Expanding Networked Communities to Advance Research Competitiveness

VIPIN CHAUDHARY (OAC)
MURIEL POSTON (BIO)
CHAITAN BARU (CISE)
DAWN TILBURY (ENG)
EVAN HEIT (HER)
JUAN MEZA (MPS)

MAY 21, 2018

Big Data and Cyberinfrastructure:

Building Jurisdictional Capabilities to Align with NSF Strategies

PANEL
EPSCoR PI Meeting

Credit: *NSF/Nicolle Rager Fuller*

Outline

What new forms of research within your directorate are enabled by the data revolution.?
How do you prepare the next generation of scientists/engineers with appropriate mix of skills within domain and BigData?
How can we leverage existing contributions to CI to ensure we connect diverse institutions, communities and individuals.

The NSF Big Ideas

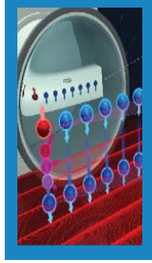
RESEARCH IDEAS

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: Multi-messenger Astrophysics



Quantum Leap: Leading the Next Quantum Revolution

Understanding the Rules of Life: Predicting Phenotype

PROCESS IDEAS

Mid-scale Research Infrastructure

NSF 2026

Growing Convergence Research at NSF

NSF INCLUDES: Enhancing STEM through Diversity and Inclusion

" … bold questions that will drive NSF's long-term research agenda -- questions that will ensure future generations continue to reap the benefits of fundamental S&E research. "

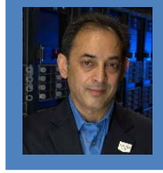
Big Ideas => Big Cyberinfrastructure Challenges & Opportunities

Engage NSF's research community in the pursuit of ***fundamental research in data science and engineering****, the development of a cohesive, federated, national-scale approach to* ***research data infrastructure****, and the development of* ***a 21st-century data-capable workforce****.*

Multi-agency effort to maximize the ***benefits of High Performance Computing (HPC)*** *for scientific discovery and economic competitiveness:* ***coherence in modeling/simulation and data analytics technologies****; viable post-Moore Law path for HPC; a capable and sustainable HPC ecosystem.*

NSF Office of Advanced Cyberinfrastructure

Program Staff


Manish Parashar*
Office Director

Amy Friedlander
Deputy Office Director

Bill Miller
Science Advisor
(On Detail)

Beth Plale*
Science Advisor
Public Access

Alejandro Suarez
Cooperative Agreements

Scott Sellars
AAAS S&T Policy Fellow

Computing	Data	Software	Networking & Cybersecurity	Learning & Workforce Development

Bob Chadduck

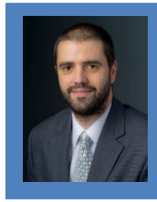
Amy Walton

Vipin Chaudhary*

TBD

Sushil Prasad*

Ed Walker


Stefan Robila*

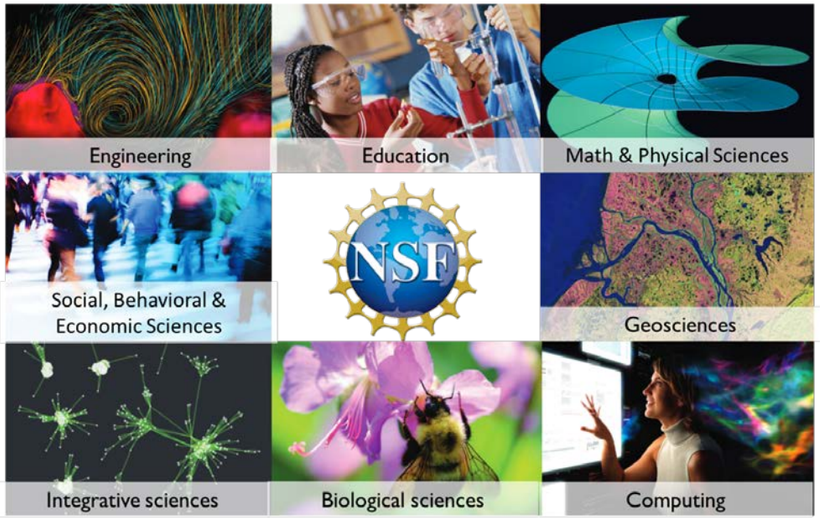
Rajiv Ramnath (Part-Time)*

Kevin Thompson

Join NSF/OAC: Multiple Program Officer openings

* IPA Appointment

CISE/OAC – Transforming the Frontiers of Science & Society



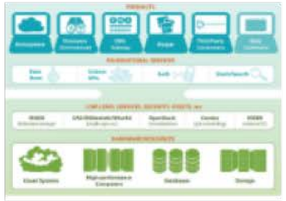
Foster a cyberinfrastructure ecosystem to transform computational- and data-intensive research across all of science and engineering

- Cyberinfrastructure Research & Research Cyberinfrastructure

CI-Enabled Instrumentation

Computing Resources

Data Infrastructure

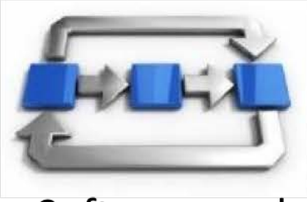


Gateways, Hubs, and Services

Cloud Resources & Services

R&E Networks, Security Layers

Coordination & User support



Software and Workflow Systems

Pilots, Testbeds

People, organizations, and communities

CISE/OAC – Transforming the Frontiers of Science & Society

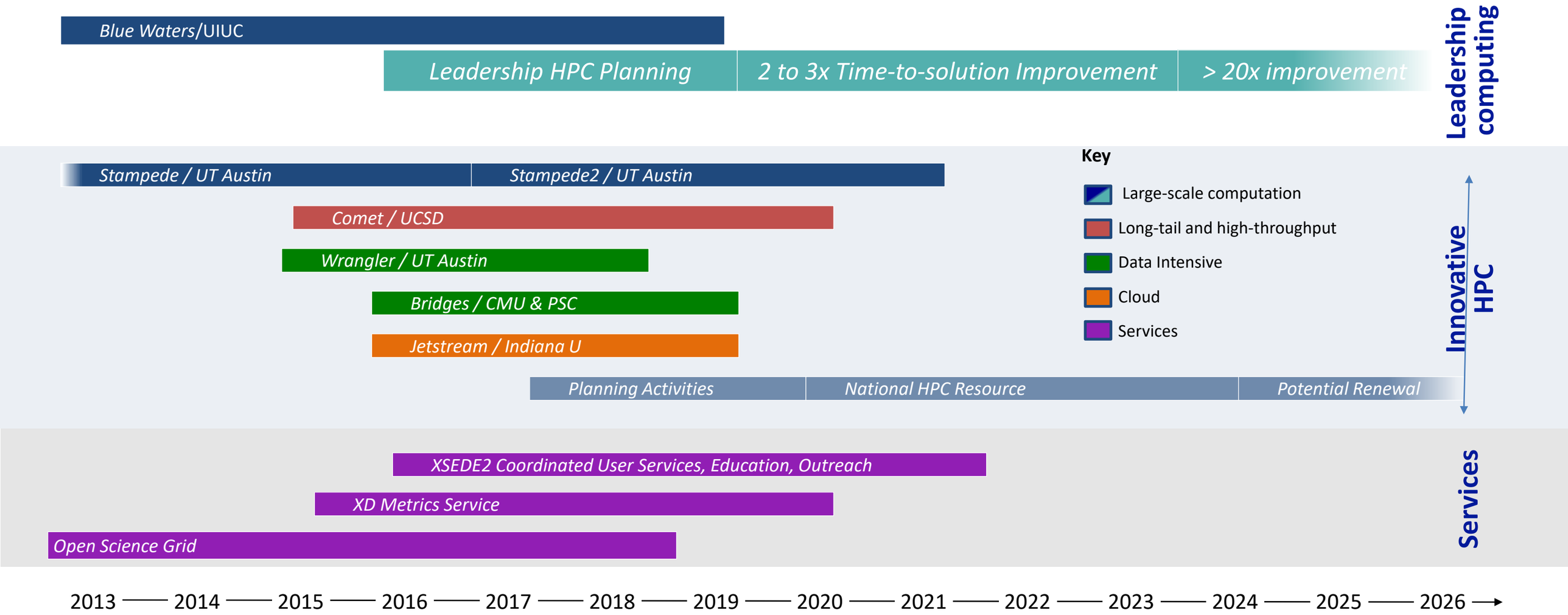
Computing	Advanced resources and services at all scales – MRI (clusters); Innovative HPC; Leadership Class; XSEDE coordination and user services; Research
Data	Data Building Blocks (DIBBS) Program
Software	Software Infrastructure for Sustained Innovation (SI2)
Networking & Cybersecurity	Campus Cyberinfrastructure (CC*), International Research Network Connections (IRNC), Cybersecurity Innovation for CI (CICI)
Learning & Workforce Devel	Training-based Workforce Development for Advanced Cyberinfrastructure (CyberTraining), CAREER, CRII

Cyberinfrastructure for Sustained Scientific Innovation (CSSI)

Emerging Opportunities

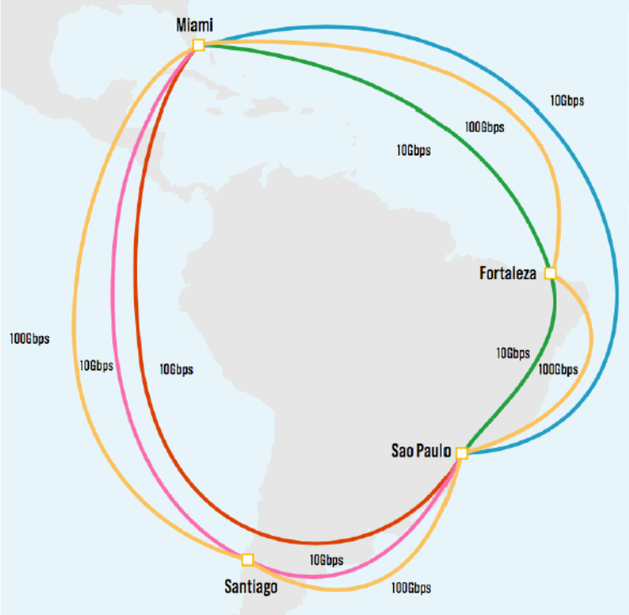
Cyberinfrastructure for Emerging Science and Engineering Research (CESER), Public Access

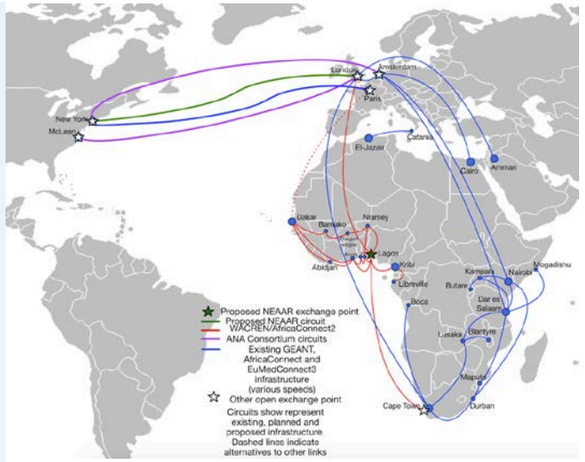
OAC-Funded National Computing Ecosystem

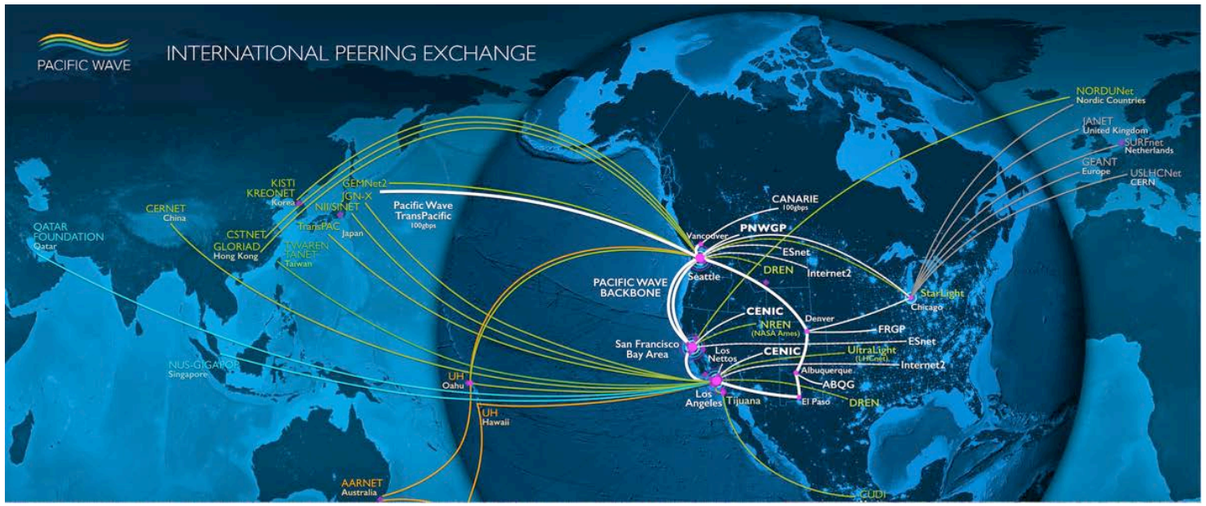


Networking Programs and Investments

- Networking as a fundamental layer and underpinning of CI
- CC* - Campus Cyberinfrastructure
  - Campus networking upgrades, e.g., re-design to scienceDMZ at campus border, 10/100Gbps
  - Innovation program
- IRNC – International R&E Network Connections
  - Scientific discovery as a global collaborative endeavor
    - Provide network connections linking U.S. research with peer networks in other parts of the world – 100Gbps links, software defined exchanges
    - Supports all R&E US data flows (not just NSF-funded)
  - Includes performance flow measurement, monitoring, training

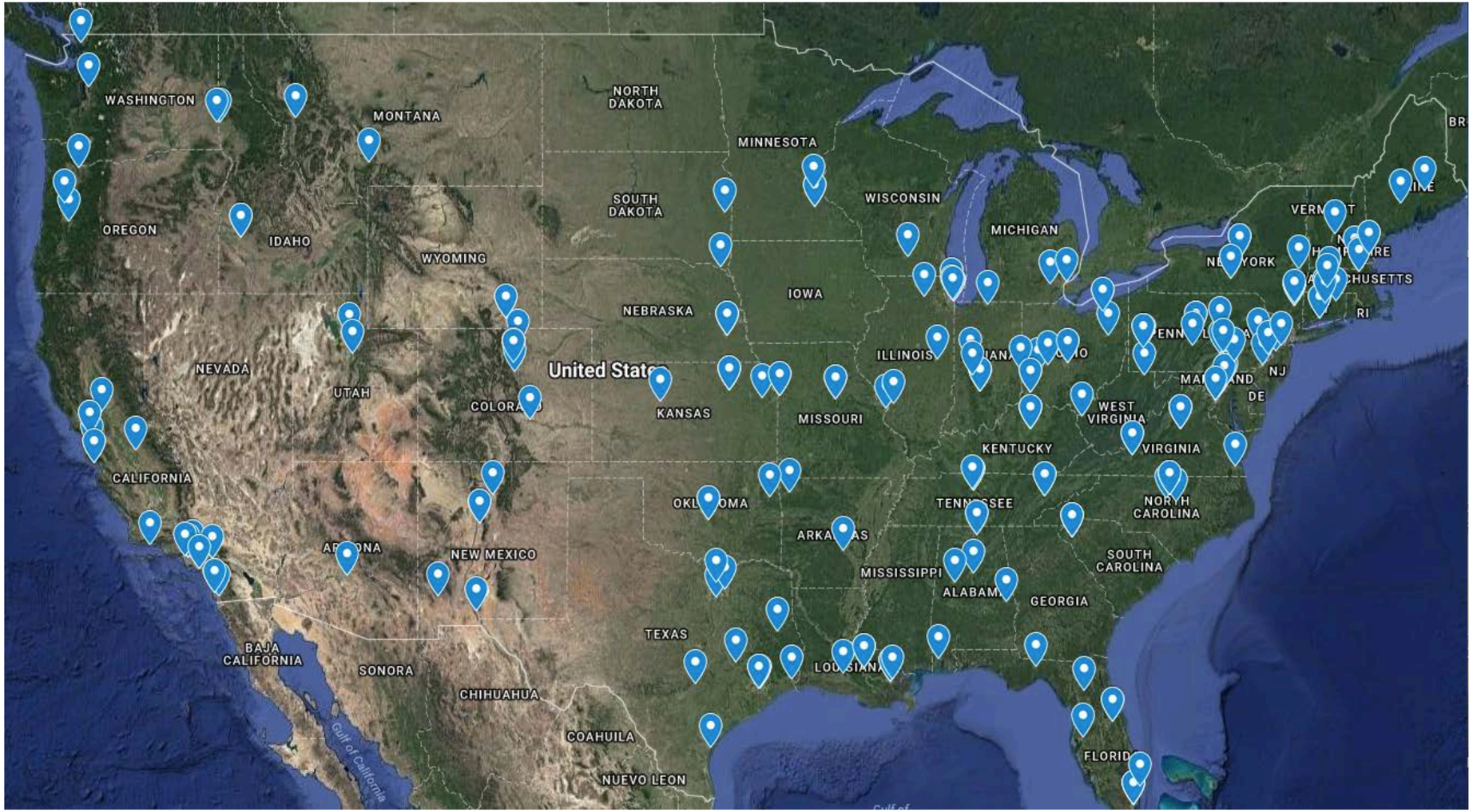






Building the Research Network Substrate at the Campus Level

CC* (2012-2017): 210+ awards across 44 states



OAC Cybersecurity

- Addressing the distinct cybersecurity requirements of science and engineering, including the large facilities
- Portfolio includes 2 types of awards:
  - Secure and Trustworthy Cyberspace (SaTC)
    - Funds basic security and privacy research
    - OAC focuses on Transition to Practice (TTP) projects, which aim to transition applied cybersecurity research projects into adoption and use in operational CI environments
  - Cybersecurity Innovation for Cyberinfrastructure (CICI)
    - Addresses the unique cybersecurity needs of CI in support of advanced computationally intensive scientific research
- Annual NSF Cybersecurity Summit for Large Facilities and Cyberinfrastructure
  - ~120 attendees from NSF-funded science facilities at last summit (August 2017)
  - Next summit planned for August 21-23, 2018 at the Westin, Alexandria (See http://trustedci.org)

TRUSTED CI
THE NSF CYBERSECURITY CENTER OF EXCELLENCE

Bro Intrusion Prevention/Detection software

New!

Cyberinfrastructure for Sustained Scientific Innovation (CSSI)

NSF 18-531
Due: 04/18

- Cross-directorate program that encompasses the Data Infrastructure Building Blocks (DIBBs) and Software Infrastructure for Sustained Innovation (SI^2).
- Supports innovative, and integrative, development and deployment of robust, reliable, sustainable data and software CI for scientific discovery and innovation.
- Flexible and responsive to evolving needs of science and engineering research.

Elements	Small groups that will create and deploy robust capabilities to advance one or more areas of science and engineering.
Framework Implementations	Larger, interdisciplinary teams for development and application of common, sustainable CI to address shared research challenges.

- Planned CSSI categories: Planning Grants for Community CI and Community CI Implementations that aim to establish long-term CI capabilities and hubs of excellence.

www.nsf.gov/funding/pgm_summ.jsp?pims_id=505505

Learning and Workforce Development

Communities of Concern

CI Contributors, Cyber-scientists

Research and Develop new CI

CI Professionals

Deploy & support CI

CI Users

Exploit CI for Reseach

- **Research Programs in Advanced Cyberinfrastructure (CI)**
  - ***CAREER – NSF 17-537***

    Submissions to OAC doubled in FY2016, tripled in FY2017; 30 active awards

    *Now open to **non-tenure track faculty**; Senior personnel allowed*
  - ***CRII – NSF 17-552 CISE-wide***

    *Independent research by Faculty or **research scientists** in their first 3 years*
  - ***Research Experience for Undergraduates - NSF 13-542***

    *Active research participation by **undergraduate students***
- **Education/Training/Internships**
  - ***CyberTraining (NSF 18-516)***

    ***Scalable** training, education and curricular activities for **research workforce development** in advanced CI, computational and data science*
  - *Non-academic Research Internships (NSF 17-091)*
- **EAGERs, Workshops, Student Travel Grants**
  - *Exploration of Research, Education, and Broadening Participation*

Outline

OAC Overview & Update
Emerging Science Role of CI
Conclusion

An Evolving Science, CI Landscape

Evolving Science/Engineering Landscape

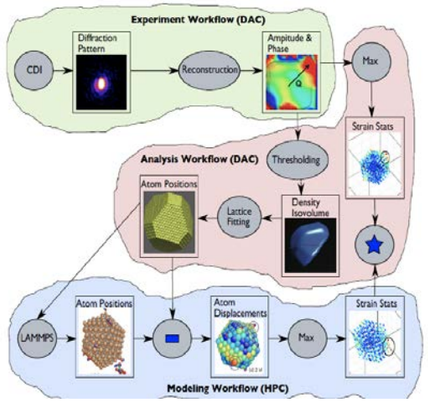
- High-resolution multi-scale, multi-physics simulations / Data-driven models
- Streaming data from observatories, instruments
  - Disconnected from each other, CI services
- Growing “long-tail” / Growing “gateway” jobs / Increasing use of clouds

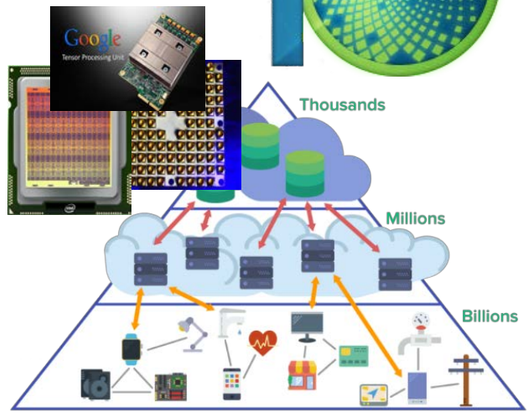
Evolving Technology Landscape

- Extreme scales / pervasive computing and data / disruptive technologies
- Novel paradigms / growing capabilities & capacities at the edges
- Unconventional software stacks
- High throughput/low-latency networks
- New concerns

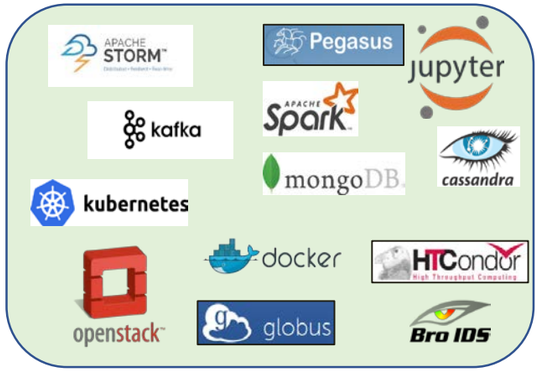
Cyberinfrastructure ecosystem must evolve

End-to-end Workflows





Instrument, Observatories, Experiential Facilities

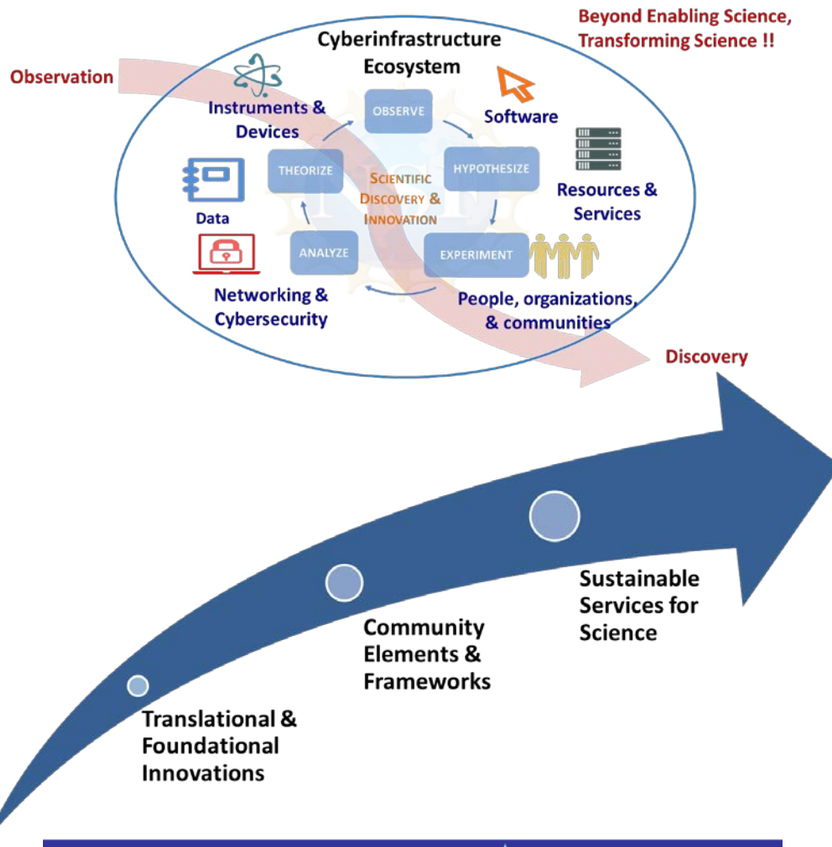


neon National Ecological Observatory Network
OOI OCEAN OBSERVATORIES INITIATIVE
ARRAY OF THINGS
LSST Large Synoptic Survey Telescope
INTERNET2

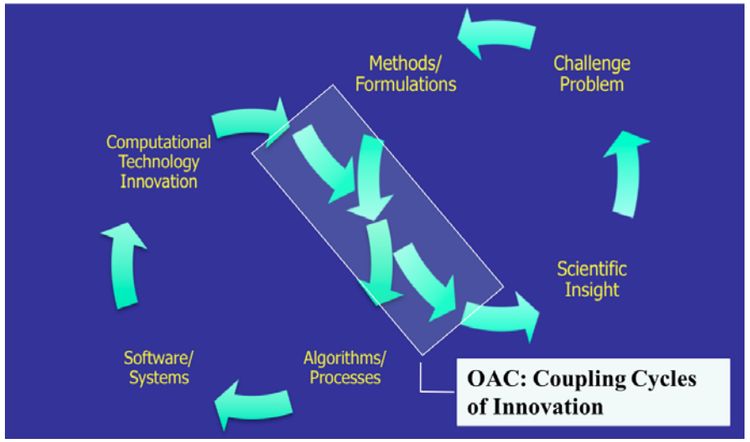
Cyberinfrastructure is central to NSF's Large Facilities…
…and touches all OAC investment areas (computing, data, software, networking, cybersecurity, learning and workforce)

Research success depends on robust, reliable, and highly connective cyberinfrastructure

Realizing a Cyberinfrastructure Ecosystem to Transform Science



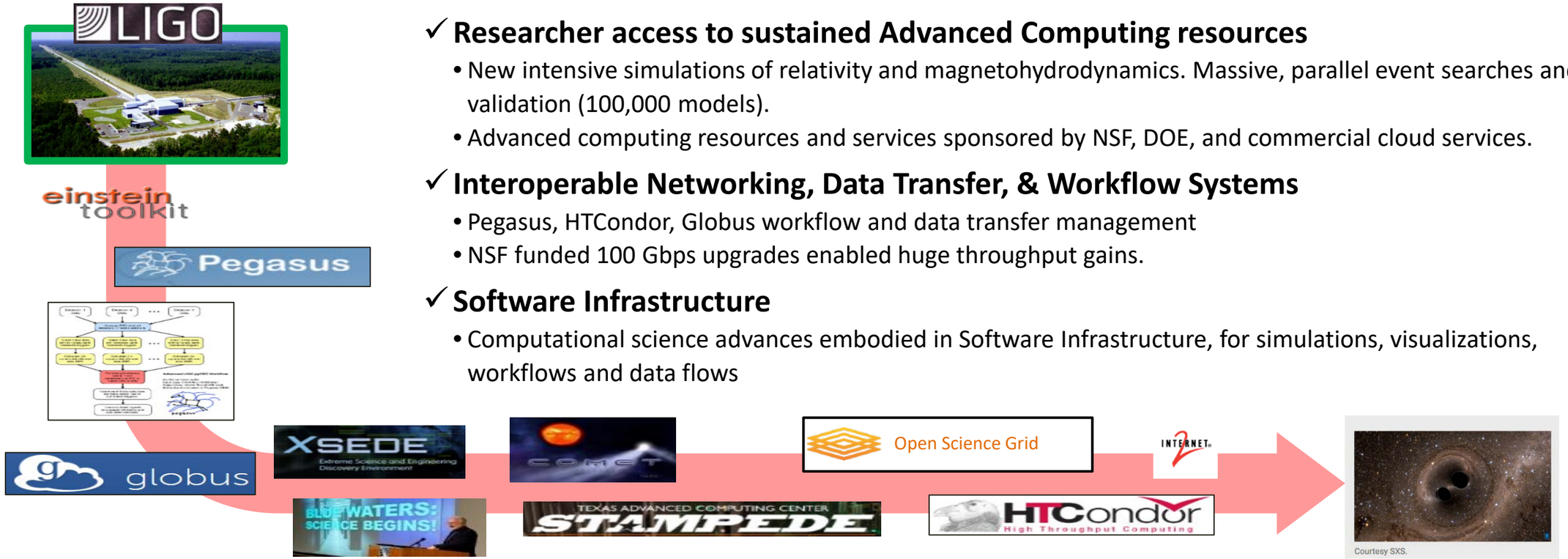
- Realize a holistic and integrated cyberinfrastructure ecosystem aimed at transforming science
- Support the translational research continuum, from catalyzing core innovations, through fostering the community tools and frameworks, and enabling sustainable cyberinfrastructure services
- Work closely with science and engineering communities to tightly couple the cycles of discovery and innovation



Transforming Science through a Cyberinfrastructure Ecosystem: Dynamic discovery pathways at scale

CI enables Big Science

Gravitational wave detection enabled by NSF investments across the computational and data science workflow



✓ **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 sponsored by NSF, DOE, and commercial cloud services.

✓ **Interoperable Networking, Data Transfer, & Workflow Systems**

- Pegasus, HTCondor, Globus workflow and data transfer management
- NSF funded 100 Gbps upgrades enabled huge throughput gains.

✓ **Software Infrastructure**

- Computational science advances embodied in Software Infrastructure, for simulations, visualizations, workflows and data flows

NSF programs: Data Building Blocks (DIBBs), Software Infrastructure (SI^2), Campus Cyberinfrastructure Network Infrastructure and Engineering (CC*NIE, DNI), and others. OSG and Pegasus are also supported by the Department of Energy.

Outline

OAC Overview & Update
Emerging Science Role of CI
Conclusion

Building on Community Input: Results of NSF CI 2030 Request for Information

Common needs expressed across science and engineering domains:

- **Advanced computing.** Growing need for on-demand computing for steering large simulations, rapid data processing, experiments; comparing simulations and observation.
- **Data Science and management.** Big Data and Machine Learning. Automated mining, analytics, visualization, provenance. Discoverability, accessibility, and reproducibility.
- **Multi-source streaming data**. Processing and integrating data from the Internet of Things (IOT) and cyber-physical systems at human, community, urban, and ecosystems scales.
- **Secure access, dynamic and high bandwidth workflows.** Technologies & approaches that scale with performance demands; storage, identity management, cybersecurity.
- **Software.** Porting, accelerating, validating algorithms and community codes. Software quality, reliability, validity, practices.
- **Training and workforce development.** For researchers and computing professionals, diversity and inclusion. CS/CI experts who collaborate closely with domain researchers.

All responses posted on CI 2030 Website: www.nsf.gov/cise/oac/ci2030/

Snapshot of CI concerns and needs across all S&T fields…

(A simple wordle of all text of all responses)

"If you want to travel fast, travel alone;
if you want to travel far, travel together"

African Proverb.

Vipin Chaudhary

Program Director, Office of Advanced Cyberinfrastructure

Email: vipchaud@nsf.gov

To subscribe to the OAC Announce Mailing List
Send an email to: OAC-ANNOUNCE-subscribe-request@listserv.nsf.gov