Cyberinfrastructure Investments and Opportunities

An update from the NSF Office of Advanced Cyberinfrastructure

Bill Miller

Science Advisor, NSF/OAC

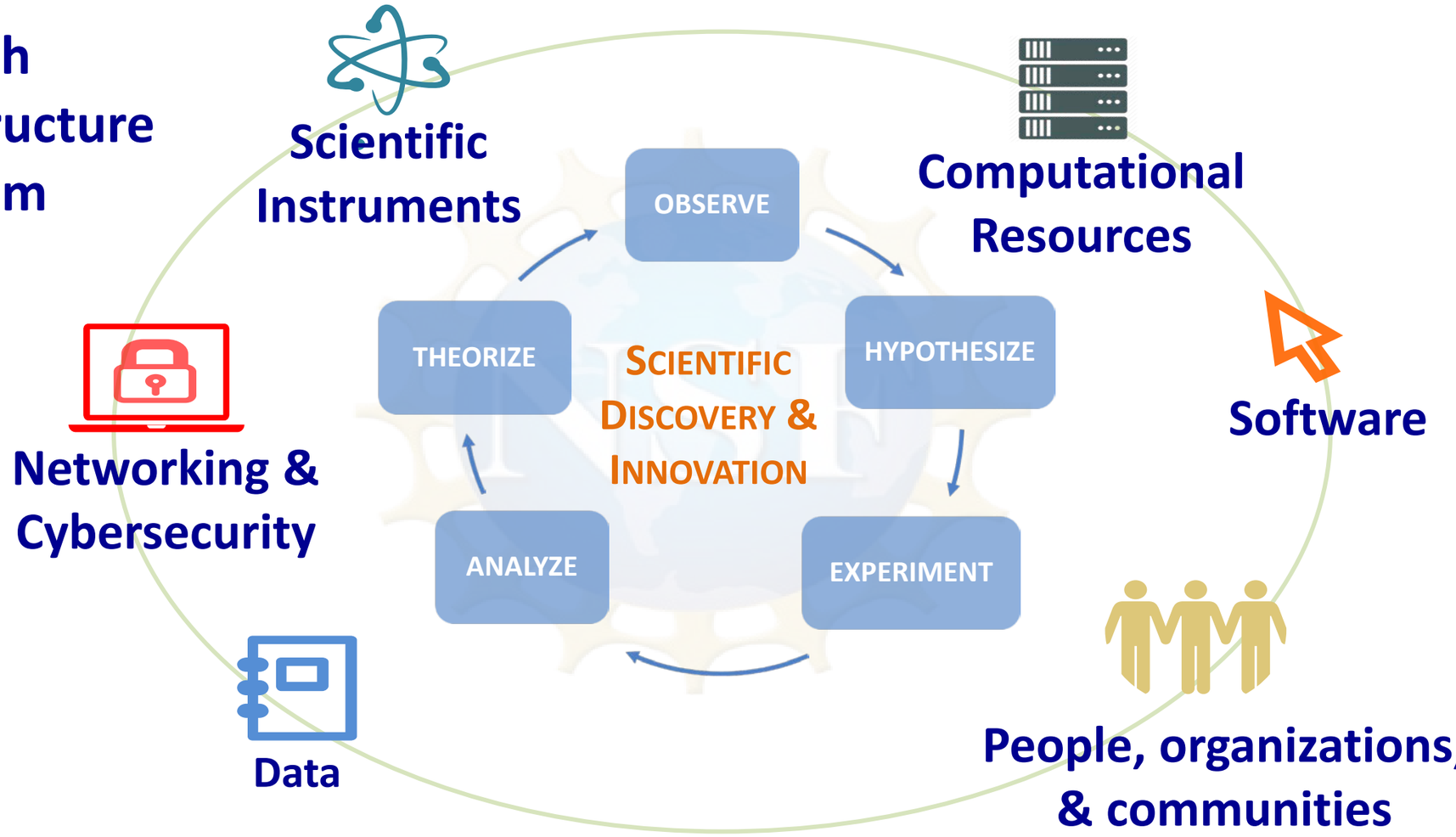
May 1, 2017

Outline

- Overview of Cyberinfrastructure
- OAC update and programs
- Looking into the Future
- Facility CI focus, upcoming workshop

NSF embraces an expansive view of Cyberinfrastructure driven by *evolving research priorities and scientific process*

Research Cyberinfrastructure Ecosystem



Drivers

- Ubiquity & scale of computational & data-intensive science.
- Ubiquity of sensors, cloud, technologies, social networks…
- Major initiatives and facilities…

NSF “Big Ideas” – all have CI implications

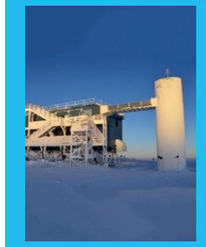
RESEARCH IDEAS

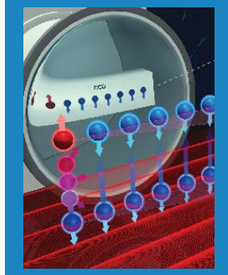
Harnessing Data for 21st Century Science and Engineering

Work at the Human-Technology Frontier: Shaping the Future

Navigating the New Arctic

Windows on the Universe: The Era of Multi-messenger Astrophysics



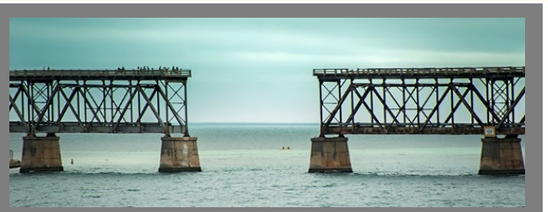


The Quantum Leap: Leading the Next Quantum Revolution

Understanding the Rules of Life: Predicting Phenotype

PROCESS IDEAS

Mid-scale Research Infrastructure

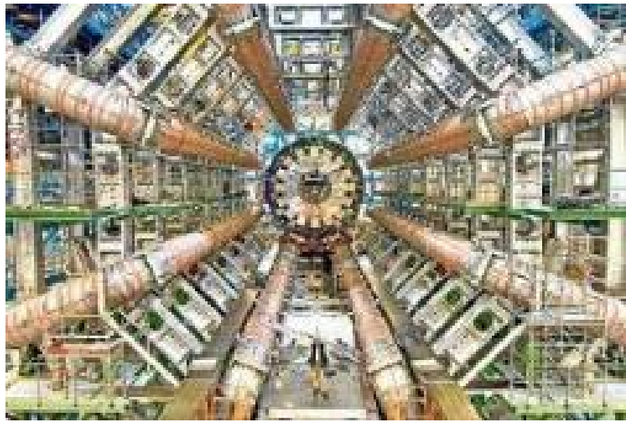


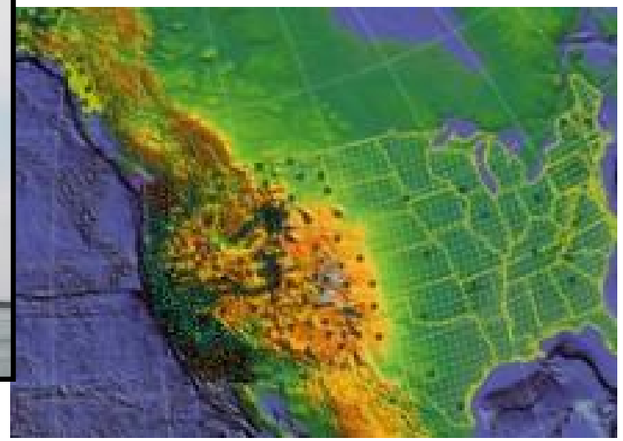
NSF 2050: Seeding Innovation

Growing Convergent Research at NSF

NSF-INCLUDES: Enhancing Science and Engineering through Diversity

NSF Facilities are Increasingly CI Driven … and dependent on Shared CI

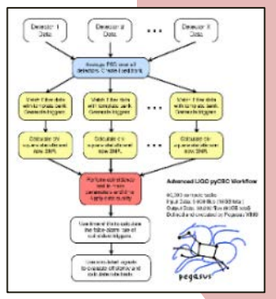




LIGO is enabled by NSF investments in instrumentation, computational science, cyberinfrastructure, and expert services

einstein toolkit

Pegasus



globus

Open Science Grid

XSEDE
Extreme Science and Engineering Discovery Environment

TEXAS ADVANCED COMPUTING CENTER
STAMPEDE

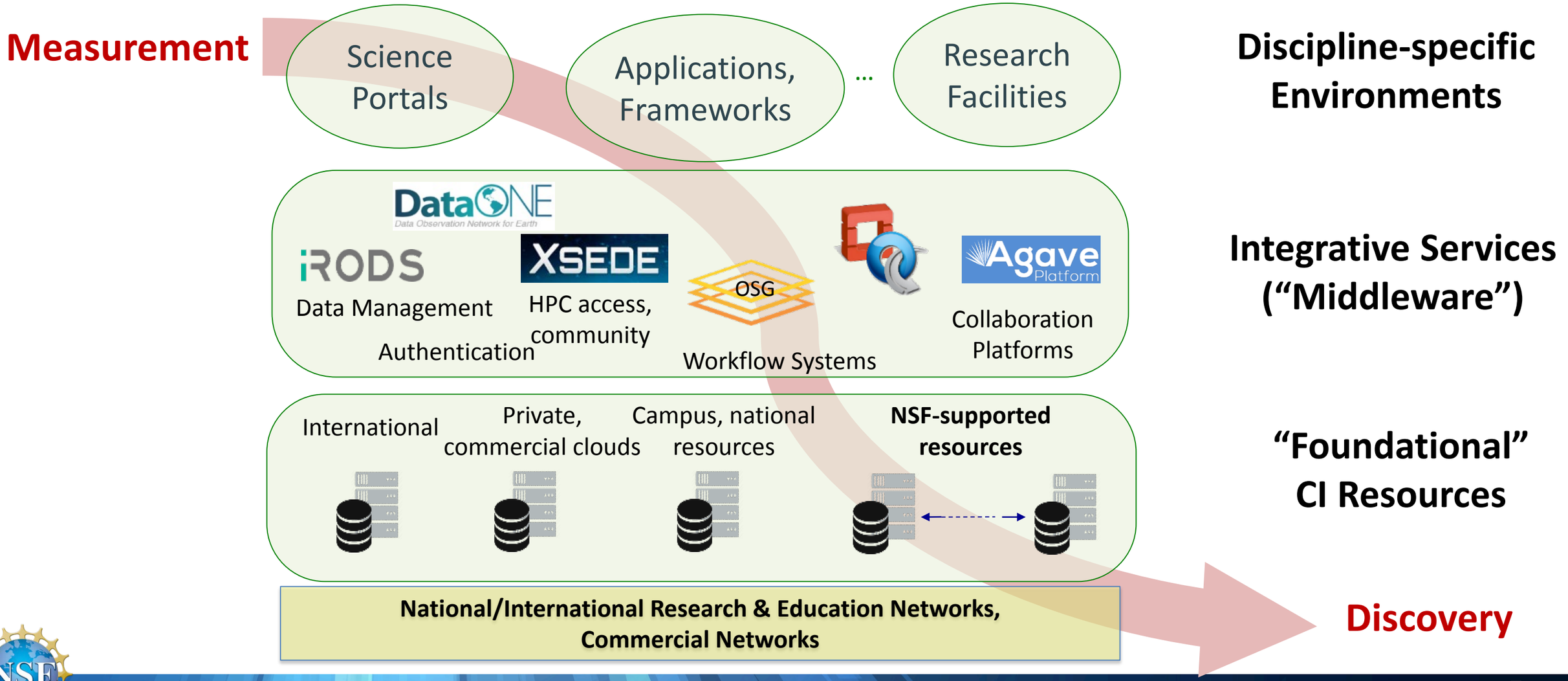
HTCondor
High Throughput Computing

- ✓ **Researcher access to sustained Advanced Computing resources**
  - New intensive simulations of relativity and magnetohydrodynamics. Massive, parallel event searches and validation (100,000 models).
  - Advanced computing resources and services: Open Science Grid (OSG); Blue Waters (UIUC); Comet (SDSC), Stampede (TACC), XSEDE allocations, AWS….
- ✓ **Interoperable Networking, Data Transfer, and Workflow Systems**
  - Pegasus, HTCondor, Globus workflow and data transfer management
  - NSF funded 100 Gbps upgrades enabled huge thoughput gains.
- ✓ **Software Infrastructure**
  - Computational science advances embodied in Software Infrastructure, for simulations, visualizations, workflows and data flows

Courtesy SXS.

NSF programs: Data Building Blocks (DIBBs), Software Infrastructure (SI2), Campus Cyberinfrastructure Network Infrastructure and Engineering (CC*NIE, DNI), and others. OSG and Pegasus are also supported by the US. DOE.

Emerging discovery pathways at scale: Architecture view



Outline

- Overview of Cyberinfrastructure
- OAC update and programs
- Looking into the Future: NSF CI 2030
- Facility CI focus, upcoming workshop

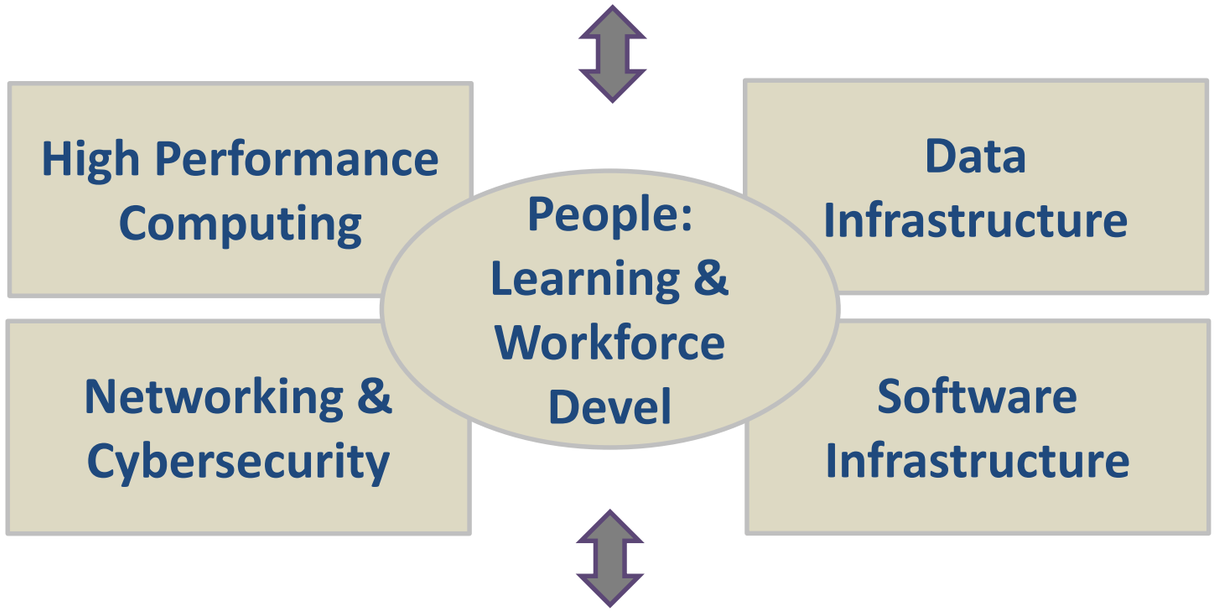
CISE Office of Advanced Cyberinfrastructure (OAC)

Mission: Accelerate discovery and innovation across all disciplines through advanced CI

Science Drivers

Domain science trends, agency priority areas, major facilities…

Investment areas



Leadership, Partnership

NSF Directorates, NSF and Federal Initiatives, Industry, International

Director: Irene Qualters
Deputy: Amy Friedlander

Science Advisor: Bill Miller

Coop Agreements: Al Suarez

High Performance Computing
Ed Walker, Bob Chadduck, Rudi Eigenmann

Data Infrastructure
Amy Walton, Bob Chadduck

Networking/Security
Anita Nikolich, Kevin Thompson

Software Infrastructure
Vipin Chaudhary, Rajiv Ramnath

Learning & Workforce Development
Sushil Prasad

www.nsf.gov/div/index.jsp?div=ACI

ACI Realignment → OAC

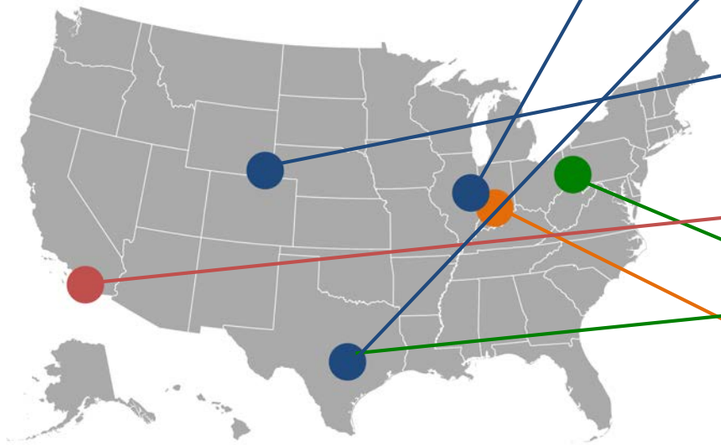
- 2013: *Office* of Cyberinfrastructure (OCI in OD) → *Division* of Advance Cyber Infrastructure (ACI in CISE).
- 2016 Review, including request for input from the community (DCL)
- Findings:
  - Well managed in CISE; budget has tracked CISE's
  - Leadership: visibility critical (external, internal)
- Outcomes:
  - Remain in CISE; renamed *Office* of Advanced Cyberinfrastructure (OAC), emphasizing its service to all S&E
  - Office Director participates in senior leadership meetings (SMaRT)
  - National searches for leadership (as for ADs)

High Performance Computing

NSF-supported National Computing Resources

Complements Larger Aggregate Investments from Universities and other Agencies

2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 ...



***Blue Waters*/UIUC**

Leadership HPC Planning

***Stampede*/UT Austin** | ***Stampede2*/UT Austin**

***Yellowstone*/NCAR-Wyoming** | **Cheyenne/NCAR**

Large-scale computation

Comet/UCSD

Long-tail and high-throughput

***Wrangler*/UT Austin**

Data-Intensive

***Bridges*/CMU/PSC**

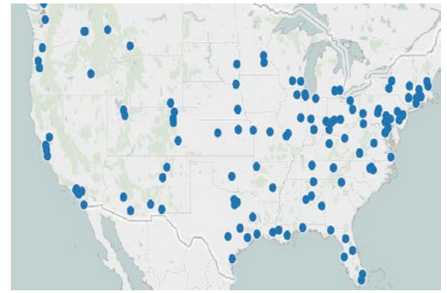
Jetstream/Indiana U.

Cloud

Networking & Cybersecurity

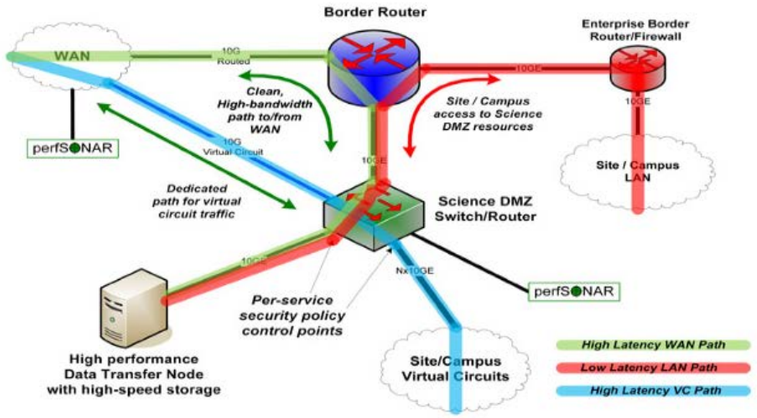
CISE/OAC Networking Programs

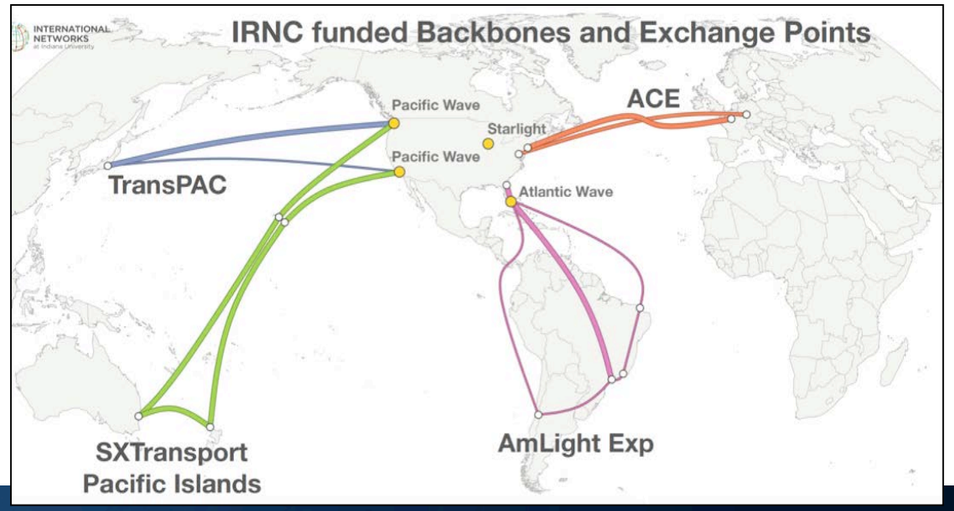
CC*



- **Fundamental layer** that enables scientific discovery at the institutional, regional and global collaborative levels.
- **Campus Cyberinfrastructure (CC*)**. Upgrading and accelerating campus networking (10/100Gbps). Re-designing campus border to Science DMZs. Innovation, + much more.
- **International R&E Network Connections (IRNC).** Link U.S. research with peer networks in other world regions. Supports all R&E US data flows (not just NSF-funded).
  - Stimulates deployment and operational understanding of emerging network technology, best practices, 100Gbps connections.

Science DMZ





Networking & Cybersecurity

OAC Cybersecurity

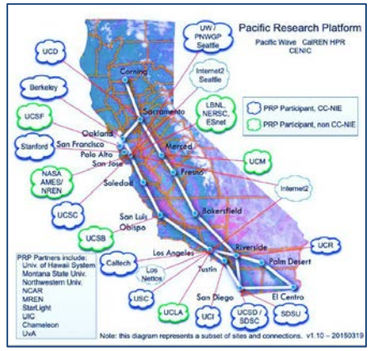
- **Cybersecurity Center of Excellence (CCoE)** – formerly Center for Trustworthy Scientific CI, www.trustedci.org.
  - Site reviews, code reviews, architecture reviews. Example engagements: Gemini, US Antarctic Program, LSST, OOI, LIGO, DKIST, NEON, Pegasus, PerfSonar, …
  - ➢ Open Science Cyber Risk Profile – asset/impact oriented approach for open science (DoE, NIH, NSF). Joint effort of CCoE & ESNet
- **Cybersecurity Innovation for Cyberinfrastructure (CICI)**. Topics: Secure and Resilient Architecture, Secure Data Provenance, Regional Cybersecurity
- **Secure and Trustworthy Cyberspace (SaTC).** – Cross-directorate program. OAC funds later stage/applied security projects that can secure scientific CI. Several "Transition to Practice" projects co-funded by Dept. of Homeland Security.
- **Annual Large Facilities Cybersecurity Summit**. ~120 attendees from NSF-funded science facilities. Next Summit: August 2017.

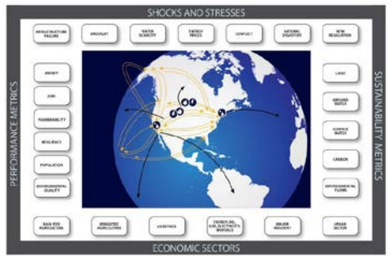
CTSC
CENTER FOR TRUSTWORTHY SCIENTIFIC CYBERINFRASTRUCTURE
The NSF Cybersecurity Center of Excellence

Example: Bro Intrusion Prevention/Detection software

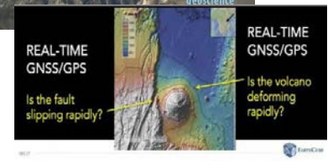
Data Infrastructure

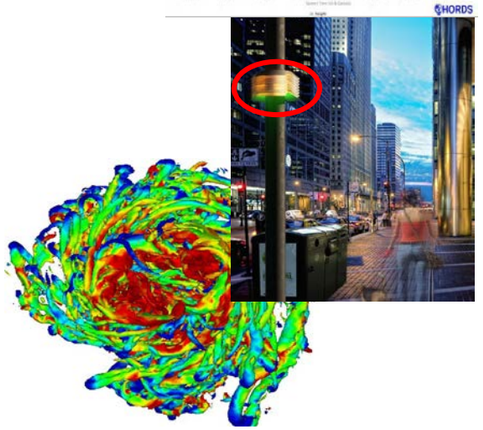
OAC Data Infrastructure: Accelerating Science, Building Community





- **Data Building Blocks (DIBBs).** Funds CI/discipline collaborations, cross-disciplinary infrastructure, built on recognized capabilities, tangible products.
  - First PI meeting, Jan 2017 on Results, Challenges, Future Directions, and Gaps to inform future investments.
- **CC* collaboration.** Example topics: multi-institution, cloud resources, sharing mechanisms.
- **EarthCube.** Collaboration with NSF GEO. Topics: Building new communities, innovative interoperable solutions that link and integrate resources, new capabilities for data capture, discovery, access, processing and analysis.
- **Innovations at the Nexus of Food, Energy and Water Systems (INFEWS)**. NSF cross-cutting activity.

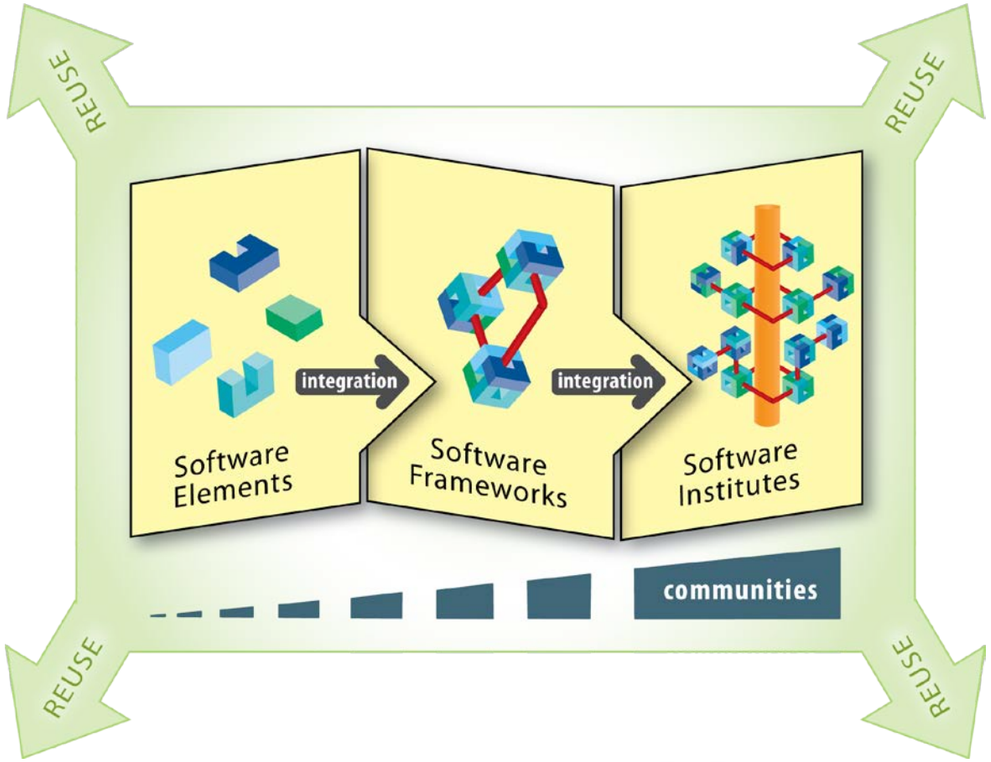


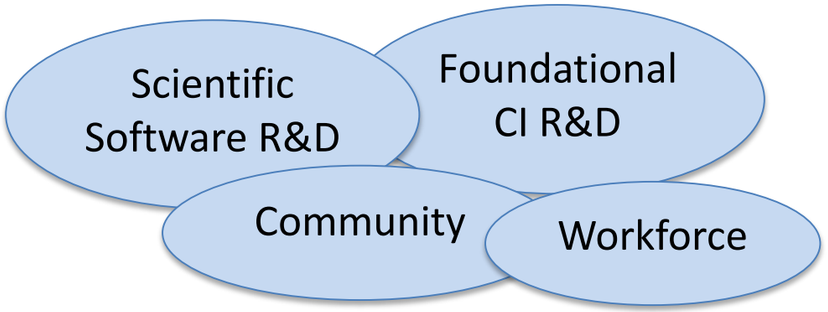


Software Infrastructure

OAC Software Investments

- **OAC Goal**: Catalyze and support unique, innovative software-intensive science ecosystems to advance research
- **Flagship - Software Infrastructure for Sustained Innovation (SI2)**. *Elements* ($500K/3 yrs), *Frameworks* ($1M/yr 3-5 yrs), *Institutes* ($3-$5m/yr 5-10 yrs).
- **Software "pipeline":**
  - R&D programs (SPX, CDS&E, DMREF, CRISP, Venture, …)
  - → Development and deployment (SI2)
  - → Outcomes: Sustainability, open source community, institutional support, education, SAAS, IP licensing, …





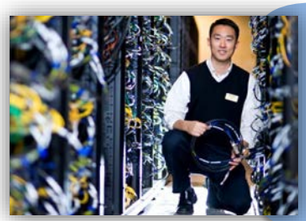
People

OAC Learning & Workforce Development

Communities of Concern

CI Contributors, Cyber-scientists

Develop new CI



CI Professionals

Deploy & support CI

CI Users

Exploit CI

New! **CyberTraining -** ***Training-based Workforce Development for Advanced Cyberinfrastructure*** (NSF 17-507)

- *Informal, scalable* training models and pilot activities - on topics in advanced CI, and computational and data-enabled science & engineering.
- OAC leads, with MPS, ENG, GEO, EHR/DGE, and CISE/CCF.
- $300K-500K over 1-3 years.
- **3 Tracks: 1**: CI Professionals. **2**: CI Contributors/Users in domain science and engineering. **3**: Undergraduate Computational & Data Science User Literacy.
- Excellent community response in the inaugural round.
- Next Deadline: *October 2017*

Outline

- Overview of Cyberinfrastructure
- OAC update and programs
- Looking into the Future
- Facility CI focus, upcoming workshop

  - ❖ National Academies Study: Future HPC
  - ❖ NSF CI 2030 survey: Science Drivers for Future CI
  - ❖ National Strategic Computing Initiative: Exascale RFI

Sunsetting in 2017

February 2012

Cyberinfrastructure for 21st Century Science and Engineering Vision and Strategic Plan

CIF21 fostered a rich NSF cyberinfrastructure ecosystem responsive to the *evolving discovery process.*

CI-Enabled Instrumentation

Computing Resources

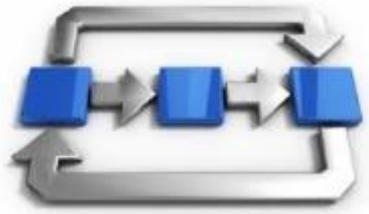
Data Infrastructure

Gateways, Hubs, and Services

Cloud Resources & Services

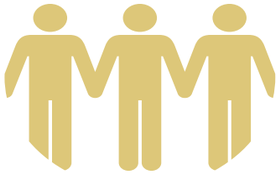
R&E Networks, Security Layers

Coordination & User support



Software and Workflow Systems

Pilots, Testbeds



People, organizations, and communities

Many parts! … Working together? … The right architectures? …
… Bottlenecks? … New pressures? …. Gaps?

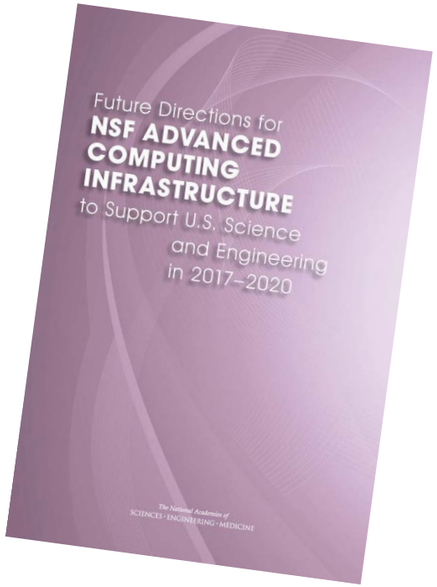
Community input is informing NSF's strategic planning refresh for advanced CI

- ✓ National Academies report on NSF Advanced Computing (2016):
  - *Future Directions of NSF Advanced Computational Infrastructure to Support US Science in 2017 – 2022*
- ✓ NSF RFI on Future Needs for Advanced Cyberinfrastructure to Support Science and Engineering Research (NSF CI 2030), (2017)
- ✓ NSF and Joint agency assessments of the NSCI Exascale RFI (2015):
  - *NSF Assessment of Responses to the Request for Information (RFI) on Science Drivers Requiring Capable Exascale High Performance Computing* *and*
  - *Joint Agency Assessment of the Responses to the RFI on Science Drivers Requiring Capable Exascale High Performance Computing*
- ✓ PI and Disciplinary Workshops, e.g.:
  - Data Building Blocks (DIBBs) 2017 PI Workshop Final Report,
  - Software Infrastructure 2017 PI Workshop
  - *2017 NSF Cyberinfrastructure for Facilities Workshop, Sept 6-7, 2017*
- ✓ Input from other NSF Advisory Committees and other bodies.

High Performance Computing

Community Analysis of Future HPC

- 2016 National Academies Study, *Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering in 2017-2020*
- Charge: "…*[E]xamine anticipated priorities and trade-offs for advanced computing for NSF-sponsored research"*
- Seven major recommendations across maintaining leadership, innovation, science needs, balance of capabilities (see backup slide).
- NSF formulating its response and actions.



Report: www.nap.edu/catalog/21886

NSF CI 2030 RFI

NSF Request for Information on Future Needs for Advanced Cyberinfrastructure to Support Science and Engineering Research

Dear Colleague Letter: www.nsf.gov/pubs/2017/nsf17031/nsf17031.jsp, Jan 5 - April 5, 2017

"NSF seeks input on scientific challenges, associated CI needs, and bold ideas to advance research frontiers over the next decade and beyond."

Question 1: Research Challenge(s) *[Including institutional challenges…]*

Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research and standing questions in the field.

Question 2: Cyberinfrastructure Needed to Address the Challenge(s).

Describe any limitations or absence of existing CI or specific advancements that must be addressed to accomplish the identified research challenge(s).

Question 3: ***Any other aspects or issues that NSF should consider.***

NSF CI 2030 RFI

First Look at Responses: Who?

Analysis is underway with NSF Advisory Committee for CI

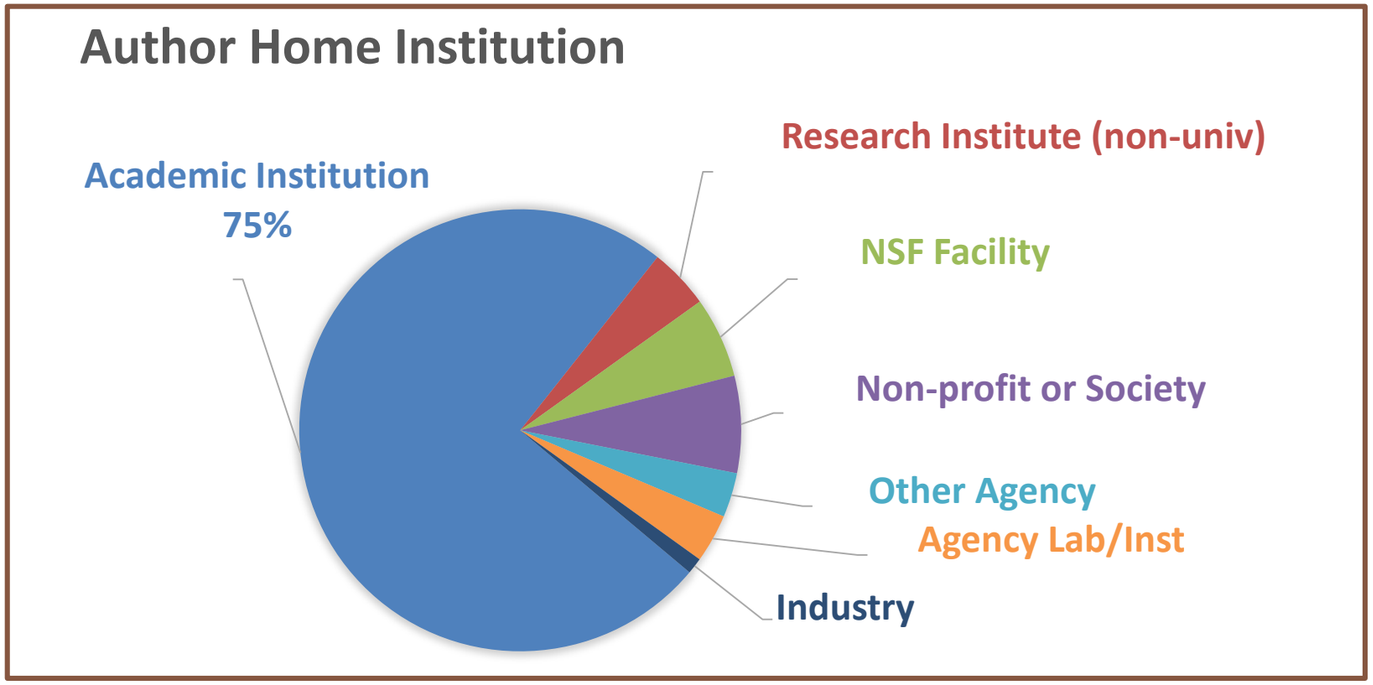
136 Submissions

- 50% Single Author
- 50% Groups (2-15 Authors)

366 Named Authors
(339 Unique)
(some were busy bees)

Geographic spread

- 39 States
- 9 Foreign contributors from 6 countries



NSF CI 2030 RFI

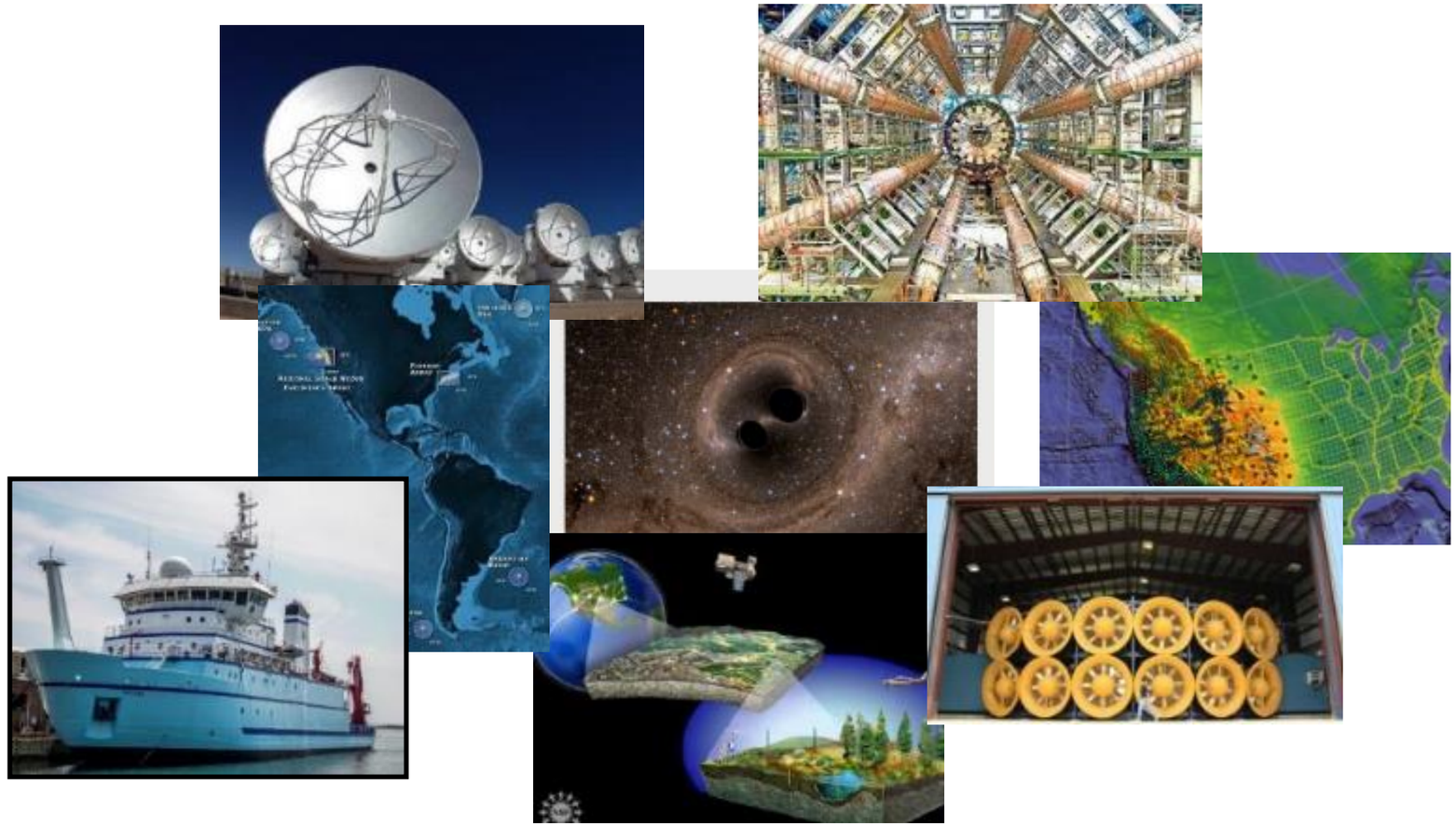
NSF Facilities are well Represented

NSF Facility Contributors

- ARF/UNOLS
- Gemini
- LHC
- LIGO
- LSST
- NCAR
- NHERI
- NHMFL
- NOAO
- NRAO
- NSO
- RCRV

Mentioned

- DKIST
- Ice Cube
- IODP
- NEON
- OOI



ID#212 - Marine Crew of the R/V Atlantis & Marine Crew of the R/V Armstrong Shipboard Scientific Support Group (SSSG), WHOI

Outline

- Overview of Cyberinfrastructure
- OAC update and programs
- Looking into the Future
- Facility CI focus, upcoming workshop

NSF Workshops on Facility Cyberinfrastructure

- **Origin.** NSF Facility POs recognized a common challenge area, and OAC began internal discussions.
- **Overall Goal**. Enable direct and synergistic interactions between the NSF large facilities and the Cyberinfrastructure (CI) communities to jointly address needs of current and future facilities.
- **Desired outcomes**: Foster *collaborations* and *communities of practice*. Encourage *sharing* of *practices and solutions*. Inform *NSF program planning*.
- **First workshop** FY 2016 (December 2015). PI: Alex Szalay/JHU. Broad agenda, and emphasized data issues. Very successful. NSF Director gave a keynote.
- **Second workshop** upcoming September 6-7, 2017, Alexandria VA. Focus will be on facility CI designs and architectures – what is being done, internal "IT" vs. external CI resources being used, common issues, opportunities.

Thanks!

…

Bill Miller
Office of Advanced Cyberinfrastructure
WLMiller@nsf.gov
703-292-7886

NAS Report on Future NSF HPC Summary of Recommendations

Leadership

1. Grow comprehensive investments in advanced computing.
2. Support full range of science requirements for advanced computing.

Meeting Needs

3. Collect community requirements; develop roadmaps to inform decisions and set priorities.
4. Adopt approaches to consider investments in an integrated way with associated research.

On the Cutting Edge

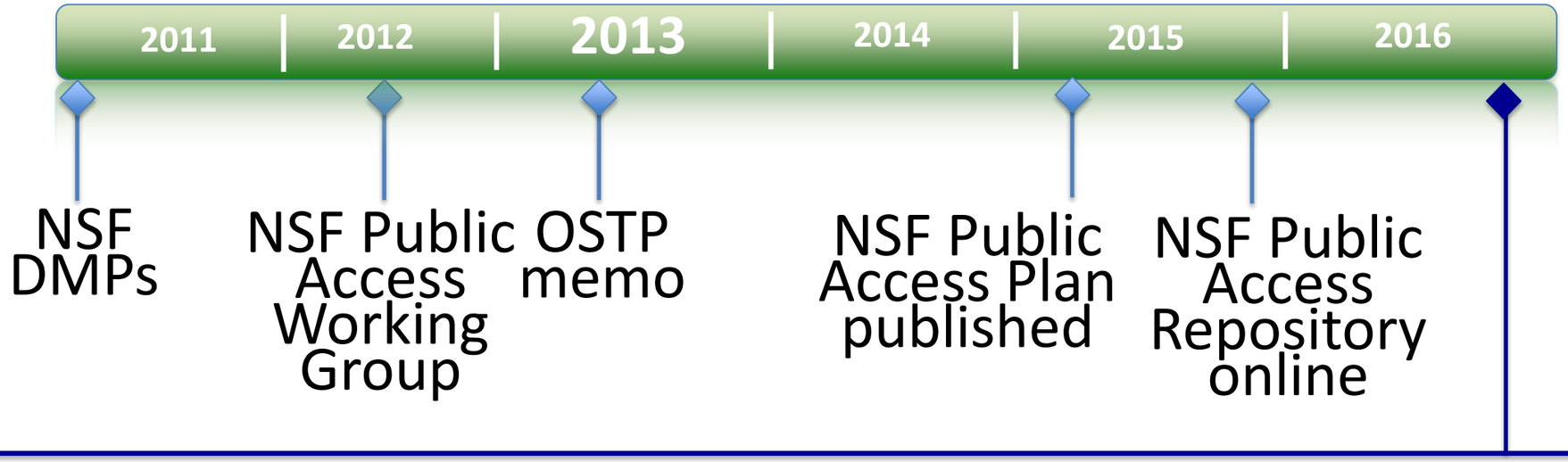
5. *Software.* Support development and maintenance of expertise, scientific software, and software tools relevant to advanced computing resources.
6. *Next-generation capabilities.* Make modest in next-gen hardware, software technologies to explore new ideas for next-gen capabilities. Adoption of radical new technology takes time.

Sustainability

7. Manage advanced computing investments in a predictable and sustainable way.

Report: www.nap.edu/catalog/21886

Public Access: recent USG activities (1/17)



US Government-wide (Jan. 2017)

- All subject Federal agencies have posted public access plans (compliance with 2013 OSTP memo)
- Published: “Principles for Promoting Access to Federal Government-Supported Scientific Data and Research Findings Through International Scientific Cooperation”
- New Open Science Working Group formed (NIH, NSF co-chairs)