+

Panel on Big Data Challenges and Opportunities

Dr. Chaitan Baru

Senior Advisor for Data Science,
Directorate for Computer & Information Science & Engineering

National Science Foundation

NSF's Perspective and Role

The NSF funds basic, curiosity driven research

To promote the progress of science;

to advance the national health, prosperity, and welfare;

to secure the national defense….

National Science Foundation
WHERE DISCOVERIES BEGIN

NSF: Research to Infrastructure Continuum

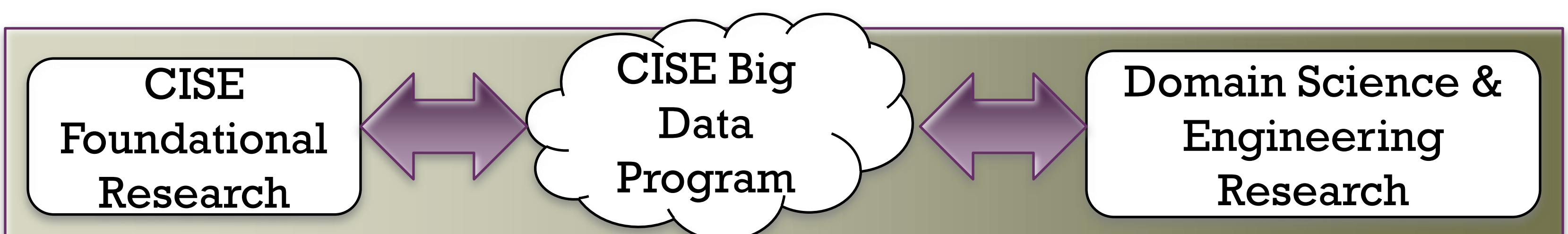
- NSF has programs across the spectrum
  - From foundational research programs to science facilities and cyberinfrastructure
  - ACI – Advanced CyberInfrastructure is a Division in CISE
- And in education and training

Big Data in the CISE Context

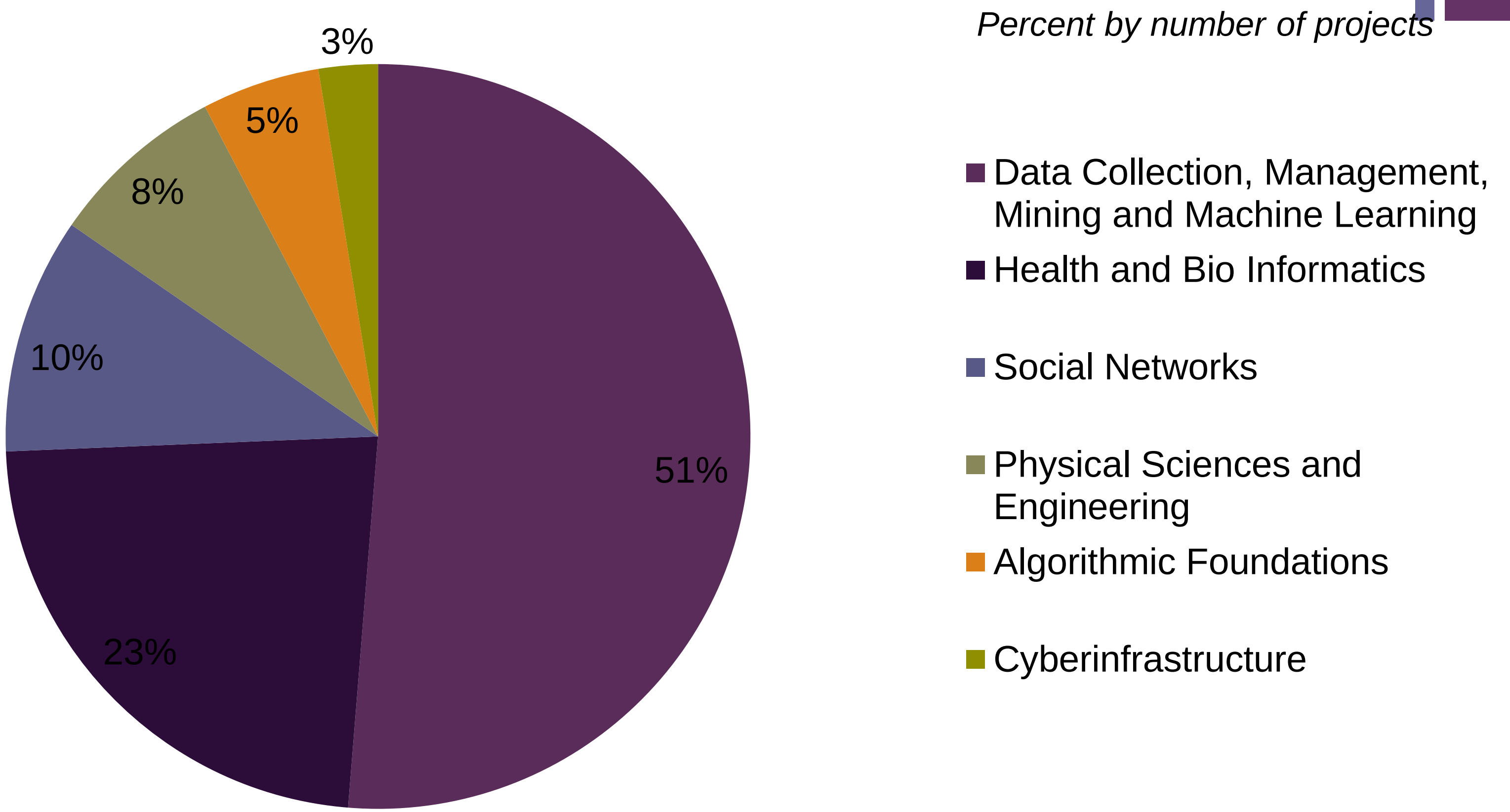
- Big Data is about Data…

  → Data comes from the domains…

    → Big Data research needs to tie intimately to the domains



2013: Core Techniques and Technologies for Advancing Big Data Science & Engineering



In 2013, NSF and NIH awarded 45 projects
ranging from $250K/year for up to 3 years to $1M/year for up to 5 years.

2014: Critical Techniques and Technologies for Advancing Big Data Science and Engineering

- <u>**Two**</u> categories for submission
  - <u>Foundational</u>: Fundamental, novel techniques, theories, methodologies and technologies of broad applicability
  - <u>Innovative Applications</u>: Novel techniques, theories, methodologies, and technologies of interest to at least one specific application

In 2014, NSF awarded ~40 projects
ranging from $300K to $1.5M, for 3-5 years.

Data Infrastructure Building Blocks

- **Two** categories for submission
  - Pilot Demonstration: Addresses broad community needs of interest either to a large number of researchers within a research domain, or extending beyond that to encompass other disciplines
  - Early Implementation: Addresses interest to multiple research communities in multiple scientific and engineering domains

In 2014, NSF awarded 17 projects
ranging from $580K to $4.8M, for 3-5 years.

NSF Research Traineeship (NRT)

- **Purpose:** create and promote new, innovative, effective, and scalable models for STEM graduate student training and prepare scientists and engineers of the future, particularly in emerging STEM fields vital to the nation.

- Addresses NSF priority theme on Data Science

$24M-$30M anticipated total Investment by NSF.
Funding up to $3M, for up to 5 years.

Initiatives underway…need your inputs

- RFI on Big Data Hubs
- RFI on National Big Data R&D Initiative
- Planning Big Data Workshops focusing on new areas, e.g.
  - Big Data and IoT
  - Big Data Reproducibility
  - Data Science Training and Education

CISE RFI: Big Data Regional Innovation Hubs

Accelerating the Big Data Innovation Ecosystem

- Continue and scaleup partnerships established by the Data2Action event and stimulate, track, and help sustain new regional and grassroots partnerships around Big Data.

- Big Data Hubs could, for example:
  - Focus on solutions to specific global and societal challenges by convening stakeholders across sectors to partner in results-driven programs and projects.
  - Act as a matchmaker between the various academic, industry, and community stakeholders to help drive successful pilot programs for emerging Big Data technology.

Big Data Hubs examples…

- Coordinate across multiple regions of the country, based on shared interests and industry sector engagement to enable dialogue and share best practices.
- Aim to increase the speed and volume of technology transfer between universities, public and private research centers and laboratories, large enterprises, and SMB's.
- Facilitate engagement with opinion and thought leaders on the societal impact of Big Data technologies as to maximize positive outcomes of adoption while reducing unwanted consequences.
- Support the education and training of the entire Big Data workforce, from data scientists to managers to data end-users.

➔ Respond by Nov 1 to bigdata@nsf.gov
http://www.nsf.gov/cise/news/2014-bigdata-rfi.jsp

RFI National Big Data R&D Initiative

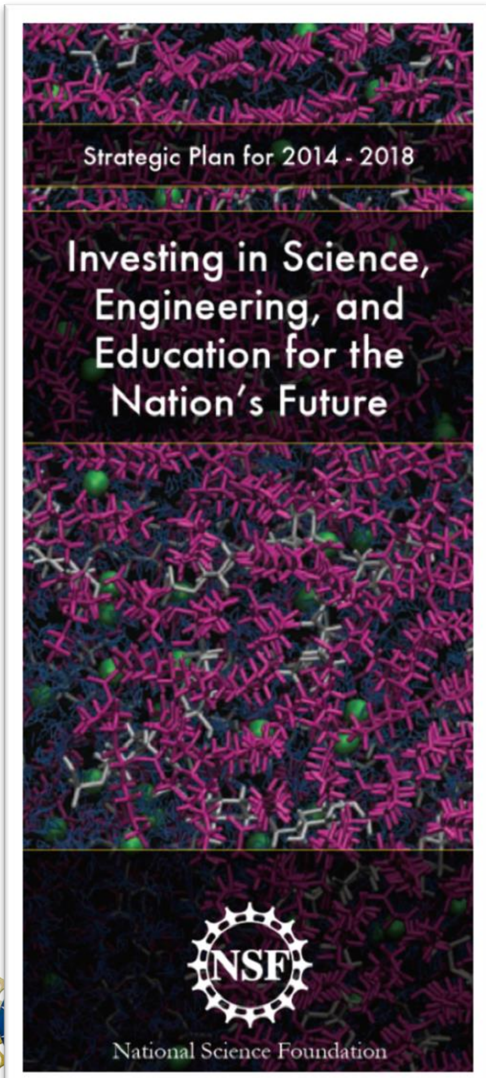
- Request For Input
  - …how we might best develop an overarching, comprehensive framework to support national-scale big data science and engineering research and education, discoveries and innovation."

- https://www.nitrd.gov/bigdata/rfi/02102014.aspx

➔ Respond by Nov 14

+

Increase the Nation's Data Science Capacity – An NSF Priority Goal



- Data science education: Improve the nation's capacity in data science by investing in the development of human capital and infrastructure.
- By September 30, 2015:
  - Implement mechanisms to support the training and workforce development of future data scientists;
  - Increase the number of multi-stakeholder partnerships to address the nation's big-data challenges; and
  - Increase investments in current and future data infrastructure, extending data-intensive science into more research communities.

Thank You!

cbaru@nsf.gov