# Realizing a Cyberinfrastructure Ecosystem that Transforms Science


Vipin Chaudhary
Manish Parashar
Sushil Prasad
Scott Sellers
Alejandro Suarez

Office of Advanced Cyberinfrastructure,

Directorate for Computer & Information Science & Engineering

National Science Foundation

PEARC'18

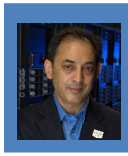
July 25, 2018

# Outline

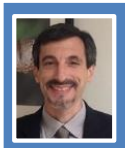

NSF/OAC Update
Emerging Science & Role of CI
Future Directions: OAC, HDR
Conclusion

# NSF Office of Advanced Cyberinfrastructure

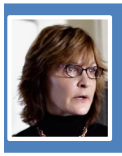
## Program Staff


Manish Parashar *
Office Director


Amy Friedlander
Deputy Office Director


Bill Miller
Science Advisor
(On Detail)

| Computing | Data | Software | Networking & Cybersecurity | Learning & Workforce Development |
|---|---|---|---|---|


Beth Plale *
Science Advisor
Public Access


Bob Chadduck


Amy Walton

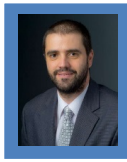

Vipin Chaudhary *

TBD


Sushil Prasad *


Alejandro Suarez
Cooperative Agreements


Ed Walker


Stefan Robila *


Rajiv Ramnath * (Part-Time)


Kevin Thompson


Scott Sellars
AAAS S&T Policy Fellow

**Join NSF/OAC: Multiple Program Officer openings**

* IPA Appointment

# NSF Budget Update


## FY 2018

- FY18 Congressionally Appropriated Budget
  - +4%
  - +$300M over FY17 budget!
- 2nd largest increase in NSF research budget increase in 15 years! (not counting ARRA)

## FY 2019

- 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

# The NSF Big Ideas

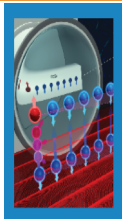
## RESEARCH IDEAS


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

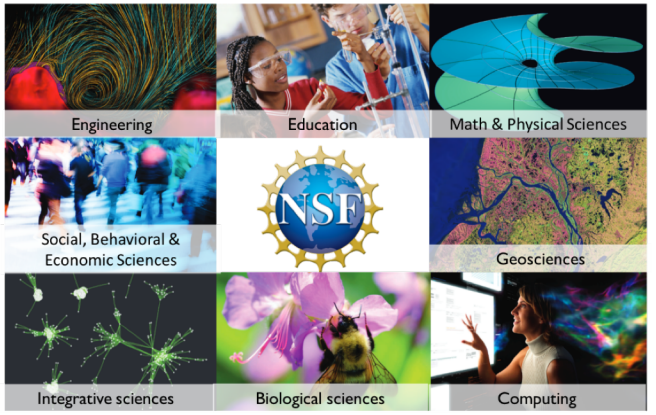
" … bold questions that will drive NSF's long-term research agenda -- questions that will ensure future generations continue to reap the benefits of fundamental S&E research. "

**Big Ideas => Big Cyberinfrastructure Challenges & Opportunities**

# 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

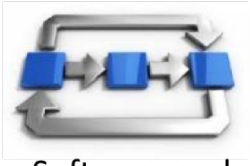

Data Infrastructure


Gateways, Hubs, and Services



R&E Networks, Security Layers

Coordination & User support


Software and Workflow Systems


Pilots, Testbeds

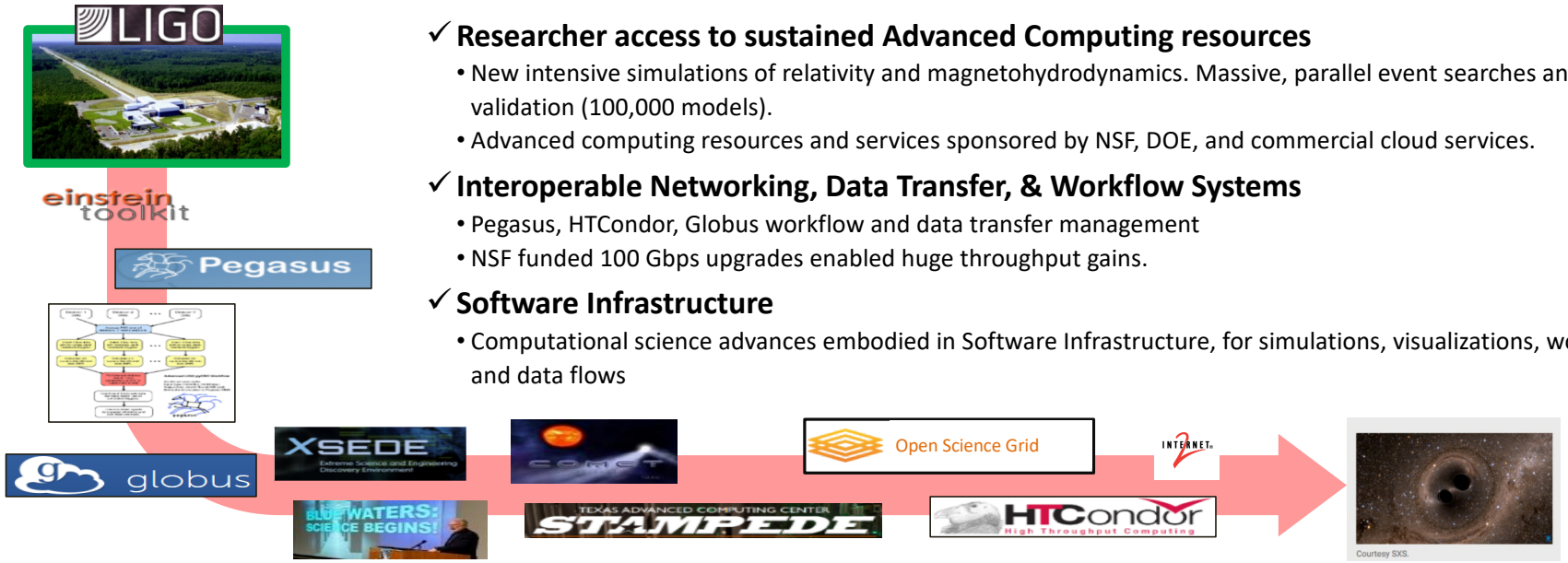

People, organizations, and communities

# 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) |
| | Cyberinfrastructure for Sustained Scientific Innovation (CSSI) |
| 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 |

# NSF Cyberinfrastructure Investments Enable 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.

# Evolving Science, CI Landscapes

## Evolving Science/Engineering Landscape

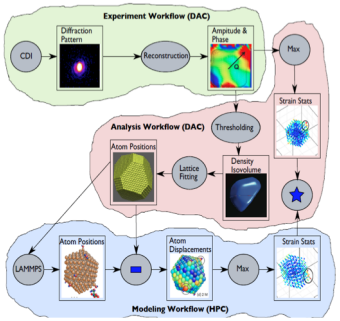
- Large scales, high-resolution, multi-scale, multi-physics simulation workflows
- Data-driven (ML-based) models, execution
- Streaming data from observatories, instruments
  - Disconnected from each other, CI services
- Rapidly growing “long-tail”, “gateway” jobs; “small” jobs dominating; increasing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

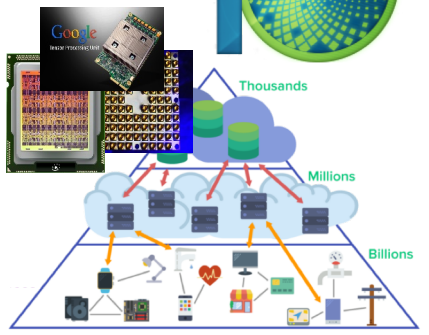
## Evolving Technology Landscape

- 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, ...)

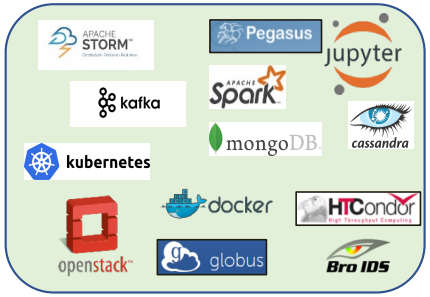
**Cyberinfrastructure ecosystem must evolve**

**End-to-end Workflows**



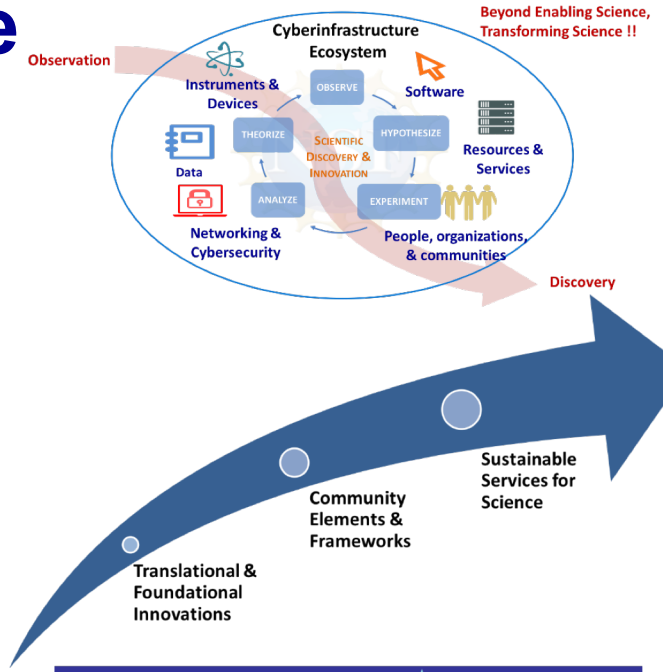

**Instrument, Observatories, Experimental Facilities**

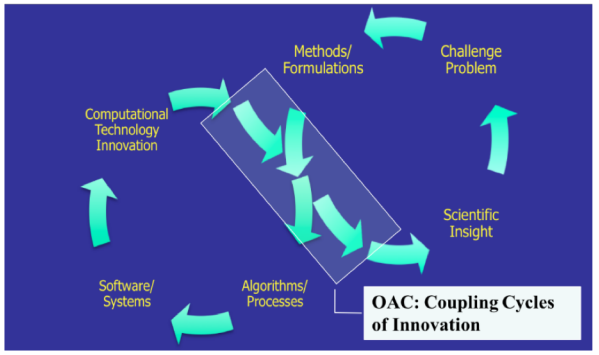


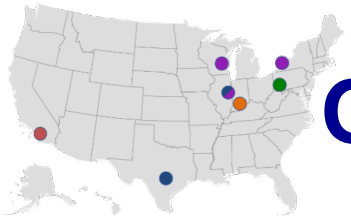

neon
OOI
OCEAN OBSERVATORIES INITIATIVE
ARRAY OF THINGS
LSST
Large Synoptic Survey Telescope
INTERNET2

# 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





# OAC-Funded Computing Ecosystem

**Leadership Computing**

- Blue Waters/UIUC
- Leadership HPC Planning
- Phase 1: 2 to 3x Time-to-solution Improvement
- Phase 2: > 20x
- LC Phase 2 System

**Innovative HPC**

- Stampede / UT Austin
- Stampede 2 / UT Austin
- Comet / UCSD
- Wrangler / UT Austin
- Bridges / CMU & PSC
- Jetstream / Indiana U

**Key**

- Large-scale computation
- Long-tail and high-throughput
- Data Intensive
- Cloud
- Services

**Services**

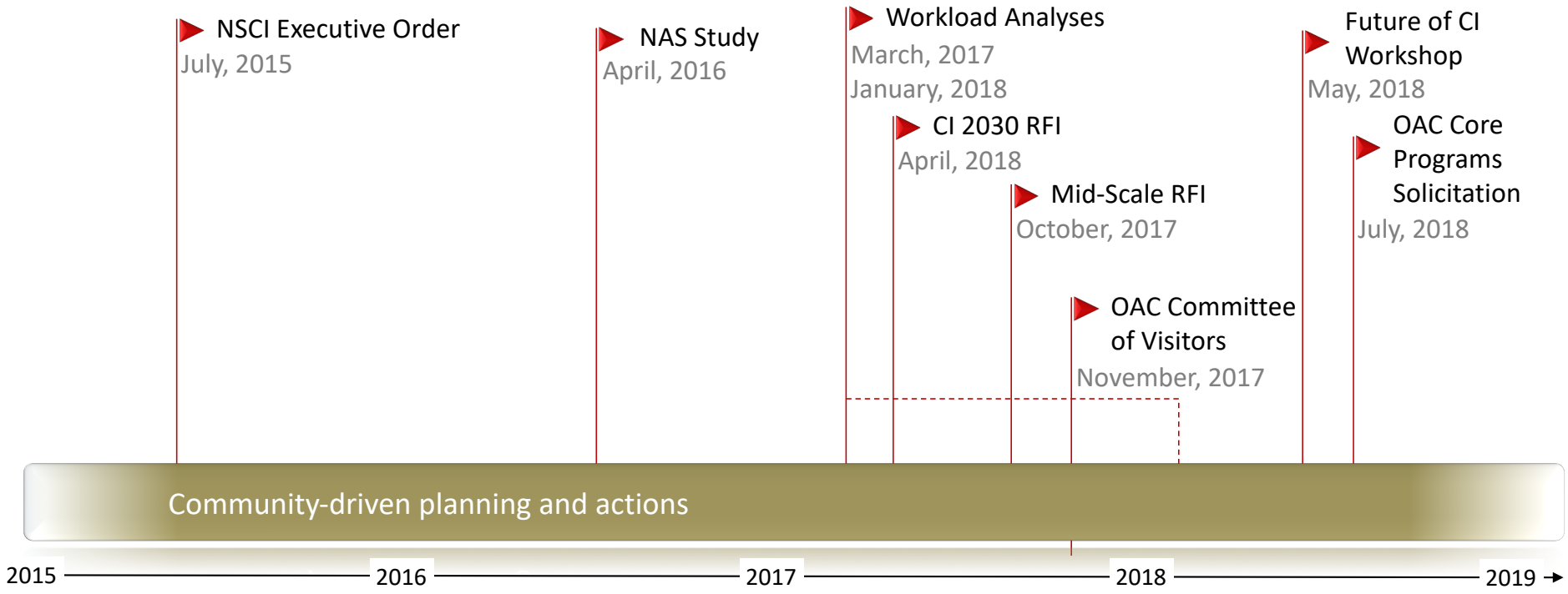
- XSEDE
- XSEDE 2 Coordinated User Services, Education, Outreach
- XD Metrics Service
- Open Science Grid

**Planning for the Future CI Ecosystem**

- Community-driven planning and actions

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

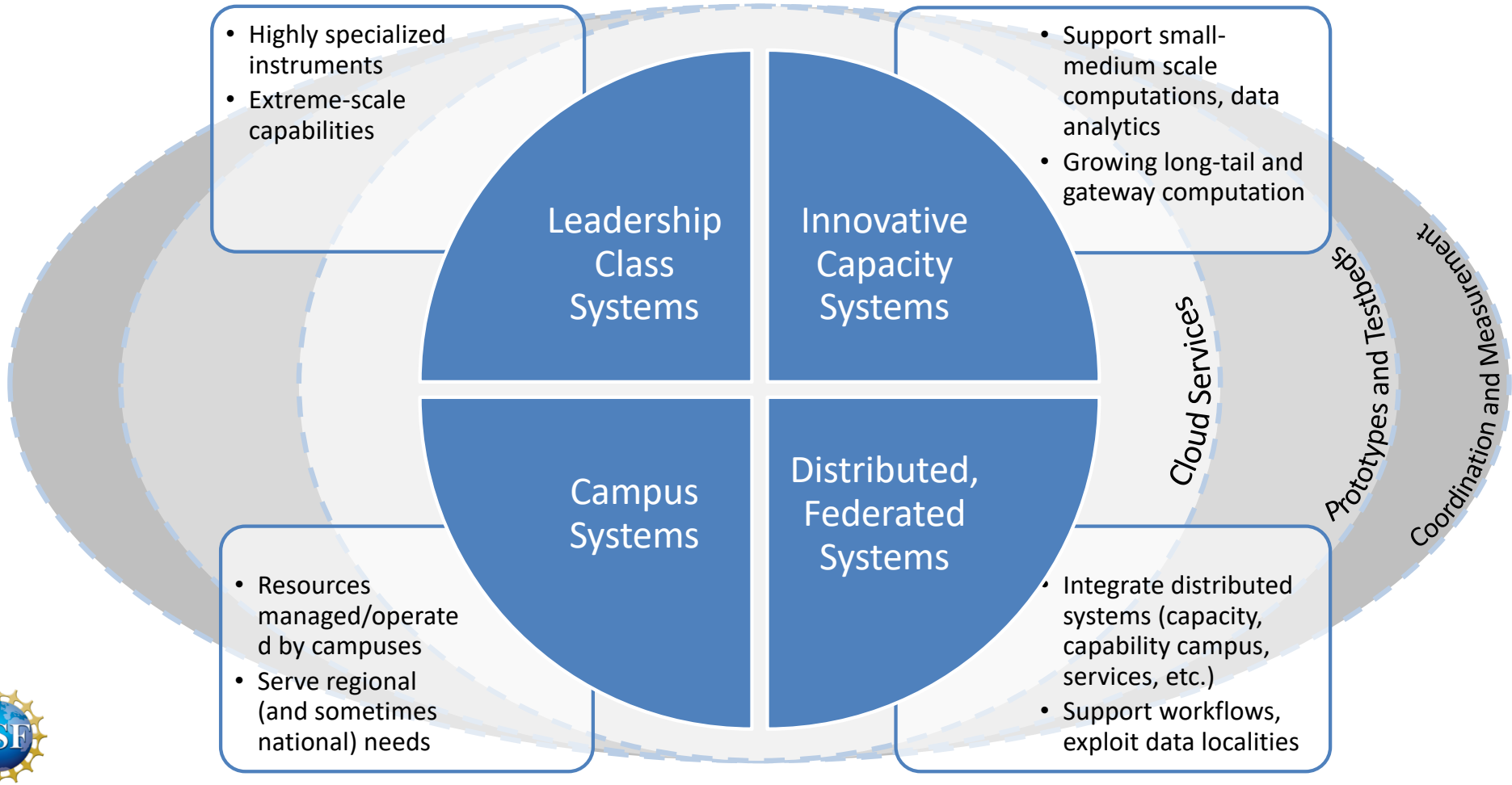
# Planning for the Future CI Ecosystem




**Key Drivers**

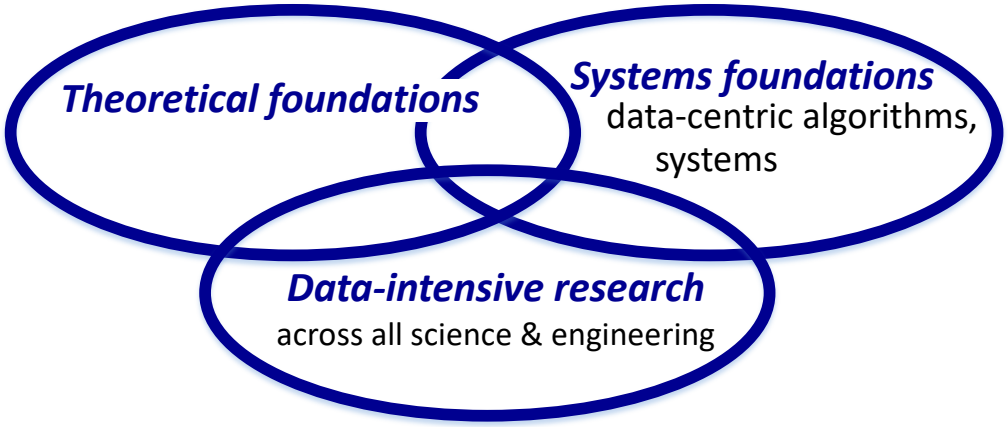
- Changing application landscape & workload profile
- Changing technology, services landscape
- Increasing availability of (exp., obs.) data
- Growing role of ML, data-driven approaches

# Elements of a Balanced CI Ecosystem*



# Harnessing the Data Revolution (HDR)

**Research** across all NSF Directorates




## Educational pathways

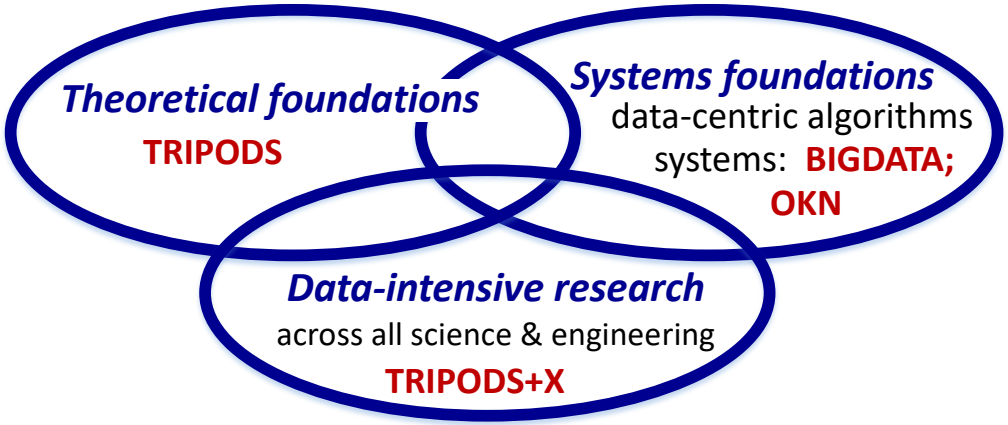
Innovations grounded in an education-research-based framework

## Advanced cyberinfrastructure

Accelerating data-intensive research.

# Harnessing the Data Revolution (HDR)

## Research across all NSF Directorates




## 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)**

OAC-Core

# OAC Core Research Program

**SOLICITATION NSF 18-567**

- ***Translational research*** (spanning design to practice) in all aspects of advanced cyberinfrastructure (CI) to transform science and engineering research
  - Multi-disciplinary, extreme-scale, driven by science and engineering research, end-to-end, or deployable as robust research CI
- ***Research Areas***
  - *Architecture and middleware for extreme-scale systems*
  - *Scalable Algorithms and Applications*
  - *Advanced Cyberinfrastructure Ecosystem*
- ***Research Communities:*** Multiple disciplinary areas supported spanning Computer as well as Computational and Data-driven Science and Engineering

- Part of CISE's coordinated core program solicitations
  - Only Small proposals in FY'19
  - Funding amount $7.5M
  - Max $500K/award
- Proposals due Nov 15, 2018
  - PI's *strongly encouraged* to send 1-page project summary for further guidance.
  - Webinar in July/Aug

# 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