Realizing a Cyberinfrastructure Ecosystem that Transforms Science

Manish Parashar
Office Director

Office of Advanced Cyberinfrastructure,

Directorate for Computer & Information Science & Engineering

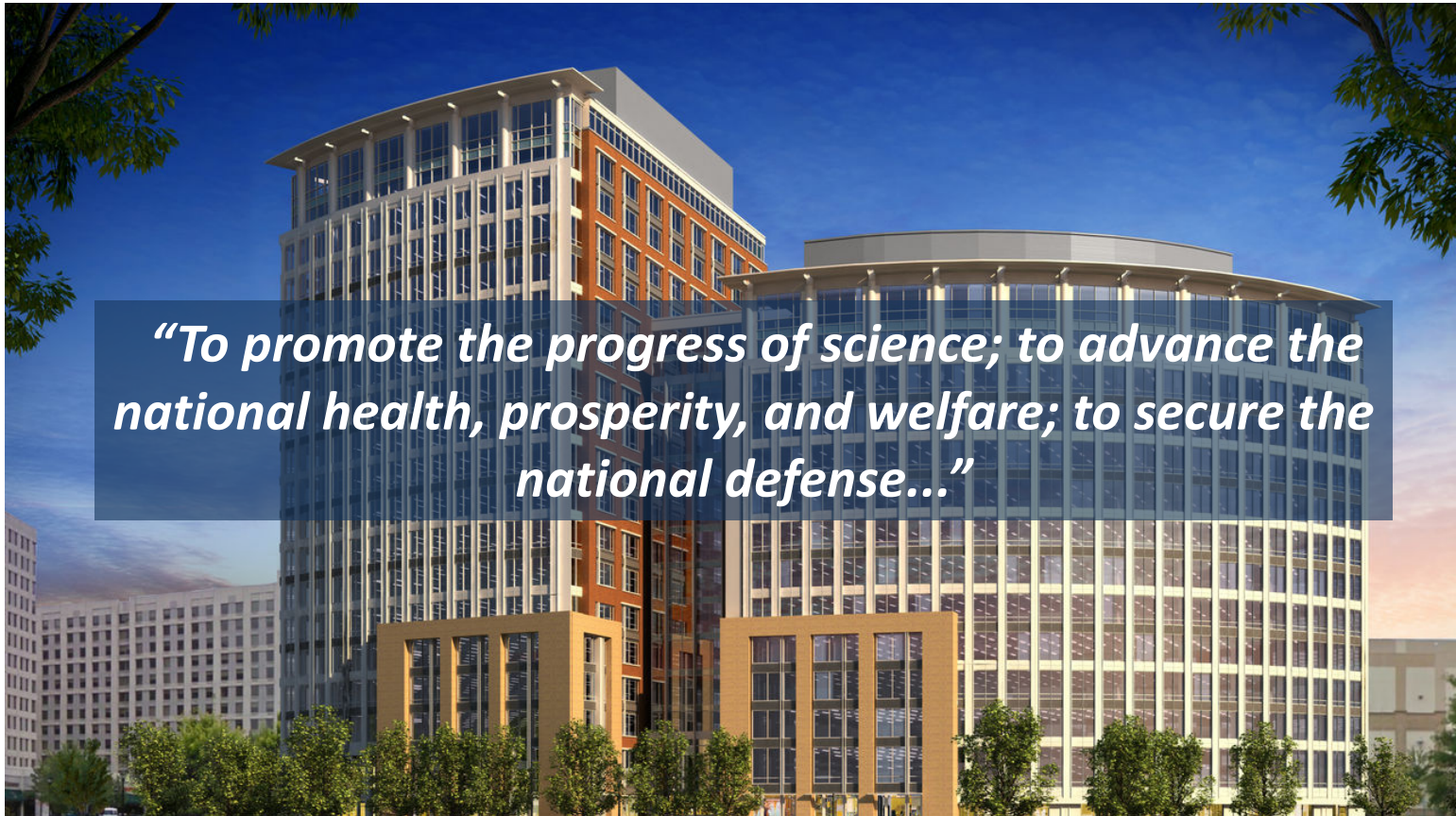National Science Foundation

SAC-PA
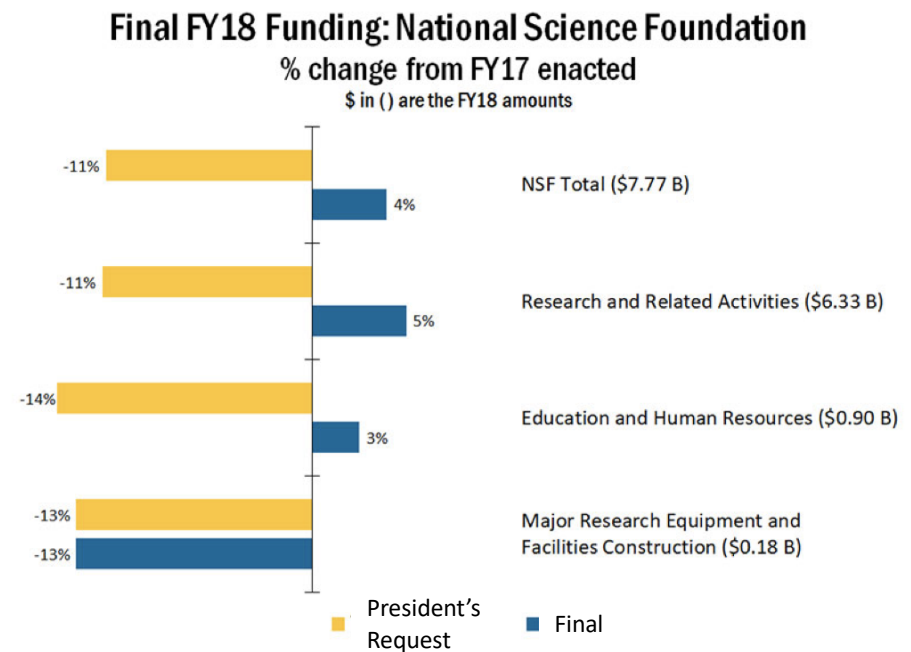
June 14, 2018

Outline

Outline

NSF Update → OAC Overview & Update → Emerging Science Role of CI → Conclusion

National Science Foundation's Mission

"To promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense..."

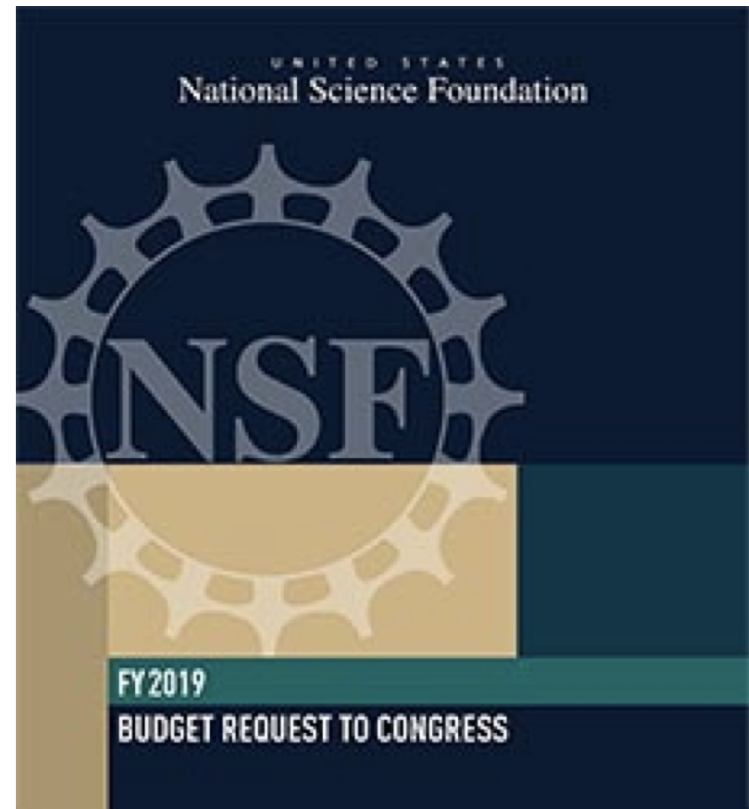# FY 2018 Budget Appropriations

Congressionally Appropriated Budget: $7.8 billion

- +4%
- +$300M over FY17 budget
    - 2nd largest increase in NSF research budget in 15 years (not counting ARRA)

Final FY18 Funding: National Science Foundation
% change from FY17 enacted
$ in ( ) are the FY18 amounts

-11% / 4% — NSF Total ($7.77 B)

-11% / 5% — Research and Related Activities ($6.33 B)

-14% / 3% — Education and Human Resources ($0.90 B)

-13% / -13% — Major Research Equipment and Facilities Construction ($0.18 B)

President's Request / Final

Source: American Institute of Physics / www.aip.org/fyi

NSF FY 2019 Budget Request

- NSF: $7.47 billion
  - Flat with respect to FY 17 Enacted

- CISE: $925.4 million
  - -1.1% from FY 17 Enacted

- Big Ideas
  - Research Ideas: $30 million each
  - Process Ideas:
    - Midscale infrastructure: $60 million

- Convergence Accelerators
  - $60 million
  - HDR, FW-HTF: $30 million each

UNITED STATES
National Science Foundation

NSF

FY 2019
BUDGET REQUEST TO CONGRESS

The NSF Big Ideas

RESEARCH IDEAS

Harnessing the Data Revolution

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

Convergence Accelerators

Accelerating Discovery through Convergence Research

Motivation: Changing nature of science research - research frontiers at intersection of existing disciplines

- Time-limited entities: accelerating impactful *convergence* research in areas of national importance
- Innovating in organizational structure to better enable frontier research
- Separate (from directorates) in leadership, budget, and programmatics; but aligned with, relying on, and building on foundational disciplinary research
- Emphasis on translational research, partnerships

HDR Convergence Accelerator

FW-HTF Convergence Accelerator

Future accelerator(s)

BIO CISE EHR ENG GEO MPS SBE

National Strategic Computing Initiative (NSCI)

Maximizing benefits of HPC for scientific discovery and economic competitiveness

- *Multi-agency effort to maximize the **benefits of High Performance Computing (HPC)** for scientific discovery and economic competitiveness*

- NSF/CISE/OAC leads interagency effort to:
  - Increase coherence between technology base used for modeling/simulation and for data analytics;
  - Establish viable path forward for HPC systems in post-Moore's Law era; and
  - Increase capacity, capability, and sustainability of an enduring national HPC ecosystem

National Strategic Computing Initiative (NSCI)

Includes CISE investments in the following programs: High Performance Computing (HPC), Scalable Parallelism in the Extreme (SPX), Quantum Leap Big Idea, and Cyberinfrastructure for Sustained Scientific Innovation – Software (CSSI)
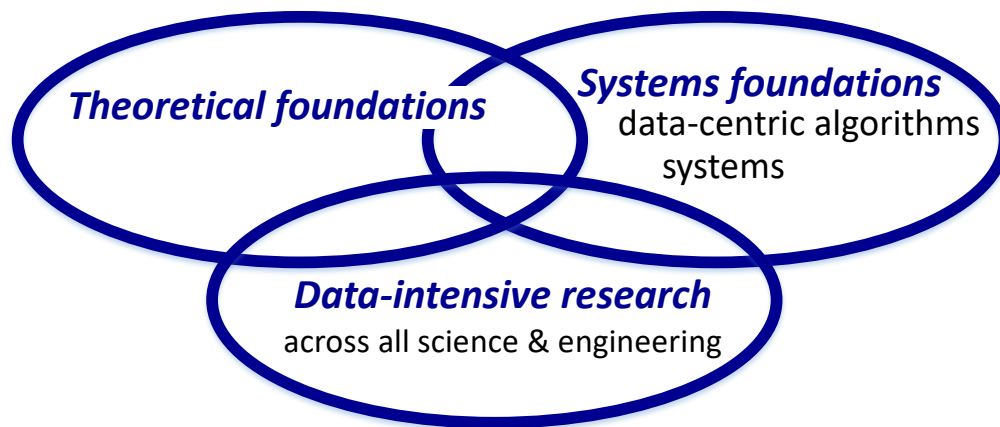
"*Engage NSF's research community in the pursuit of **fundamental research in data science and engineering**, the development of a cohesive, federated, national-scale approach to **research data infrastructure**, and the development of **a 21st-century data-capable workforce**.*"

Harnessing the Data Revolution (HDR)

Research across all NSF Directorates

Theoretical foundations

Systems foundations
data-centric algorithms
systems

Data-intensive research
across all science & engineering

Educational pathways

Innovations grounded in an education-research-based framework

Advanced cyberinfrastructure
Accelerating data-intensive research.

Harnessing the Data Revolution (HDR)

Research across all NSF Directorates

Theoretical foundations
TRIPODS

Systems foundations
data-centric algorithms
systems: **BIGDATA;**
OKN

Data-intensive research
across all science & engineering
TRIPODS+X
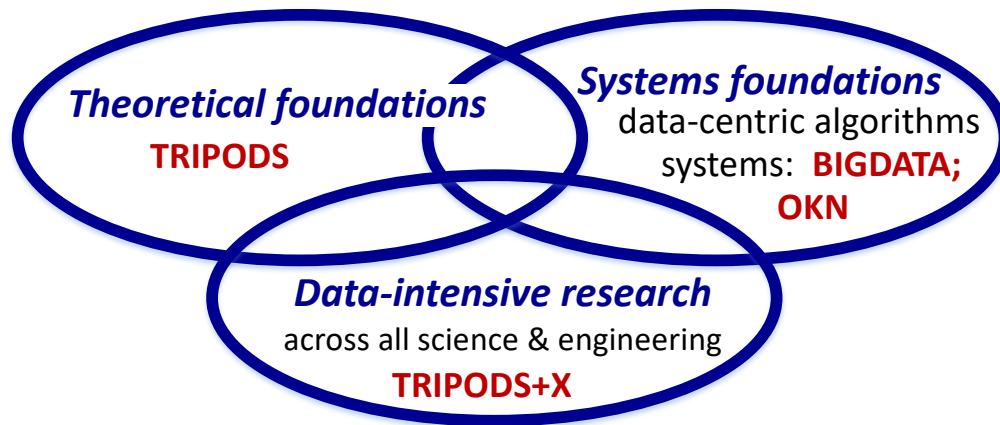
Educational pathways

Innovations grounded in an education-research-based framework
NASEM study on data science at the undergraduate level; NSF Research Traineeships; GRFP

Advanced cyberinfrastructure
Accelerating data-intensive research.
CSSI;
Scalable data-driven CI DCL;
Midscale infrastructure (Midscale RFI)

Quantum Leap: Leading the Quantum Revolution

- **Fundamentals** that advance our understanding of uniquely quantum phenomena and their interface with classical systems
- **Elements** that measure, model, control, and exploit quantum particles
- **Software systems and algorithms** that enable quantum information processing
- **Workforce**, including training a new generation of scientists, engineers

Outline

NSF Update → OAC Overview & Update → Emerging Science Role of CI → Conclusion
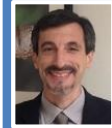
NSF Office of Advanced Cyberinfrastructure

Program Staff

Manish Parashar *
Office Director

Amy Friedlander
Deputy Office Director

Bill Miller
Science Advisor
(On Detail)

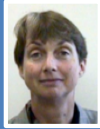| Computing	Data	Software	Networking & Cybersecurity	Learning & Workforce Development

Bob Chadduck

Amy Walton

Vipin Chaudhary *

TBD

Sushil Prasad *

Beth Plale *
Science Advisor
Public Access

Ed Walker

Stefan Robila *

Rajiv Ramnath * (Part-Time)

Kevin Thompson

Alejandro Suarez
Cooperative Agreements

Scott Sellars
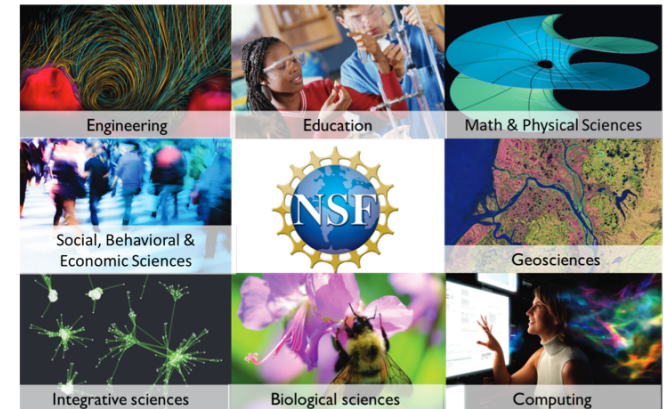AAAS S&T Policy Fellow

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

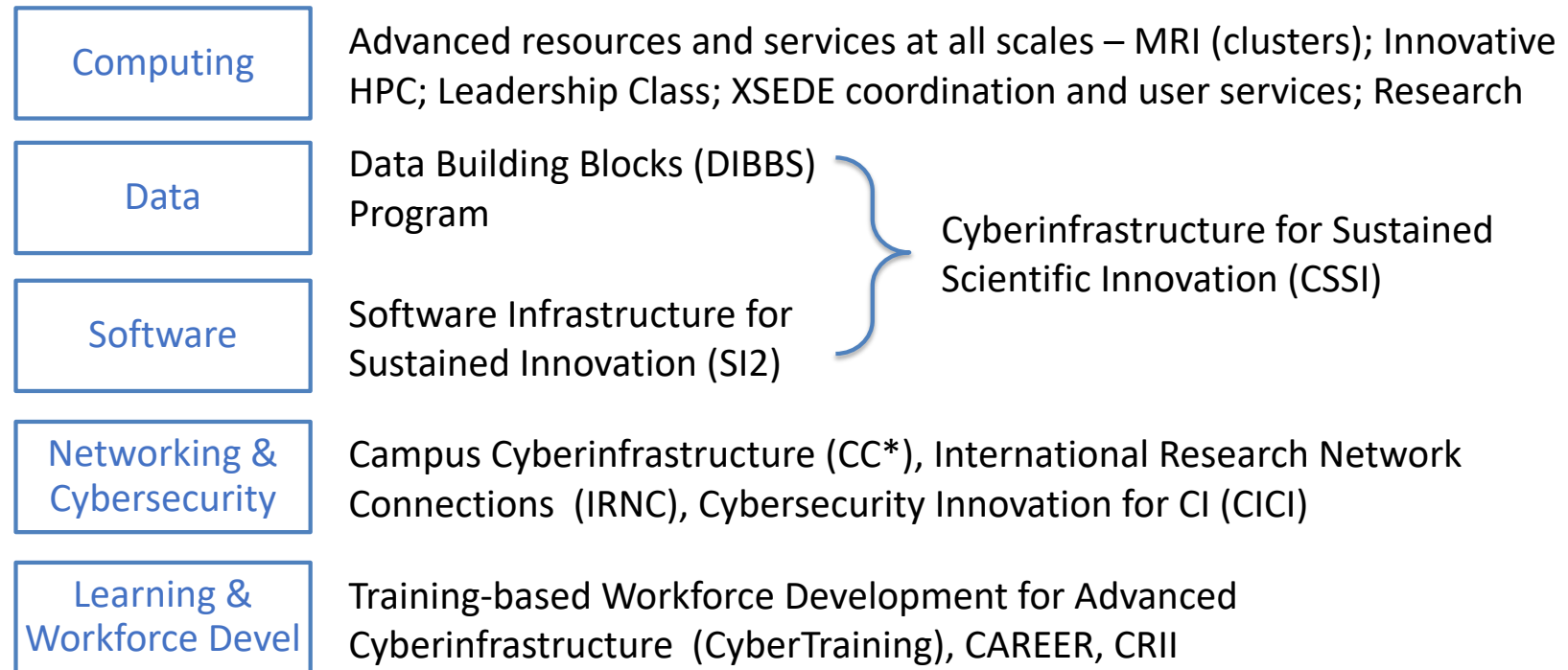R&E Networks, Security Layers

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 |

Cyberinfrastructure for Sustained Scientific Innovation (CSSI)

| 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 |

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

Outline

NSF Update ▸ OAC Overview & Update ▸ Emerging Science Role of CI ▸ Conclusion

Evolving Science, CI Landscapes

Evolving Science/Engineering Landscape

- Large scales, high-resolution, multi-scale, multi-physics simulation workflows
- Data-driven (ML-based) models, execution
- Streaming data from observatories, instruments
  - Disconnected from each other, CI services
- Rapidly growing "long-tail", "gateway" jobs; "small" jobs dominating; increasing use of clouds

Evolving Technology Landscape

- Extreme scales / pervasive computing and data
- Rapidly evolving / disruptive technologies
- Novel paradigms / growing capabilities & capacities at the edges
- Unconventional software stacks
- High throughput/low-latency networks
- New concerns (precision, correctness, reproducibility, reliability, energy, security, ...)
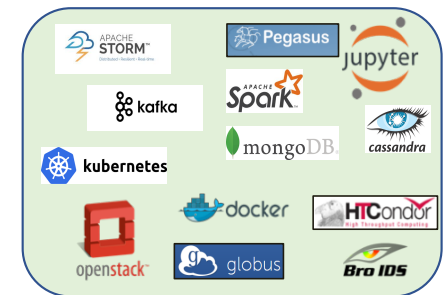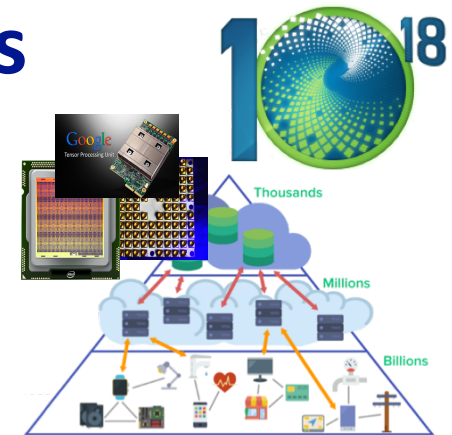
Cyberinfrastructure ecosystem must evolve

End-to-end Workflows
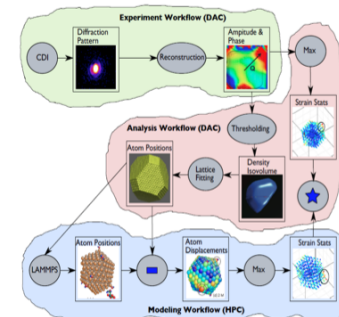
Instrument, Observatories, Experimental Facilities

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

23

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 <u>on-demand computing</u> for steering large simulations, rapid data processing, experiments; comparing simulations and observation.

- **Data Science and management.** <u>Big Data</u> and <u>Machine Learning</u>. Automated mining, analytics, visualization, provenance. Discoverability, accessibility, and reproducibility.

- **Multi-source streaming data**. Processing and integrating data from the <u>Internet of Things</u> (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/

24

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, and other stakeholders to tightly couple the cycles of discovery and innovation
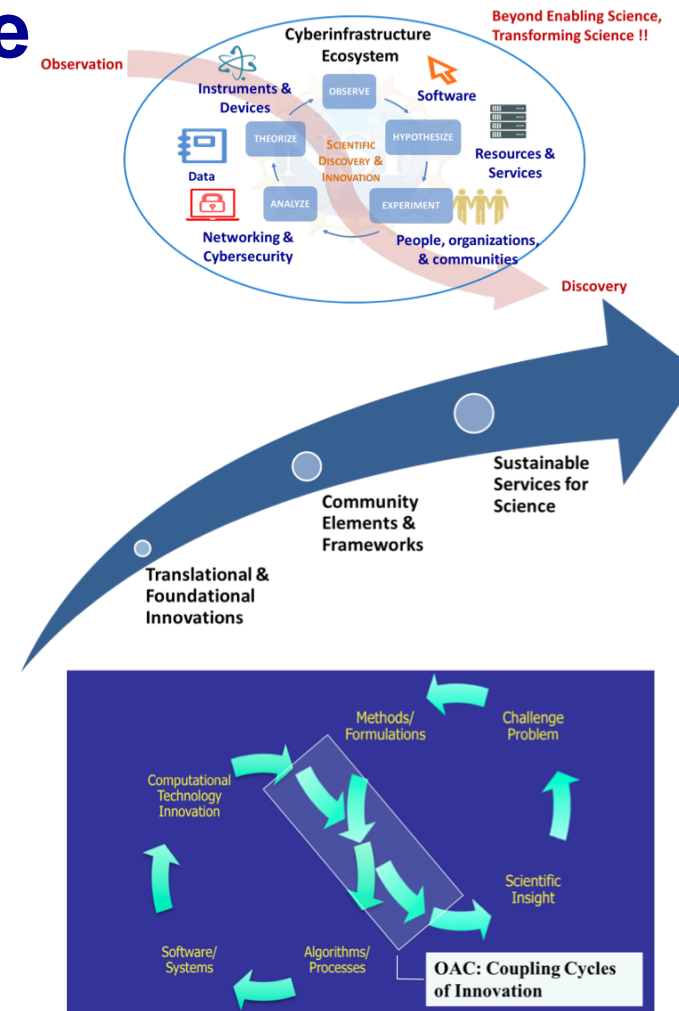
Cyberinfrastructure for Sustained Scientific Innovation (CSSI)

New!

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

Public Access

Foundation-wide Initiative housed in OAC

- Implement NSF Public Access Plan (NSF 15-52): Today's Data; Tomorrow's Discoveries
  - Instantiated requirements in NSF policies, including terms and conditions of awards
  - Deployed NSF Public Access Repository (PAR), hosted by DOE/OSTI

- DCL: Advancing Long-term Reuse of Scientific Data
  - Community, Data Reuse and Socio-Technical Infrastructure tracks
 - Supports Conferences, EAGERS (Due 05/23/18)

NSFS 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

NSF 18-060

Dear Colleague Letter: Advancing Long-term Reuse of Scientific Data

April 6, 2018

Dear Colleagues:

Through this Dear Colleague Letter (DCL), the National Science Foundation's (NSF) Office of Advanced Cyberinfrastructure (OAC) announces its intention to support initial exploratory activities toward the creation of social and technical infrastructure solutions that further NSF's commitment to public access. These solutions are a means to accelerate the dissemination and use of fundamental research results in the form of data that will advance the frontiers of knowledge and help sustain the Nation's prosperity well into the future.
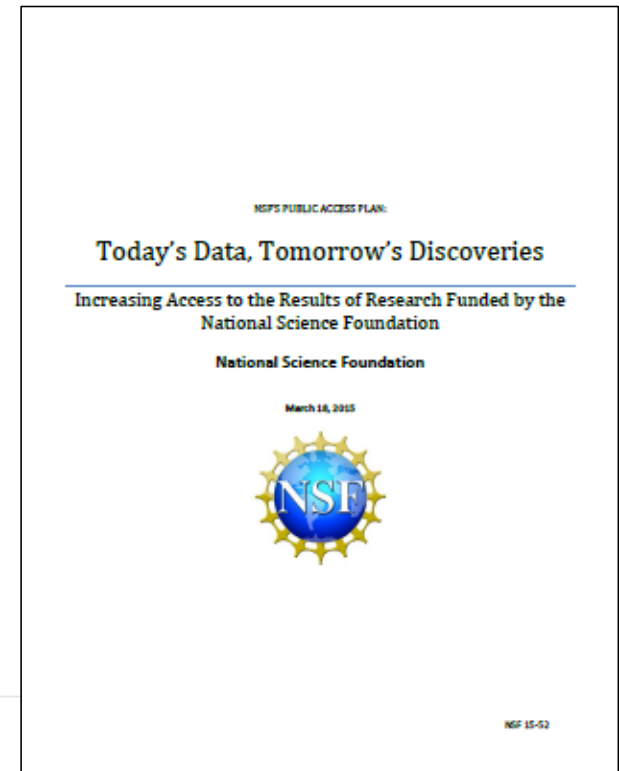
DCL: Scalable Cyberinfrastructure to Accelerate Data-Driven Science and Engineering Research (NSF 18-076)

- Scalable data-driven cyberinfrastructure (CI) exemplars that will accelerate discovery for science and engineering research communities, capitalizing on and enhancing existing NSF priority investments
  - Have the potential to rapidly expand or scale capacity and impact within 18 months

- Examples of potential topics include (but are not limited to):
  - Incorporating streaming data, intelligent data delivery, and real-time feedback loops between data collection and processing to enable design of smart infrastructures and provision of real-time information for better analysis, visualization, and discovery
  - Enriching scientific value of community data via integration of diverse and distributed datasets from multiple instruments in novel ways to enhance processing, analysis, sharing, and new science pathways

- Proposals due by June 20, 2018

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

CI enables Big Science

Gravitational wave detection enabled by NSF investments across the cyberinfrastructure ecosystem

✓ **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

NSF Update → OAC Overview & Update → Emerging Science Role of CI → Conclusion

Conclusion

- Science and society are being transformed by compute and data – an integrated cyberinfrastructure ecosystem is essential

- Rapidly changing application requirements; resource and technology landscapes
  - Our cyberinfrastructure ecosystem must evolve in response

- Lets build a holistic and integrated cyberinfrastructure ecosystem aimed at transforming science

Join the conversation

- OAC Webinar Series
  - 3rd Thursday @ 2PM ET
- OAC Newsletter
- Follow us on Twitter @NSF_CISE

Stay informed

- Join the OAC, CISE Mailing Lists
  - Learn about NSF events, programs, webinars, etc.
  - Send email to:
    - oac-announce@listserv.nsf.gov
    - cise-announce-subscribe-request@listserv.nsf.gov

Get involved

- Reviews proposals, serve on panels
- Visit NSF, get to know your programs and Program Officers
- Participate in NSF workshops and visioning activities
- Join NSF: serve as Program Officer, Division Director, or Science Advisor

"Make no little plans; They have no magic to stir men's blood ..."

Daniel H. Burnham, Architect and City Planner Extraordinaire, 1907.

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

African Proverb.

THANKS!

Manish Parashar

Office Director, Office of Advanced Cyberinfrastructure

Email: mparasha@nsf.gov

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