NSF Big Data Programs and Activities: Harnessing the Data Revolution

Chaitan Baru
Senior Advisor for Data Science, CISE
National Science Foundation

*"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**."*

NSF's Big Data / Data Science Programs

Policy

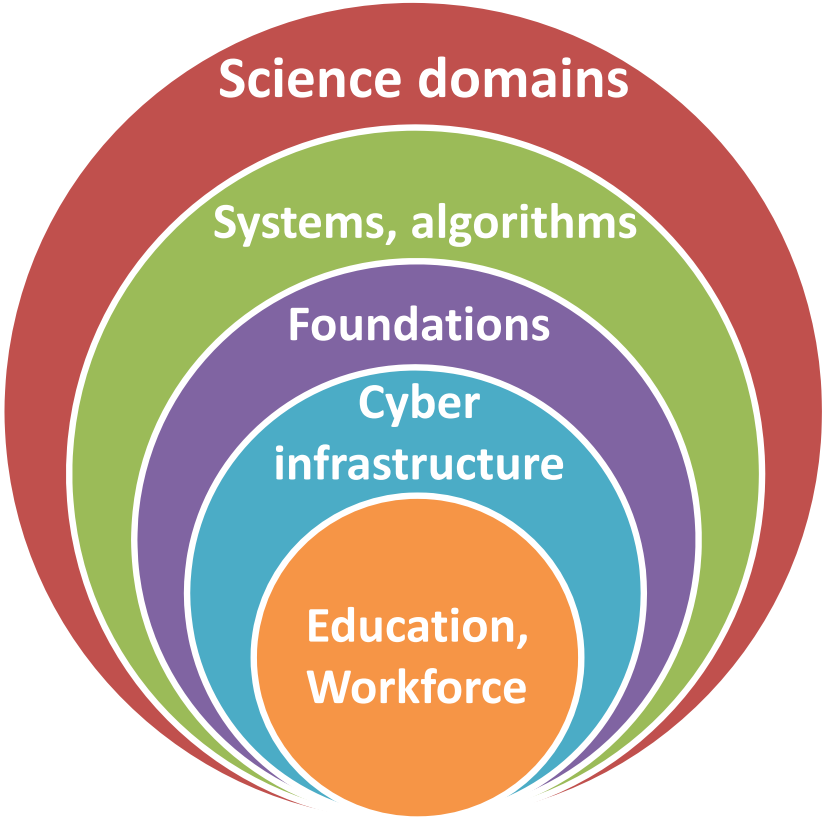
NSF BIGDATA Program: An evolving research program

- **2012: First year of program. NSF + NIH.**
- **2014: NIH launches BD2K. NSF Proposal Categories:**
  - DKA: Data and Knowledge Analytics
  - DKM: Data and Knowledge Management
  - CSD: Computational Scientific Discovery
  - IA: Innovative Applications
- **2015: Categories**
  - Foundations : Innovative Applications : Combined
    - 63% : 20% : 17%

NSF BIGDATA…

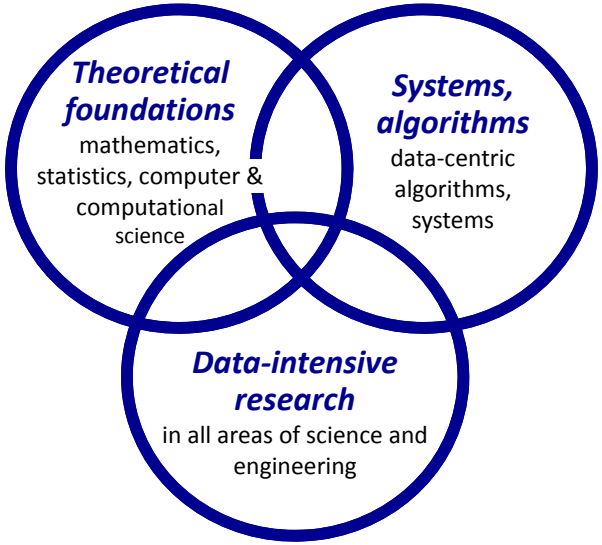
- **2016: Foundations : Innovative Applications**
  - 52% : 48%
- **2017: Partnership with cloud vendors**
  - AWS, Google, Microsoft (upto $3M each. Total: upto $9M)
  - 70:30 split in funding, NSF funds : Cloud resources
  - Minimum cloud resources: $100K
  - Maximum cloud resources: ~$860K (for a $2M proposal)
  - Proposals currently under review

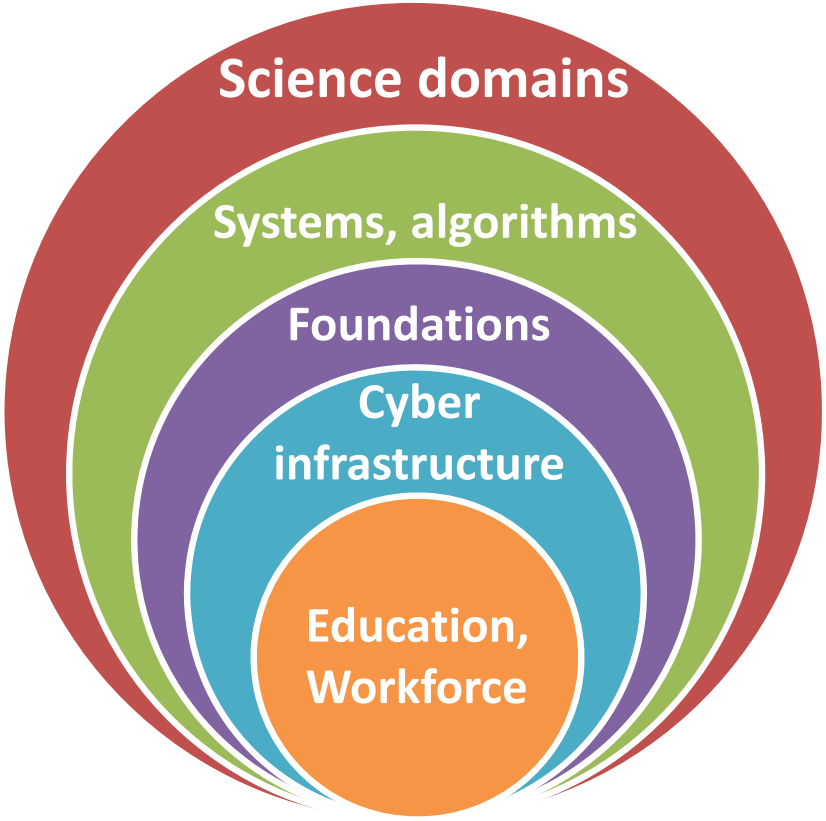
Harnessing the Data Revolution: five themes



Harnessing the Data Revolution: five themes

Research across all NSF Directorates





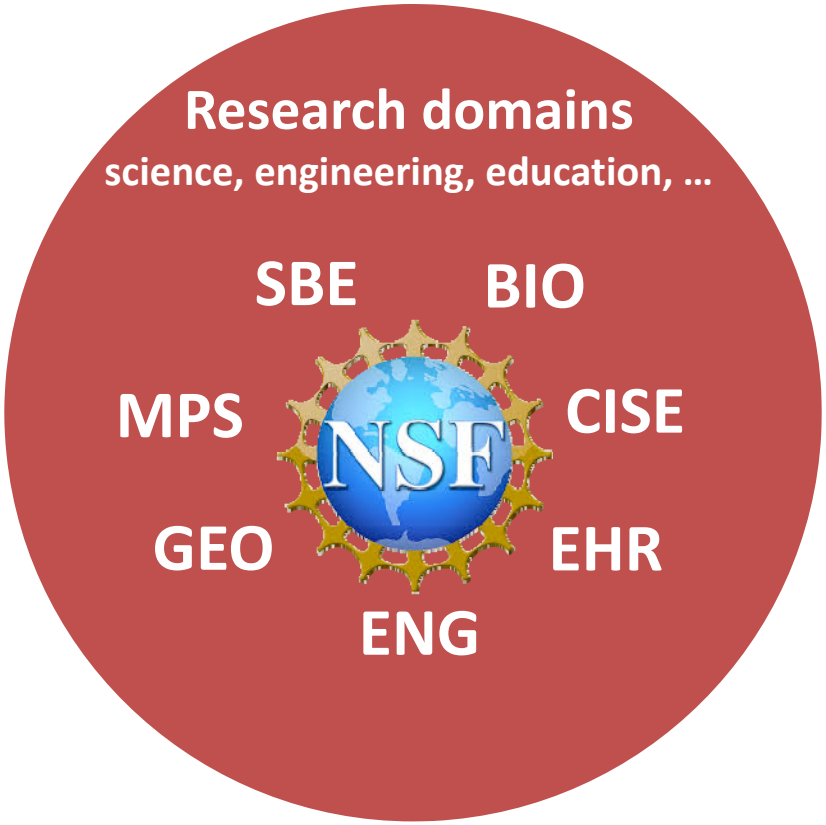
Educational pathways

Innovations grounded in an education-research-based framework

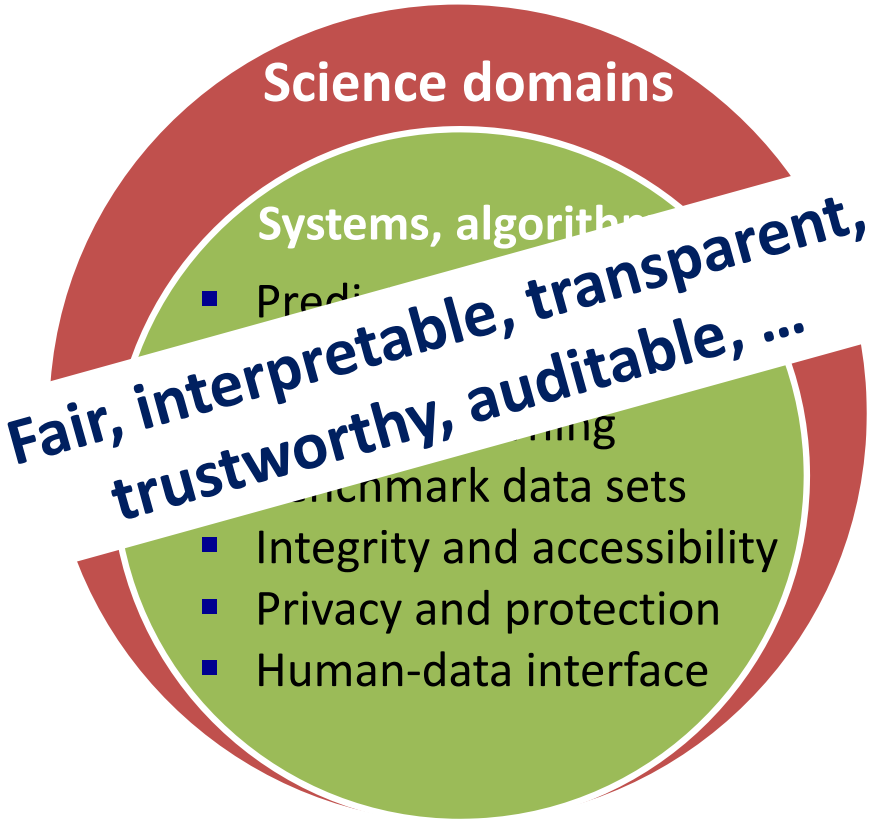
Advanced cyberinfrastructure

Accelerating data-intensive research

Harnessing the Data Revolution: Domains

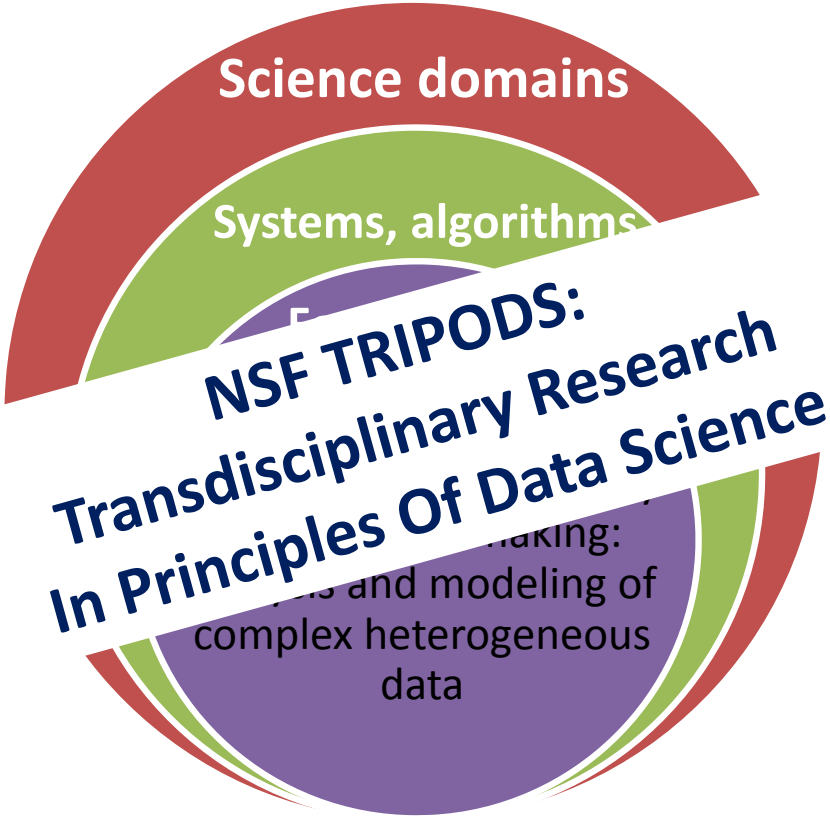


Harnessing the Data Revolution: Systems

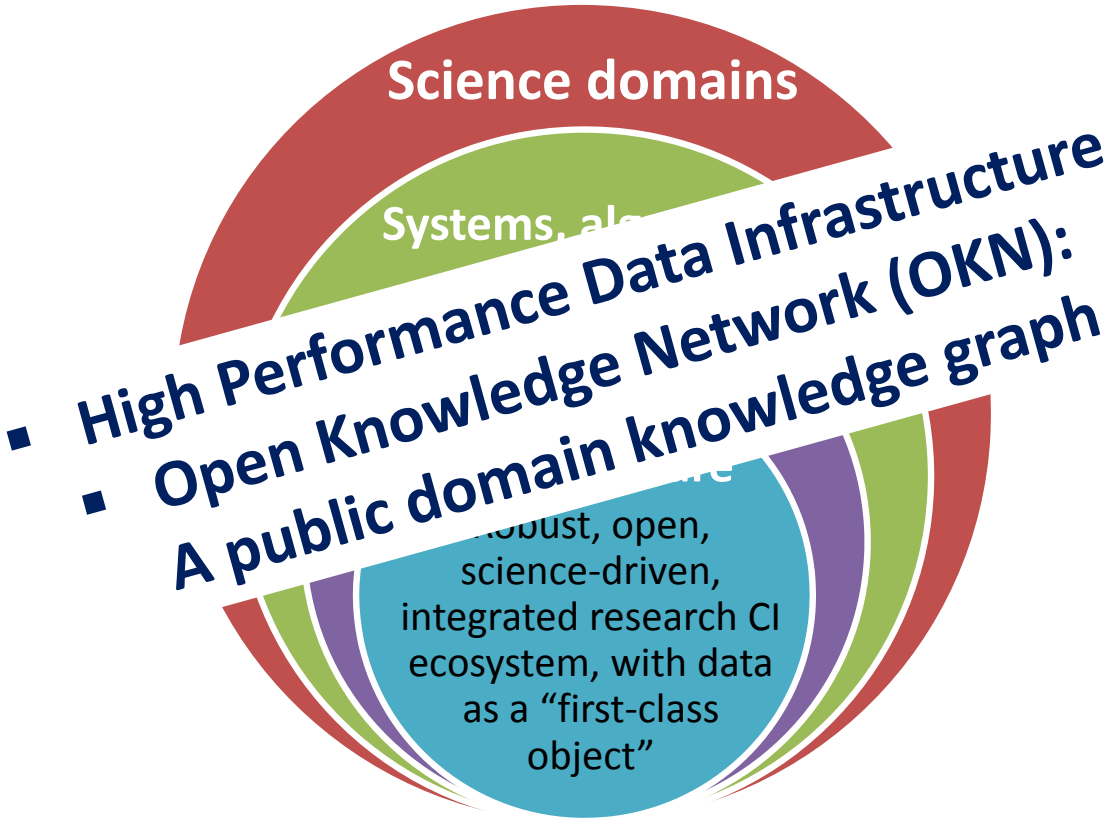


Harnessing the Data Revolution: Foundations

- Requires close collaboration among CS, Math, Stats
- Phase I: 3 years, ~10 "Proto centers"
- Phase II: ~3 Large, national centers. Connections with applications domains



Harnessing the Data Revolution: Cyberinfrastructure



- **High Performance Data Infrastructure**
- **Open Knowledge Network (OKN):**
- **A public domain knowledge graph**

OKN: Semantic information infrastructure

OKN: An open web-scale knowledge network

- Semantically-linked concepts, data
  - To foster research on an entire class of new applications leveraging data, context, and inferences from data
- Question/answer interfaces, dialog-based interactions, explanatory/story-telling interfaces
- Joint academia, industry, government workshops
  - July 2016, Washington, DC
  - Feb 2017, Sunnyvale, CA
  - Oct 2017, NLM, Bethesda, MD (planned)

Harnessing the Data Revolution: Education

National Academy of Sciences
Workshops on
Envisioning the Data Science Discipline:
The Undergraduate Perspective

Putting it all Together: Translational Data Science

Development, application of data science in the science and other applications domains

NSF is uniquely positioned to execute on the convergent, full-breadth of HDR activities

Thank You!