

Funder perspective: NSF Data Related Activities and Priorities



Dr. Shree Mishra

Program Director
Division of Atmospheric and
Geospace Sciences
Geosciences Directorate

August 8th, 2018



Overview

1. Cyberinfrastructure (CI) in the context of NSF Geosciences
2. NSF 10 big ideas and NSF initiatives
3. HDR



NSF Geoscience Directorate

Dr. Bill Easterling
Office of the
Assistant
Director



Atmospheric & Geospace Sciences (AGS)

- Atmosphere
- Geospace
- NCAR/Facilities

Earth Sciences (EAR)

- Deep Earth Processes
- Surface Earth Processes
- Infrastructure & Facilities

Ocean Sciences (OCE)

- Ocean Sciences
- Marine Geosciences
- Integrative Programs

Polar Programs (PLR)

- Antarctic Sciences
- Arctic Sciences
- Antarctic Infrastructure & Logistics
- Polar Environment, Safety, & Health

Integrative & Collaborative Education & Research (ICER)

- INFEWS, PREEVENTS, CIF21 & other Cross-Foundation Programs
- Education & Diversity
- International Collaborations & Partnerships

Home to cross-cutting Geo-CI projects such as EarthCube, CSSI, etc.



NSF GEO Cyberinfrastructure (GEO CI) Working Group

The GEO CI group coordinates the activities of the four GEO divisions (EAR, AGS, OCE, and OPP) and the CISE Office of Advanced Cyberinfrastructure (OAC) to support computational and data-driven infrastructure, training, and applications for enabling potentially transformative geoscience research.

EAR: Eva Zanzerkia and Dena Smith

AGS: Shree Mishra and Irfan Azeem

OCE: Mike Sieracki and Mete Uz

OPP: Marc Stieglitz

CISE/OAC: Amy Walton

AAAS S&T Policy Fellow: Raleigh Martin



Changing Expectations for Data Management

New federal data management policies, such as the National Science Foundation Public Access Plan (NSF 15-52), are emerging at federal agencies.

- Many scientific journals have new data archiving and citation policies
- Open scientific data sharing is increasingly expected by scientific communities
- Ensuring the open availability of data, however, involves overcoming various challenges:
 - Scientific resources must be collected and documented
 - Repository services must be supported and maintained
 - Governance, including legal issues relating to copyright and resource ownership, must be established
- AGU's "Enabling FAIR Data" project, which is coordinating stakeholders to address geoscience data sharing challenges



Data Management Challenges

- Data Policies - vary greatly across the Geoscience Directorate
- Long standing resources from large observation systems, community science
 - timeline; uniqueness of data
- Diversity of data types cause challenges
 - Model outputs, software
 - Physical samples
 - Discrete data in distributed curation
 - Continuous, real-time data streams
- Three groups that require different solutions – large facilities/field campaigns; individual PIs; community science/citizen science (public participation)



CI 2030 Request for Information (RFI)

NSF 17-031

Dear Colleague Letter: Request for Information on Future Needs for Advanced Cyberinfrastructure to Support Science and Engineering Research (NSF CI 2030)

January 5, 2017

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.



CI 2030 RFI (cont'd)

- *GEO received 14% of total response*
- **Responses of primary GEO interest:** 18 (AGS: 3, EAR: 7, OCE: 8, OPP: 0)
- **Authors:** 72, ~50% from univs., ~50% orgs. & agency labs
- **Key concerns:**
 - code optimization for earth-system models
 - big data wrangling
 - dealing with I/O inefficiencies
 - increased bandwidth for remote operations
 - improved organizational efforts
 - workforce development



NSF 10 Big Ideas

RESEARCH:

- Harnessing the Data Revolution for 21st Century **HDR**
- The Future of Work at the Human Technology Frontier **FW-HTF**
- Windows on the Universe (nature of matter and energy) **WoU**
- The Quantum Leap: Leading the next quantum Revolution **QL**
- Understanding the Rules of Life **URoL**
- Navigating the New Arctic **NNA**

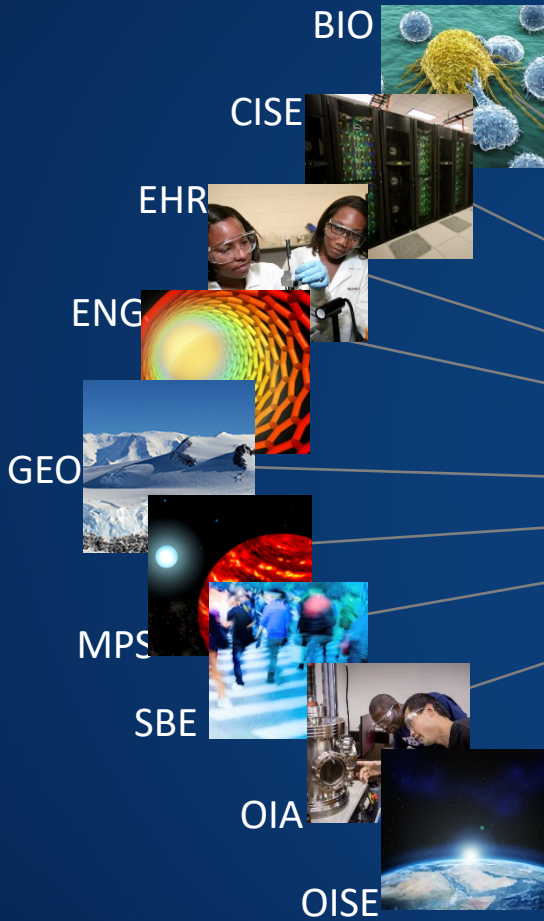
PROCESS BIG IDEAS

- NSF INCLUDES
- Growing Convergence Research
- Mid-scale Research Infrastructure (\$4 M through \$70 M)
- NSF 2026 Fund



Directorates Offices

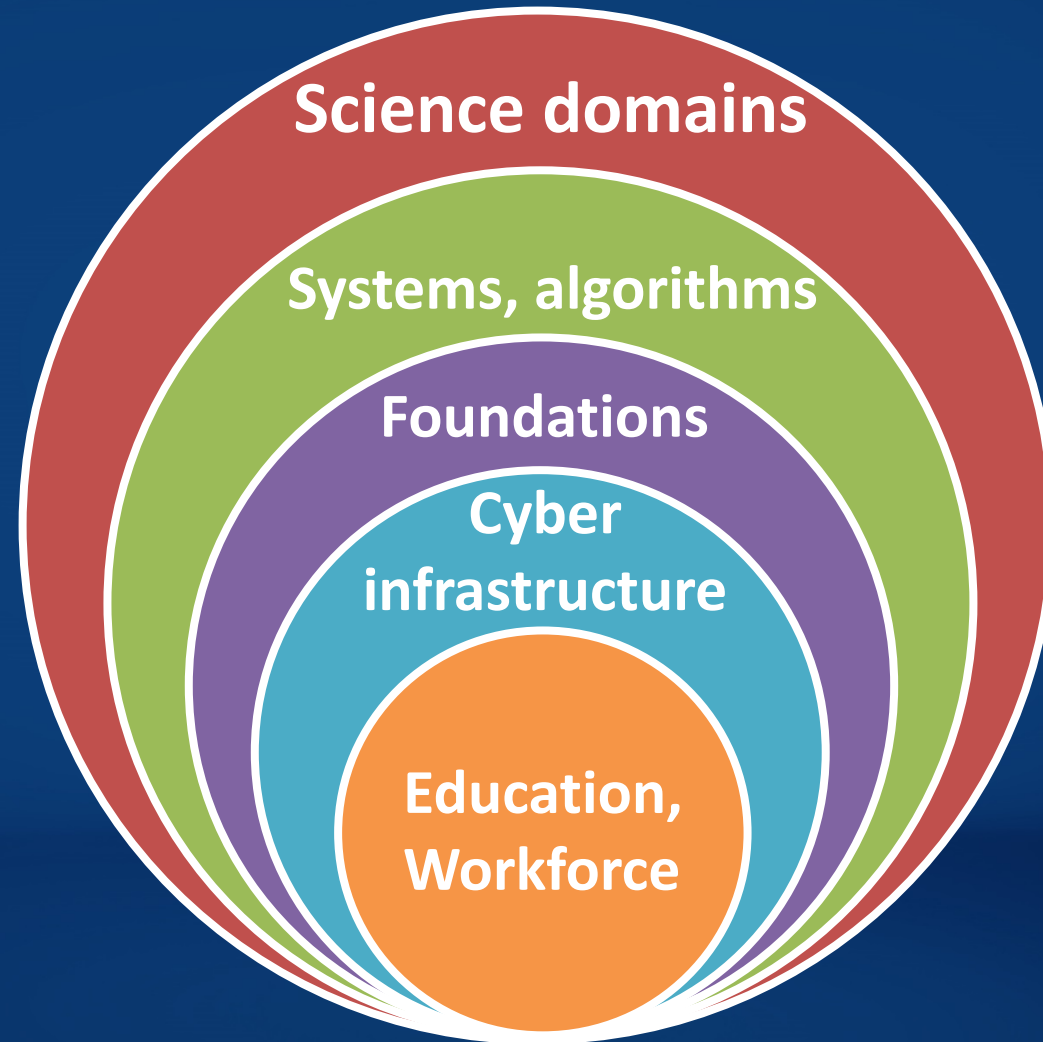
Big Ideas



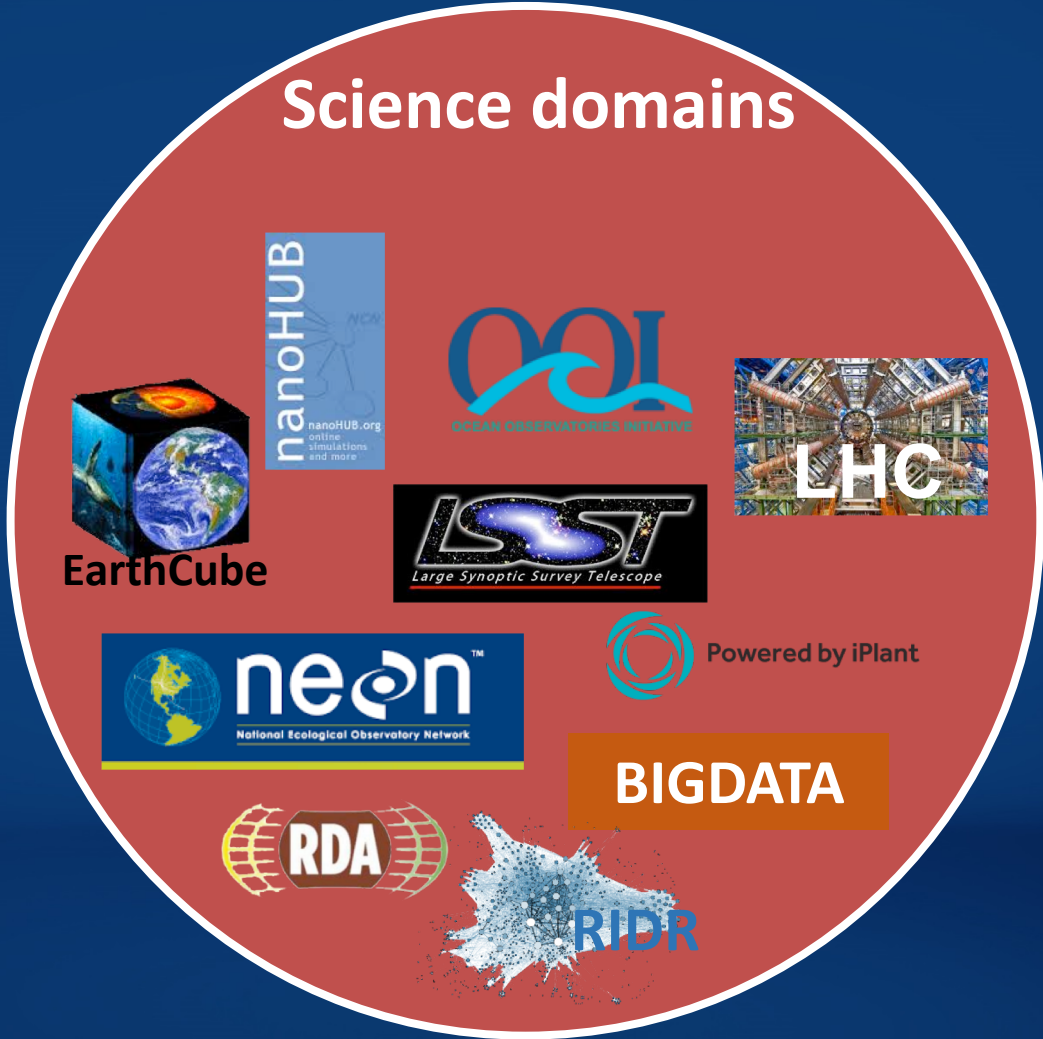
*“Engage NSF’s research community in the pursuit of **fundamental research in data science and engineering**, the development of a cohesive, federated, national-scale approach to **research data infrastructure**, and the development of a **21st-century data-capable workforce**.”*



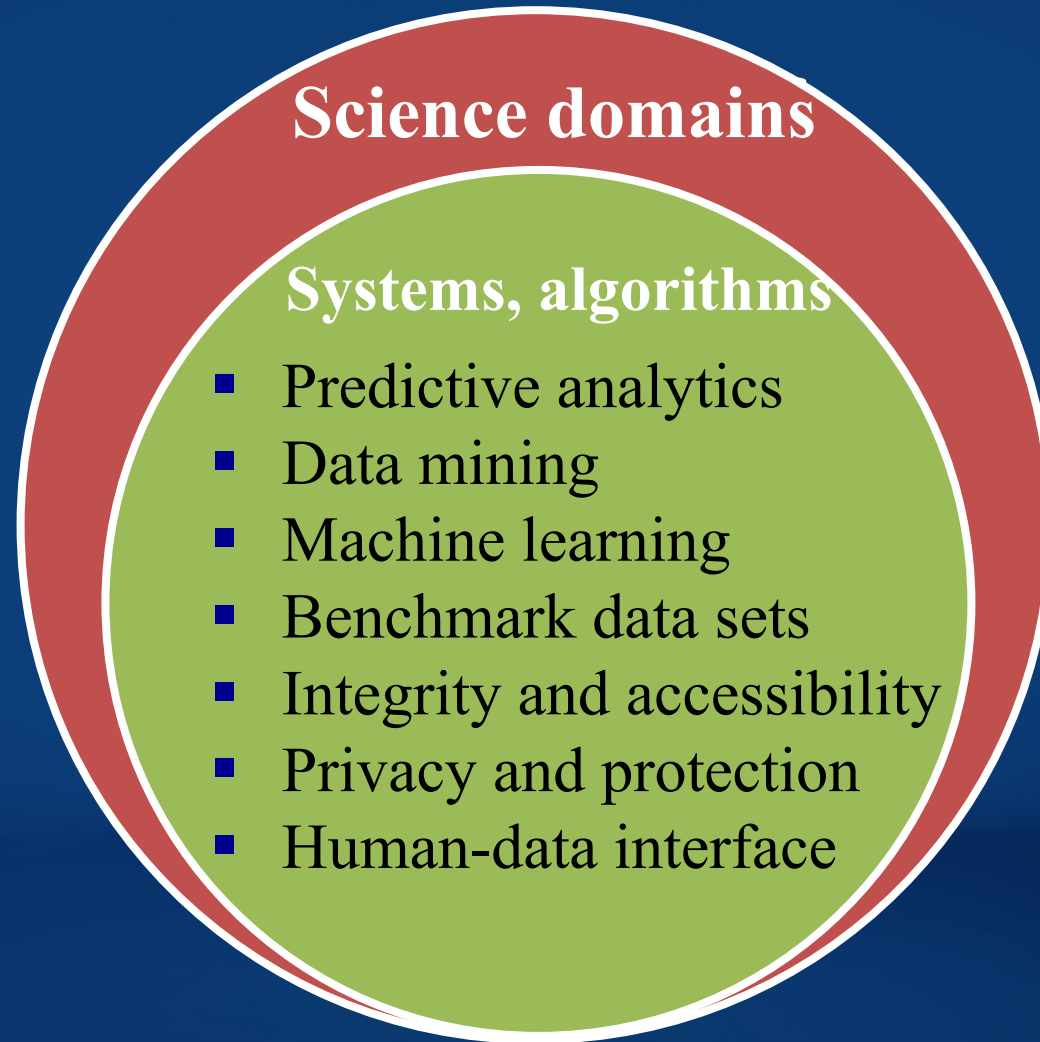
Harnessing the Data Revolution: five themes



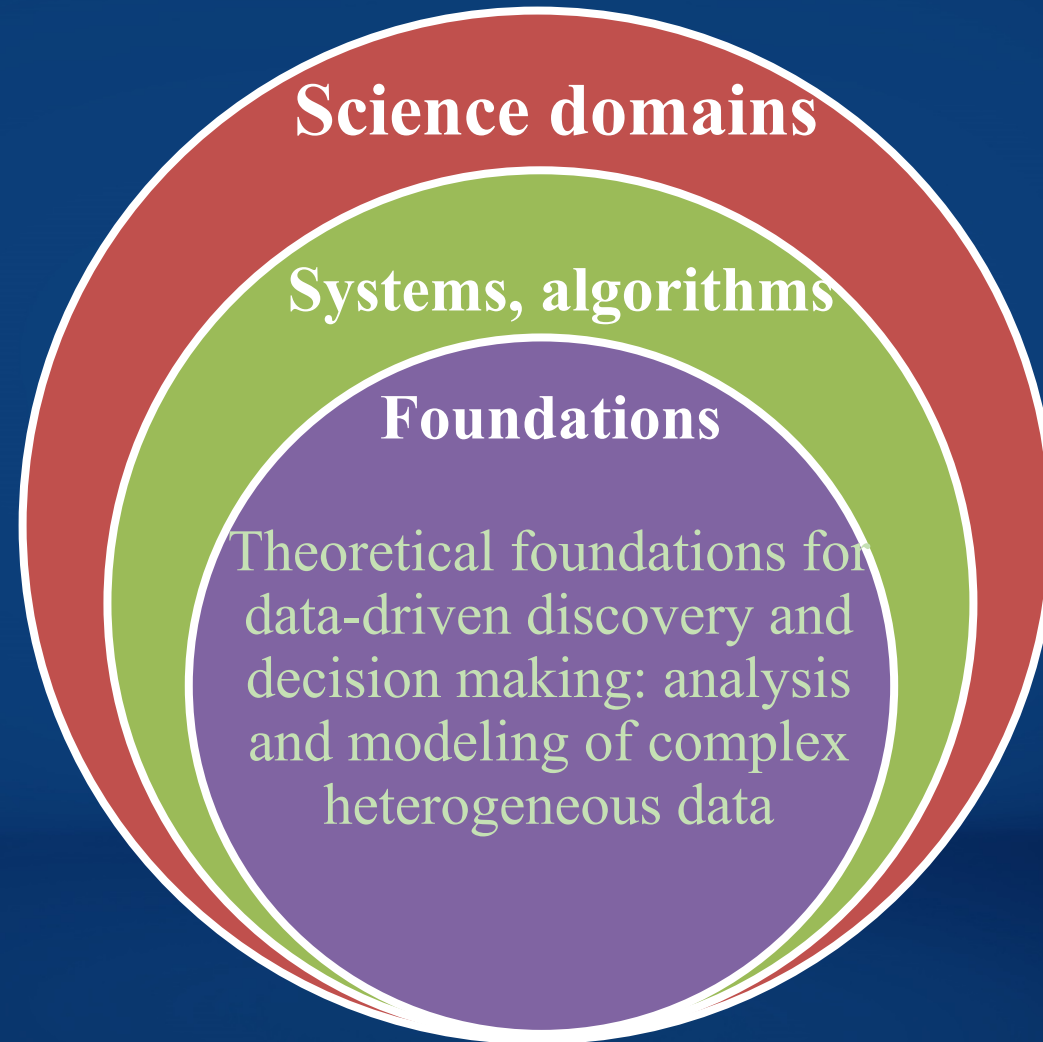
Harnessing the Data Revolution: five themes



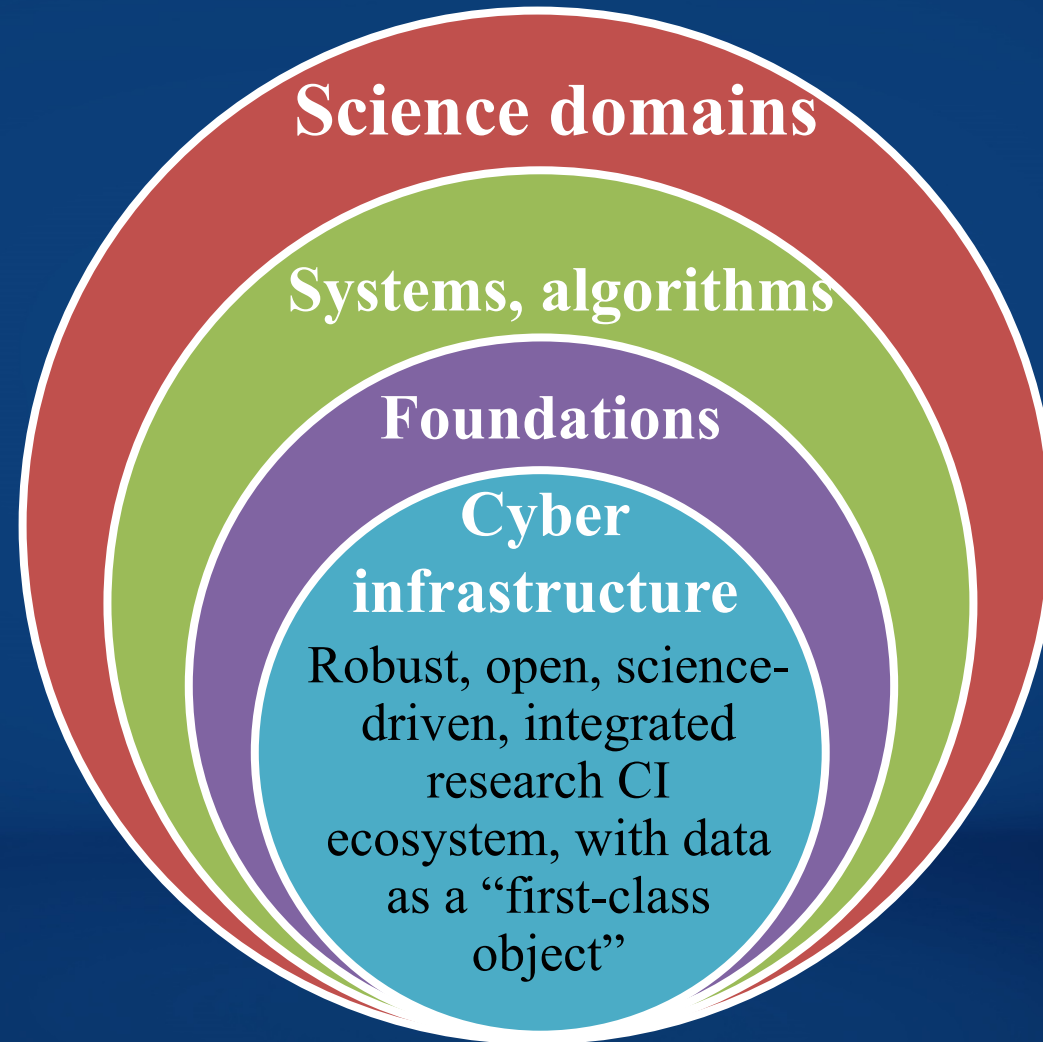
Harnessing the Data Revolution: five themes



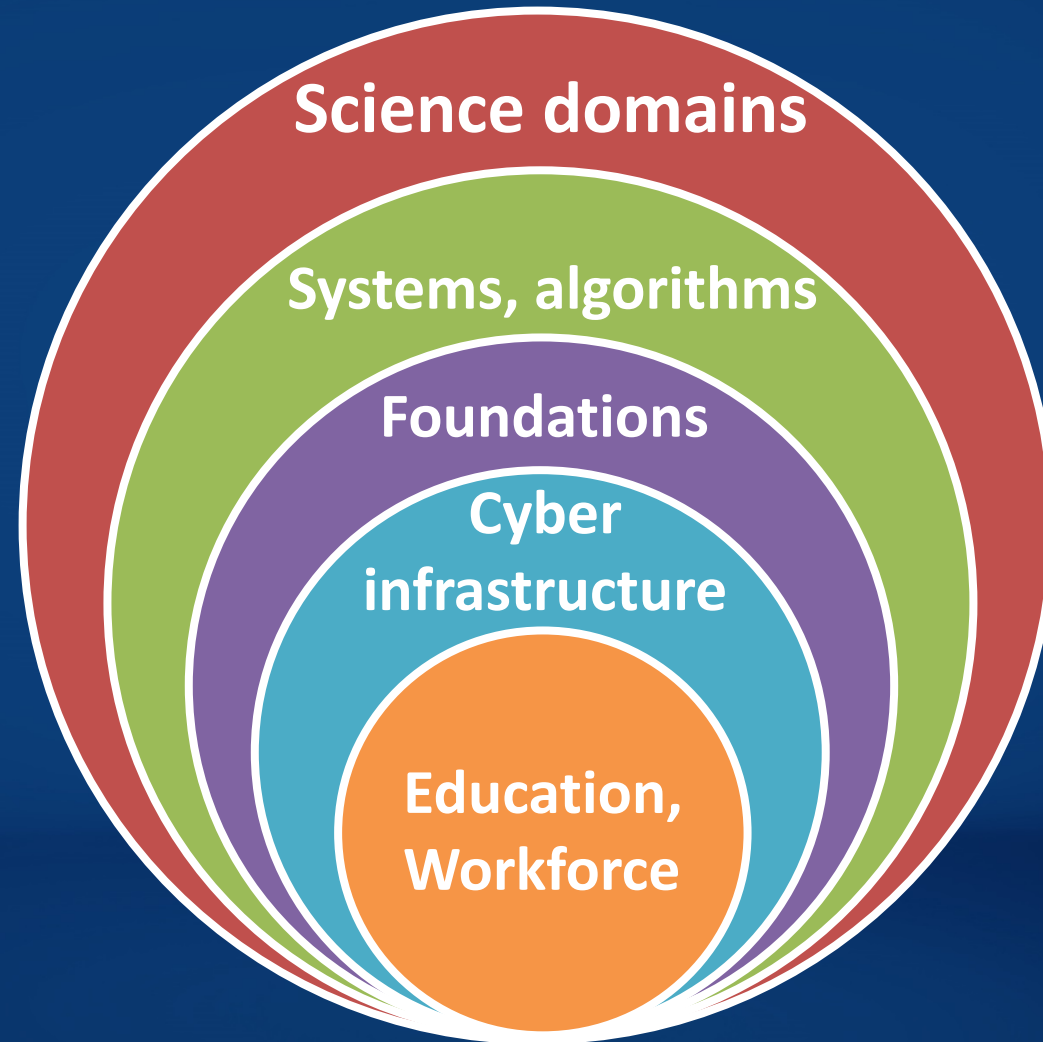
Harnessing the Data Revolution: five themes



Harnessing the Data Revolution: five themes



Harnessing the Data Revolution: five themes



HDR Education and Workforce

Education,
Workforce,
Outreach

- *Envisioning the Data Science Discipline: The Undergraduate Perspective, NASEM study, workshops*
 - Intermin report: <https://www.nap.edu/catalog/24886/>, Final report: 5/18
- NSF Workshop on *Keeping Data Science Broad: Negotiating the Digital and Data Divide*, Oct, 2017
 - Part of *Keeping Data Science Broad* series, NSF South Big Data Hub



FY 19 Budget Request

FY 19 Budget Request: \$ 7.47 B

FY 18 Appropriation: \$ 7.8 B

NSF's 10 BIG IDEAS FY 2019 REQUEST FUNDING

(Dollars in Millions)

Big Ideas	FY 2019 Request
Research Ideas	\$180.00
Harnessing the Data Revolution for 21st- Century Science and Engineering - HDR (CISE/ITR) ¹	30.00
Navigating the New Arctic - NNA (GEO/ICER)	30.00
The Future of Work at the Human-Technology Frontier - FW-HTF (ENG/EFMA) ¹	30.00
The Quantum Leap - QL (MPS/OMA)	30.00
Understanding the Rules of Life - URoL (BIO/EF)	30.00
Windows on the Universe - WoU (MPS/OMA)	30.00
Process Ideas	\$102.50
Growing Convergence Research - GCR (IA)	16.00
Inclusion across the Nation of Communities of Learners of Underrepresented Discoverers in Engineering and Science - NSF INCLUDES (EHR)	20.00
Mid-Scale Research Infrastructure (IA)	60.00
NSF 2026 Fund (IA)	6.50
Total, NSF Big Ideas	\$282.50

¹Convergence Accelerator funding will also support the Big Ideas HDR and FW-HTF in the amount of \$30 million for each, in addition to the amounts above. The Convergence Accelerator funding will be managed by IA, and the Research Ideas funding will be managed by CISE and ENG, respectively, as shown above. For more information on Convergence Accelerators, refer to the Agency Reform section of the Overview chapter. For more information on NSF's Big Ideas, refer to the Big Ideas section of the Overview chapter.



Takeaways

- Community input is extremely valuable to NSF
- NSF seeks community input through DCLs, RFIs, conference townhalls, workshops such as GeoDaRRS
- Please participate whenever possible and tell us how NSF can help advance science



Questions?

Feel free to email me at: sumishra@nsf.gov

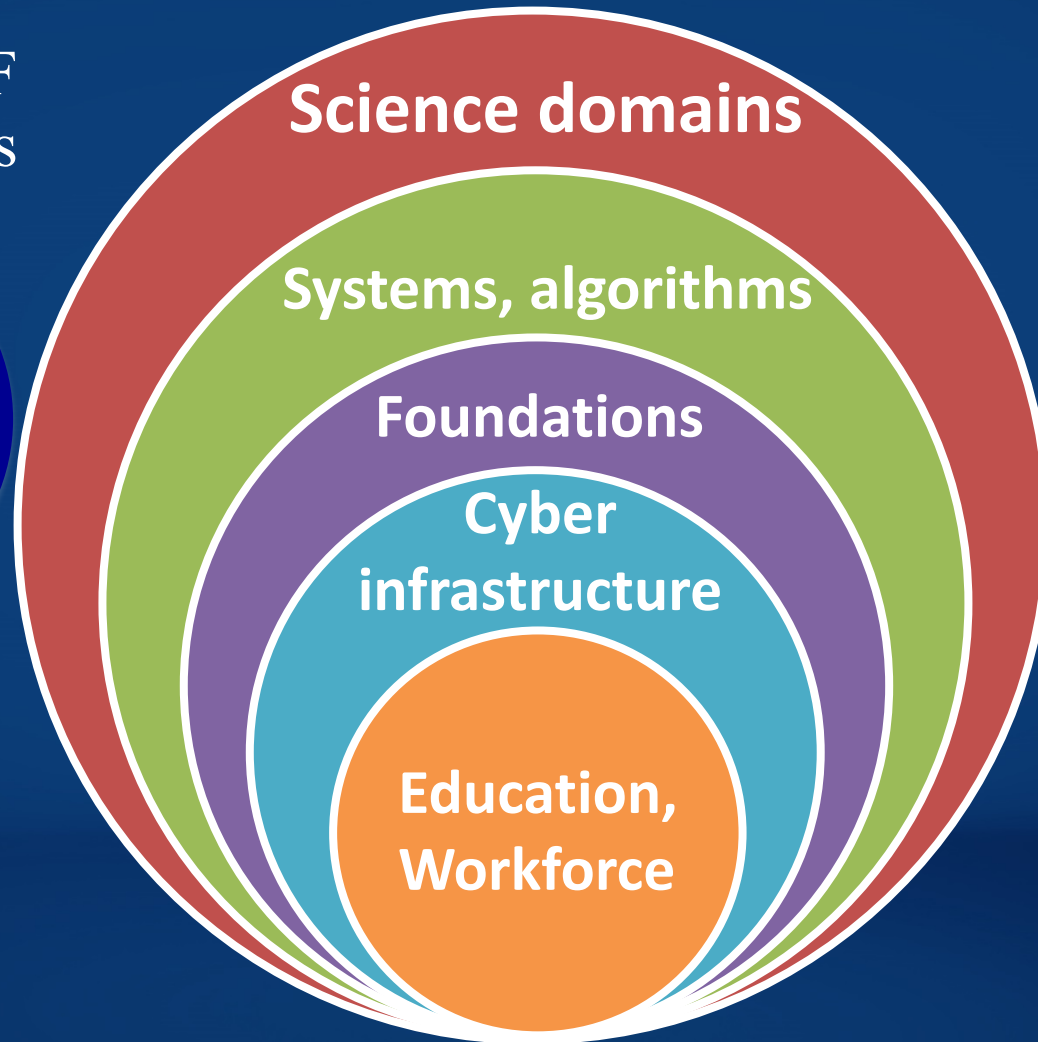
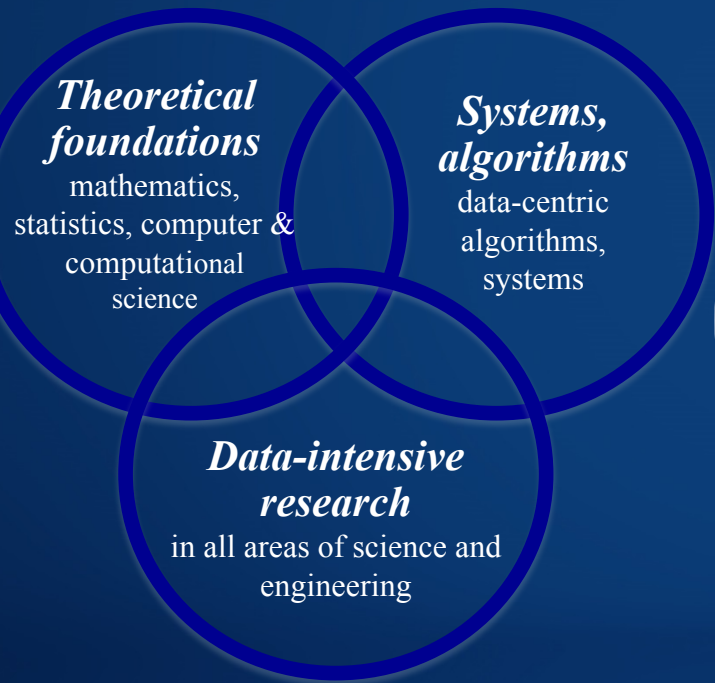


<https://www.nsf.gov>



Harnessing the Data Revolution: five themes

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 Team

HDR Steering committee

Jim Deshler, DDD, BIO/DBI
Chaitan Baru, co-chair, CISE/OAD
Robin Wright, DD, EHR/DUE
Fil Bartoli, DD. ENG/ECCS
Anjuli Bamzai, SH, GEO/AGS
Nandini Kannan, PD, MPS/DMS
Bogdan Mihaila, co-chair, MPS/OAD
Dan Sui, DD, SBE/SES

HDR Working Group Members

BIO: Karen Cone, Peter McCartney
CISE: Rajiv Ramnath, Robert Chaddock, Amy Walton
EHR: John Cherniavsky, Eamonn Kelly
ENG: Alexis Lewis, Akbar Sayeed
GEO: Eva Zanzerkia
MPS: Angela Wilson, John Schlueter, Daryl Hess
OIA: Paul Morris
OISE: Charles Estabrook
SBE: Cheryl Eavey, Dan Sui
HDR Executive Secretary: Vandana Janeja, CISE AAAS Fellow

HDR AD leads

Anne Kinney, MPS
Jim Kurose, CISE

AAAS Fellows Dominique Carter (OISE), Raleigh Martin (GEO),
Mateo Munoz (NSB), Michael Wallace (ENG), Lucy Erickson (CISE),
Tyler Kloefkorn (CISE)



EXTRAS

- NCAR Data repositories
 - Research Data Archive (RDA): <https://rda.ucar.edu>
 - EOL data archives
 - Climate data archives
 - COSMIC

