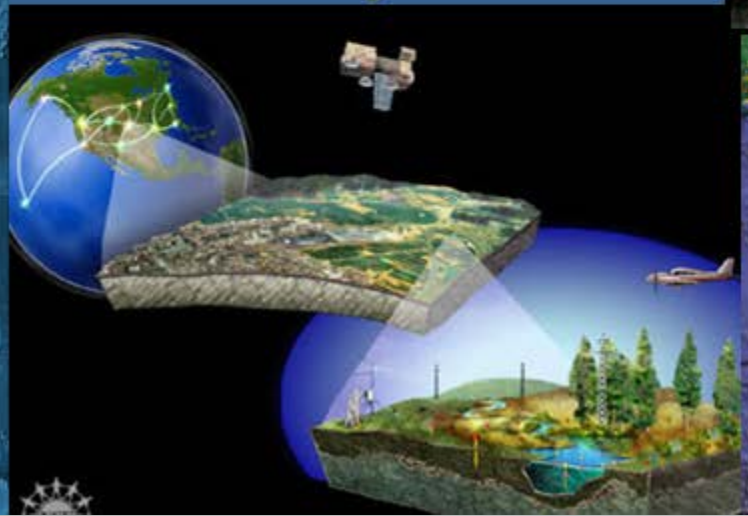
NSF: Advancing Fundamental Science & Engineering (S&E) Research & Education

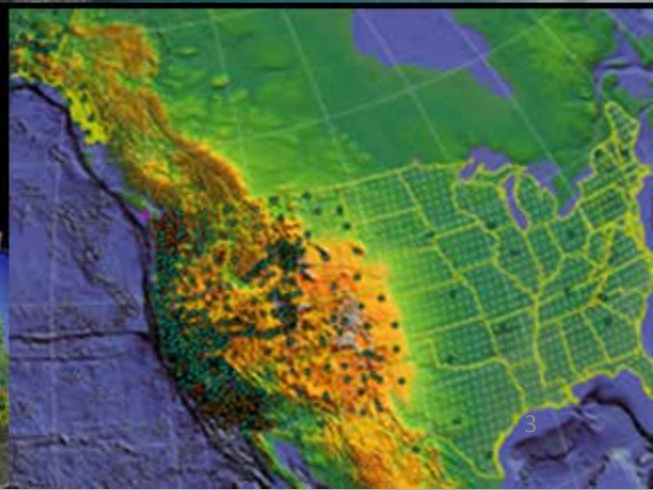
Other than the FY 2015 appropriation, numbers shown are based on FY 2014 activities.



NSF


IRMINGER SEA
PIONEER ARRAY
REGIONAL SCALE NODES
ENDURANCE ARRAY
ARGENTINE BASIN
SOUTHERN OCEAN





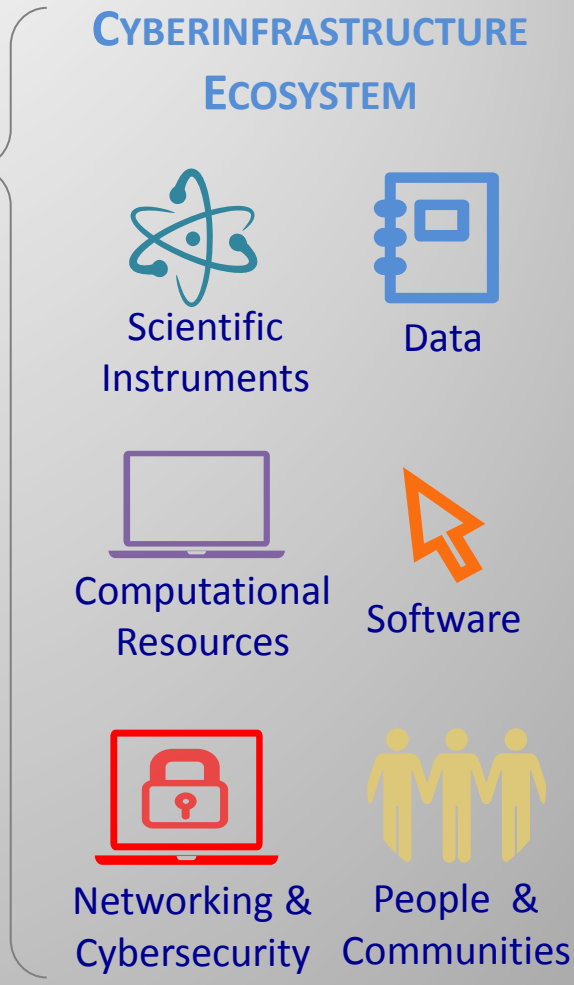
Cyberinfrastructure in NSF-Wide FY 2016 Budget Priorities

- 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

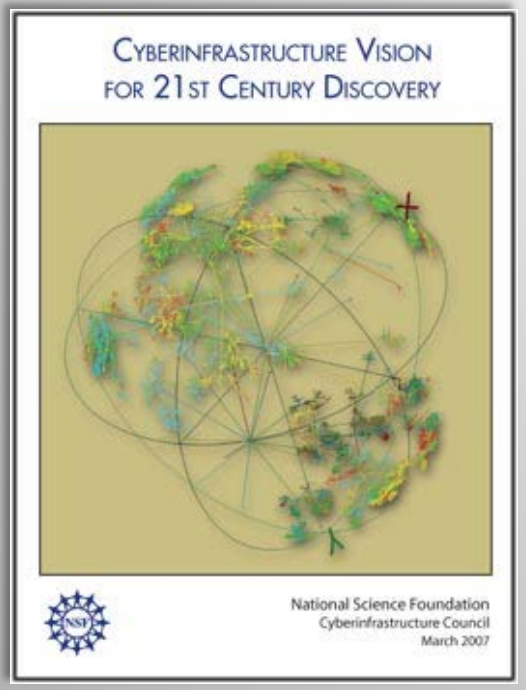
Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21)

- 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 (SI^2)
  - 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

Initial Vision (2007-2010)



NSF-Wide Task Force Reports (2009-2011)

National Academies Study (On going)

Interim Report, Oct 2014
Final Report expected Fall 2015

NSF Directorate for Computer and Information Science and Engineering (CISE)

Division of Advanced Cyberinfrastructure (ACI)

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

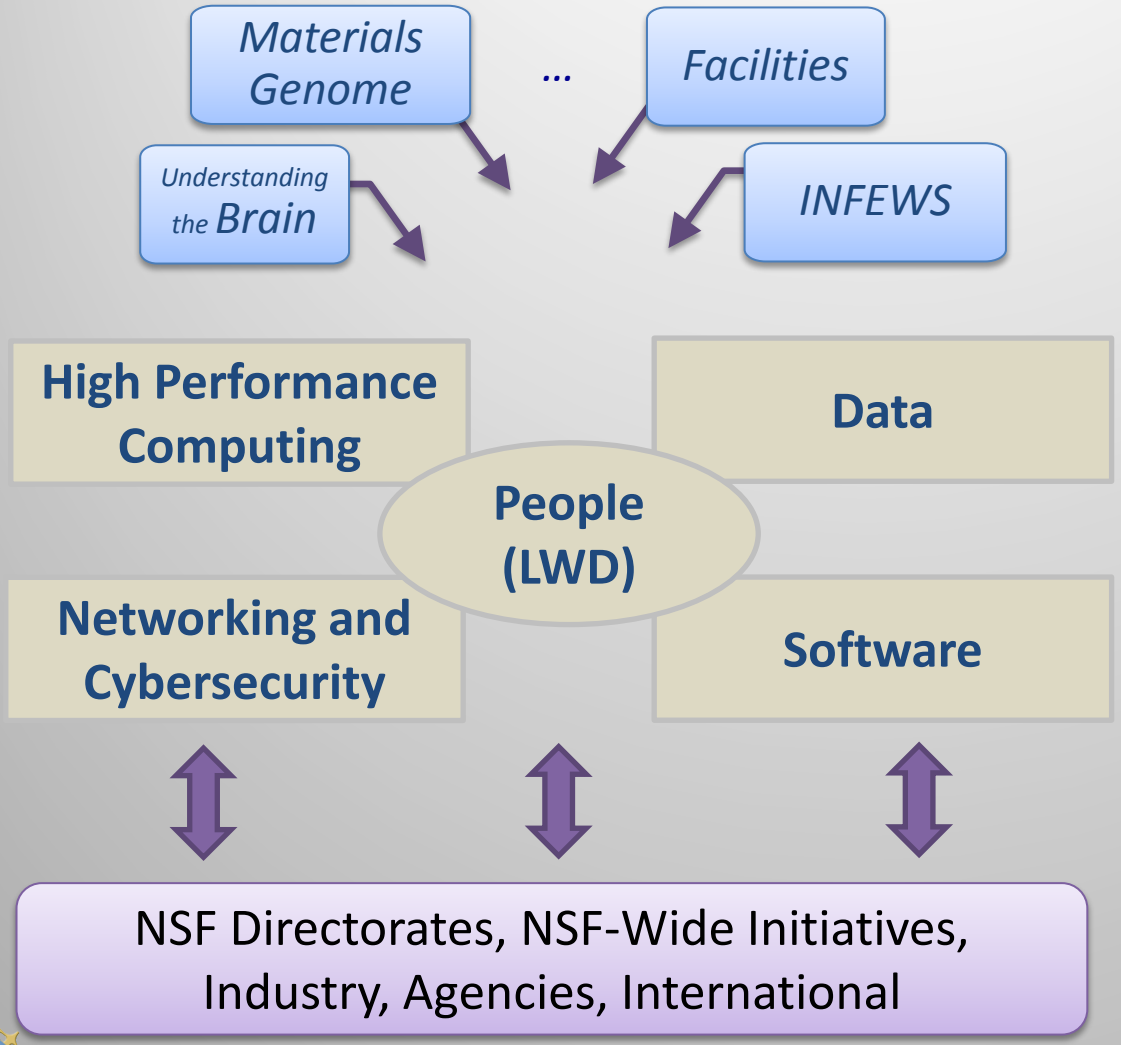
Data	High Performance Computing	Networking & Cybersecurity	Software
Bob Chadduck Amy Walton	Bob Chadduck Rudi Eigenmann Ed Walker	Anita Nikolich Kevin Thompson	Dan Katz Rajiv Ramnath

Learning and Workforce Development

Sushil Prasad

ACI: Operational View

Supporting advance CI to accelerate discovery and innovation



Science Drivers

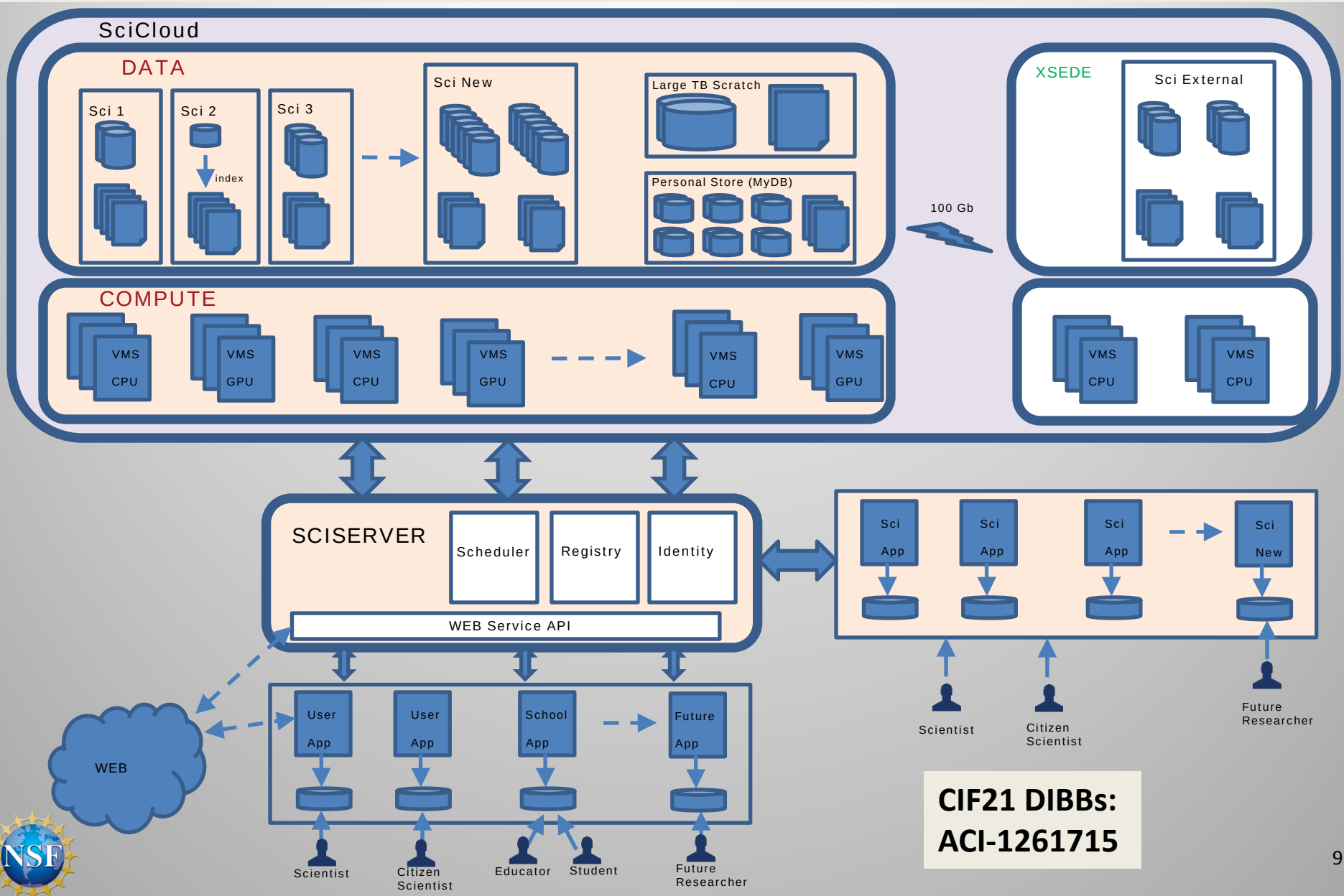
Constant exchange with NSF Directorates, Divisions and Programs

ACI investments

Convergent investments in technologies and communities to maximize impact

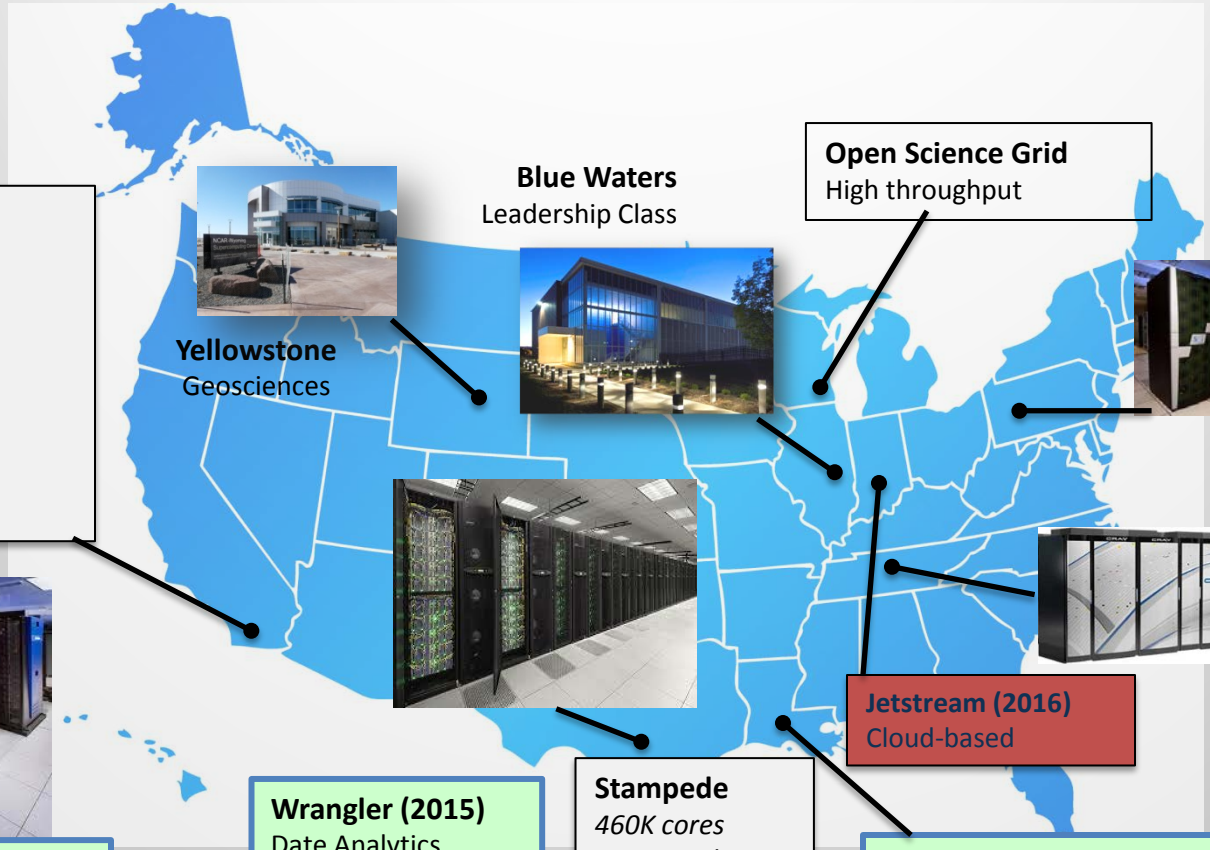
Leadership, Coordination, Partnership

Long Term Access to Large Scientific Data Sets: From SkyServer to *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

Yellowstone
Geosciences

Blue Waters
Leadership Class

Wrangler (2015)
Date Analytics

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

Maverick
Visualization
Data Analytics

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

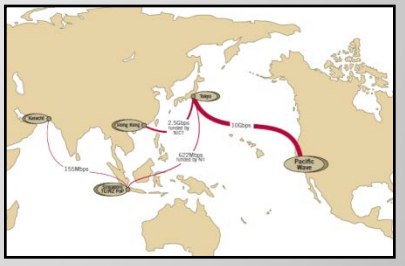
SuperMIC (2015)
380 nodes – 1PF
(Ivy bridge, Xeon Phi, GPU)

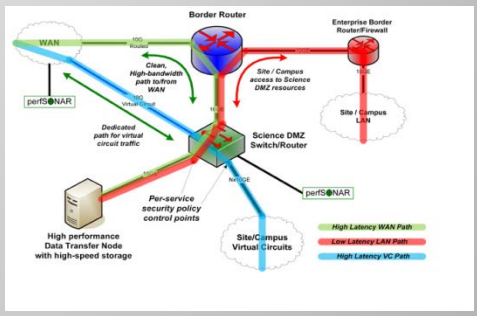
Coordination though **XSEDE**
- Resource Allocation
- Advanced User Support
- Digital Services Architecture

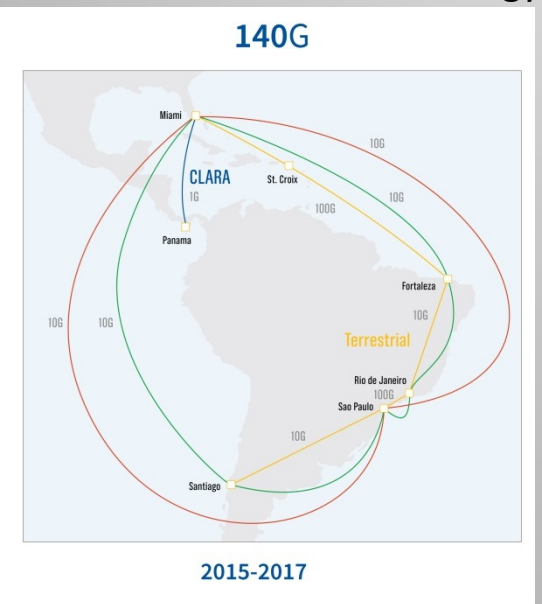
ACI Networking Programs

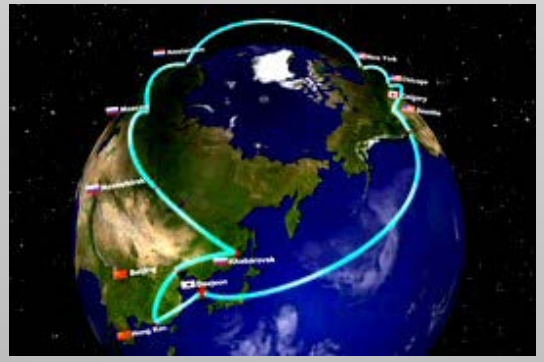
A fundamental layer underpinning CI

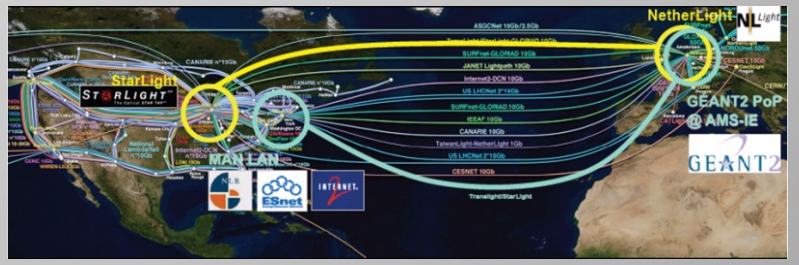
- Campus Cyberinfrastructure – Network Infrastructure and Engineering (CC-NIE/CC-IIE)
  - Campus networking upgrade (re-design to science DMZ at campus border and 10/100Gbps) and innovation program. Joint with CISE/CNS
- International R&E Network Connections (IRNC)
  - Enable global scientific collaboration. 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

Image Credit: *ThinkStock*

Image Credit: *ThinkStock*

Cybersecurity Innovation for Cyberinfrastructure (CICI)

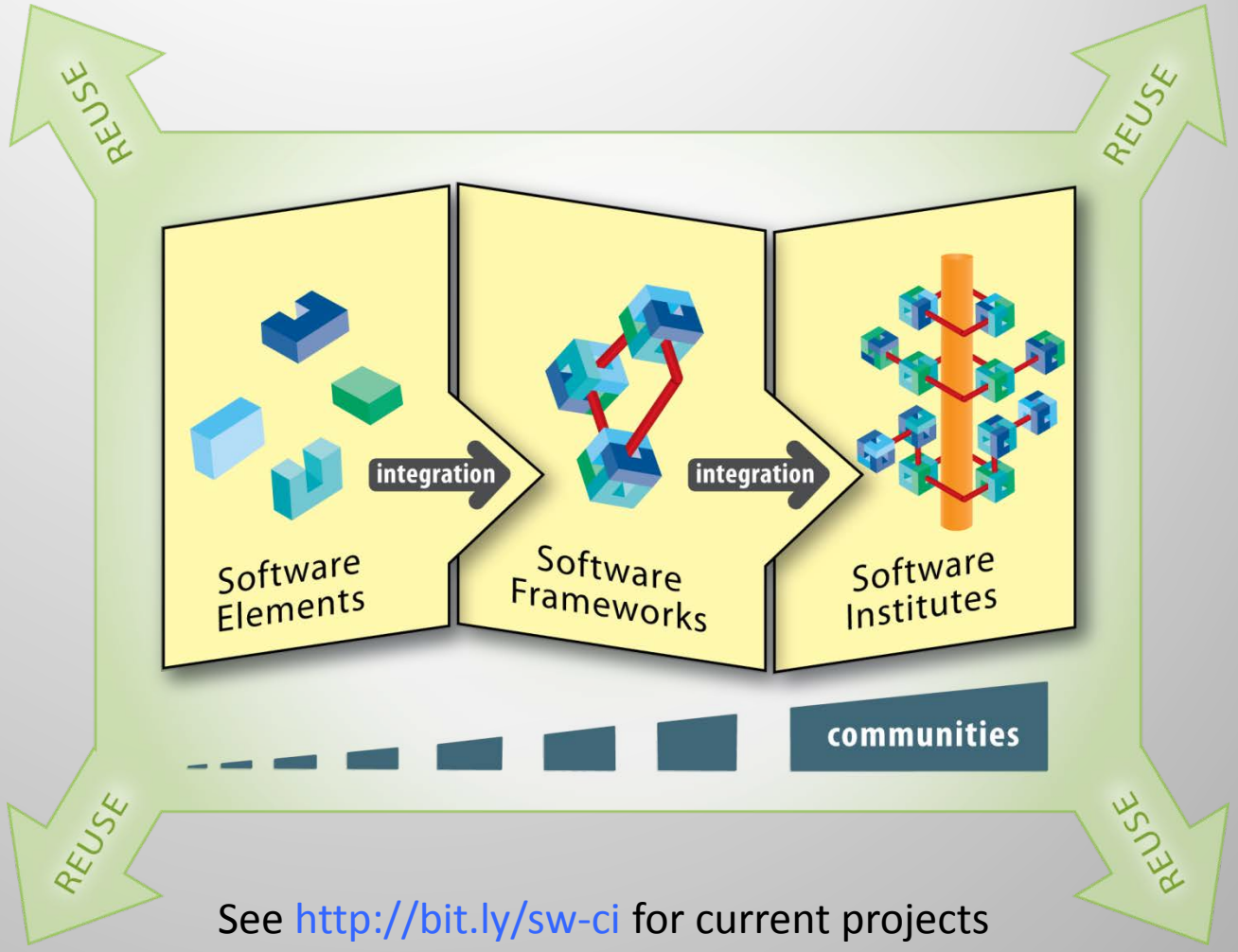
- Supports development/deployment of hardware and software technologies and techniques to protect research CI across every stage of the scientific workflow.

NSF Software Infrastructure Projects

5 rounds of funding,
65 SSEs

4 rounds of funding,
35 SSIs

2 rounds of funding,
14 S2I2
conceptualizations

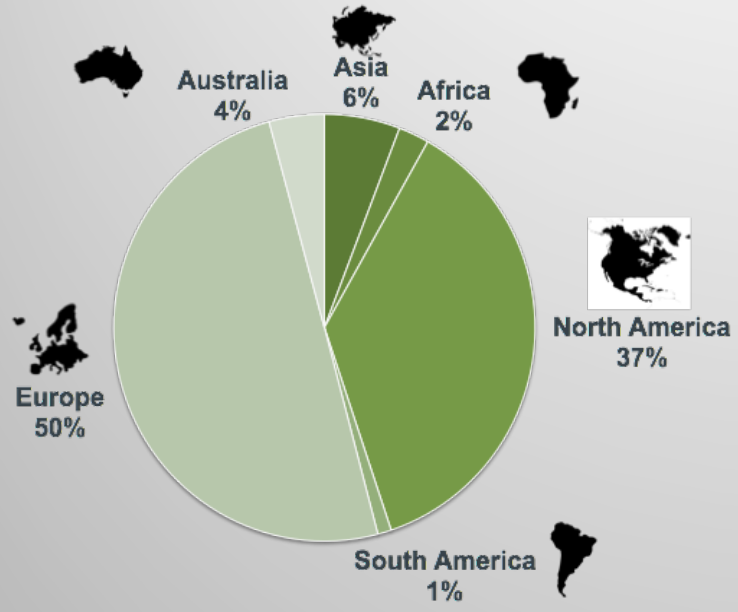


See http://bit.ly/sw-ci for current projects

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

National Strategic Computing Initiative (NSCI)

Executive Order, July 29 2015 **Lead Agencies: DOD, DOE, NSF**

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.

NSF foci

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

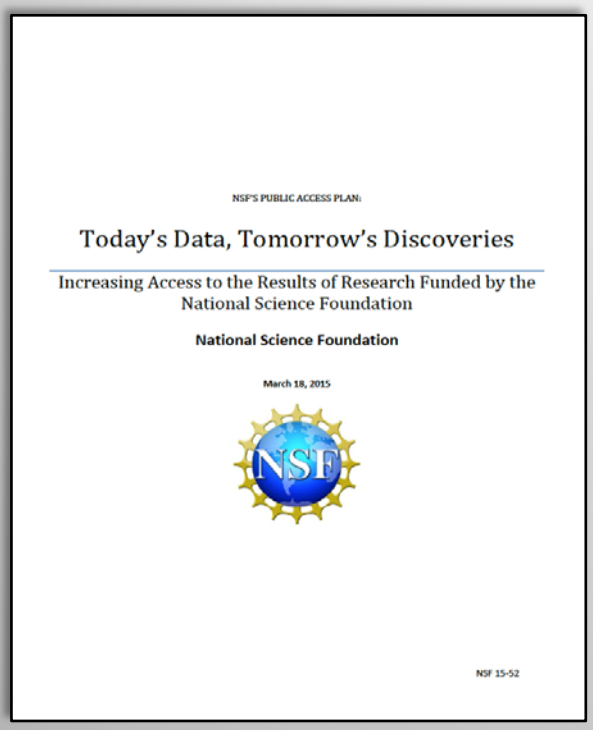
CI trends and challenges

- 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



NSF'S PUBLIC ACCESS PLAN:

Today's Data, Tomorrow's Discoveries

Increasing Access to the Results of Research Funded by the National Science Foundation

National Science Foundation

March 18, 2015

NSF 15-52

- ***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.

http://www.nsf.gov/news/special_reports/public_access/index.jsp

CI Challenge: User-Centric Viewpoint

Revolution in the scientific workflow: many interfaces to shared services

Large Facilities

Shared Data/Software Gateway Resources

Collaboration Networks

National Computing Resources

Researcher

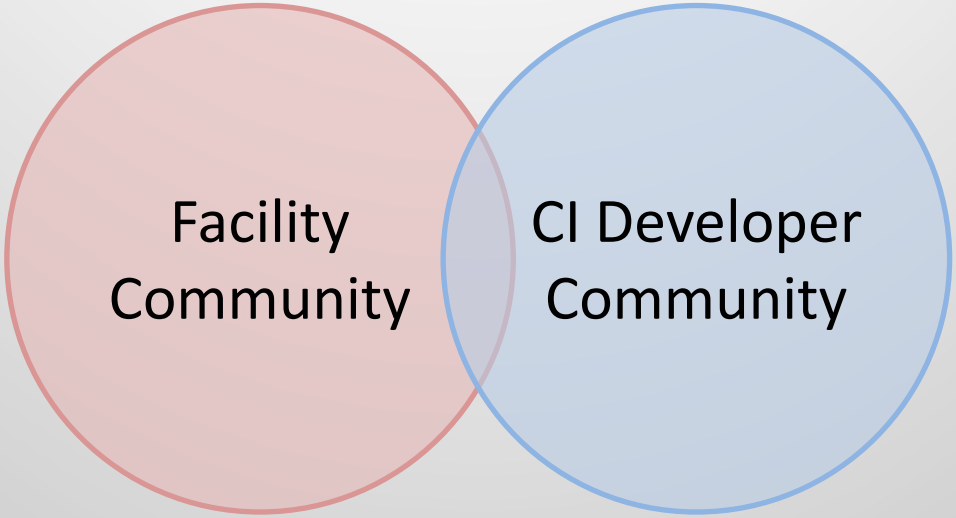
Software

Data

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