BIG DATA REGIONAL INNOVATION HUBS & SPOKES

Update on Program Activities

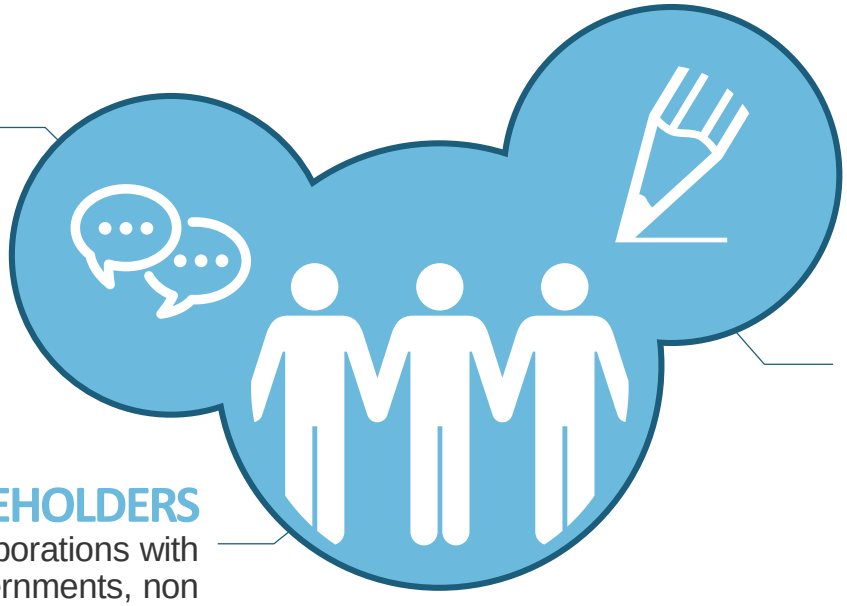
Fen Zhao

March 7, 2017

KEY TAKEAWAYS

01 THE PROGRAM

Brings together domain scientists, computer scientists, and end users to use data to solve challenges

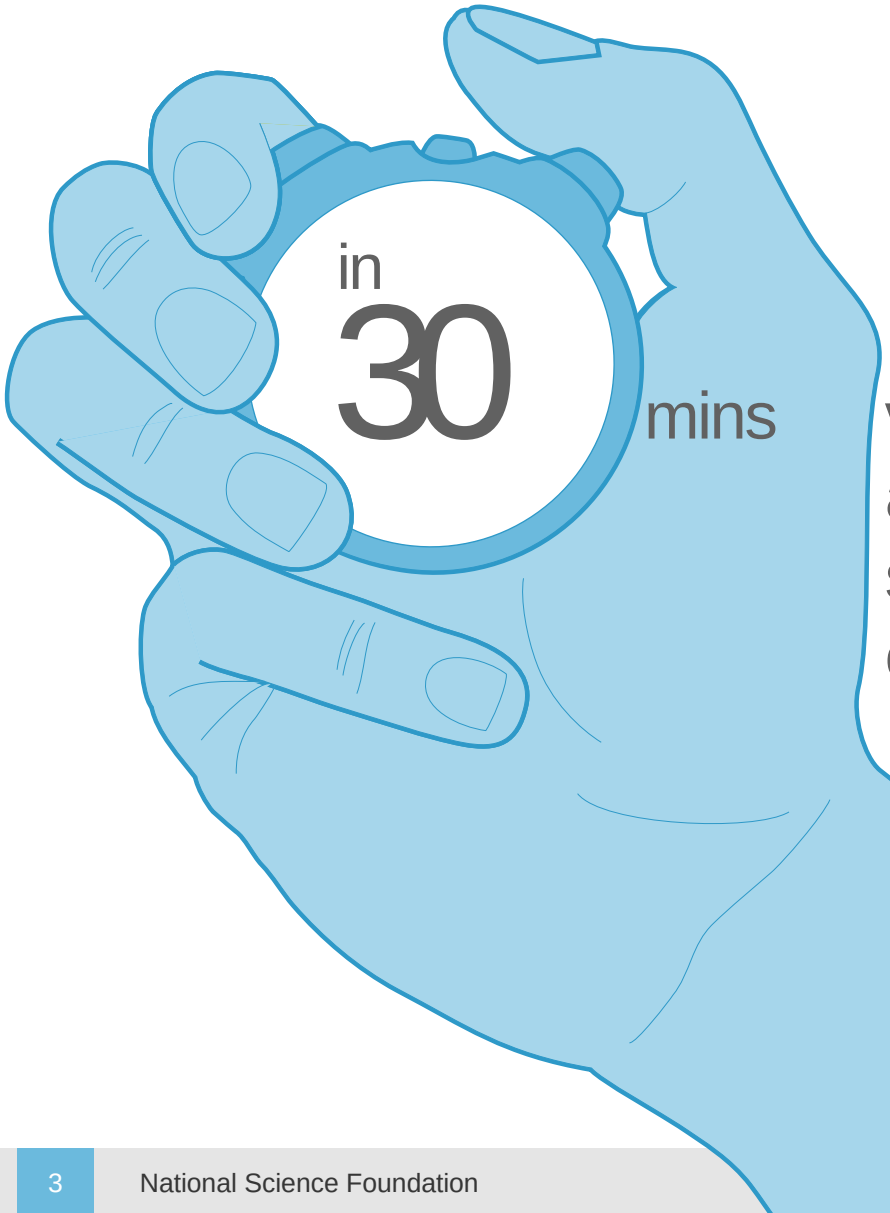


02 THE STAKEHOLDERS

Encourages collaborations with industry, state & local governments, non profits, and others that are not typical NSF participants

03 PARTICIPATION

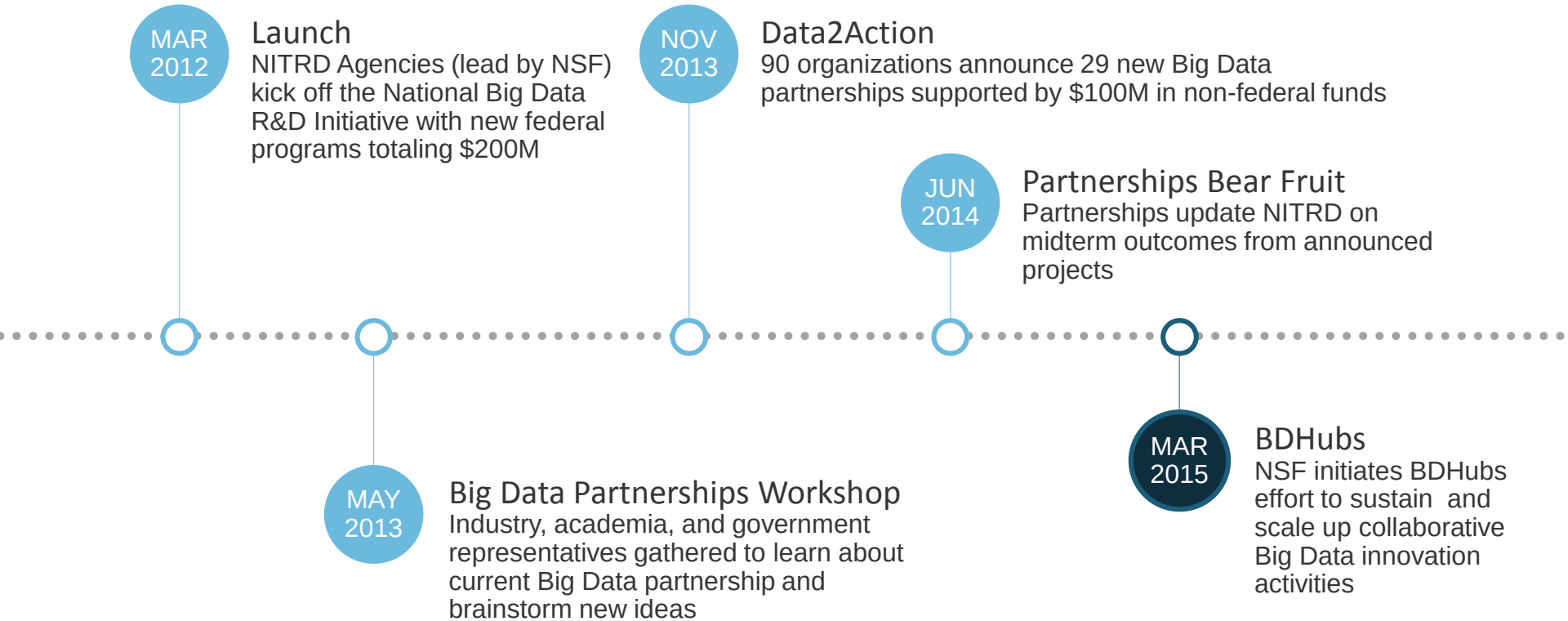
Opportunity for NASA and your communities to get involved!



vision of the BDHubs program
activities of funded Hubs
spokes awarded
opportunities for participation

WHAT IS THE HISTORY BEHIND BDHUBS?

The National Big Data R&D Initiative & Data to Knowledge to Action (Data2Action)



THE HISTORY BEHIND BDSPOKES

BD Spokes is the second phase of a long term NSF agenda for Big Data Partnerships

MAR 2015

BD Hubs Launched

BD Hubs solicitation to fund four regional Hubs is released

APR 2015

Big Data Regional Charrettes Held

Industry, academia, and government representatives gathered in four charrettes around the country

SEPT 2015

Hubs Awards Made

Awards made to coordinating institutions

NOV 2015

BD Spokes

BD Spokes solicitation released before 5th DC national charrette (bdhubs.info)

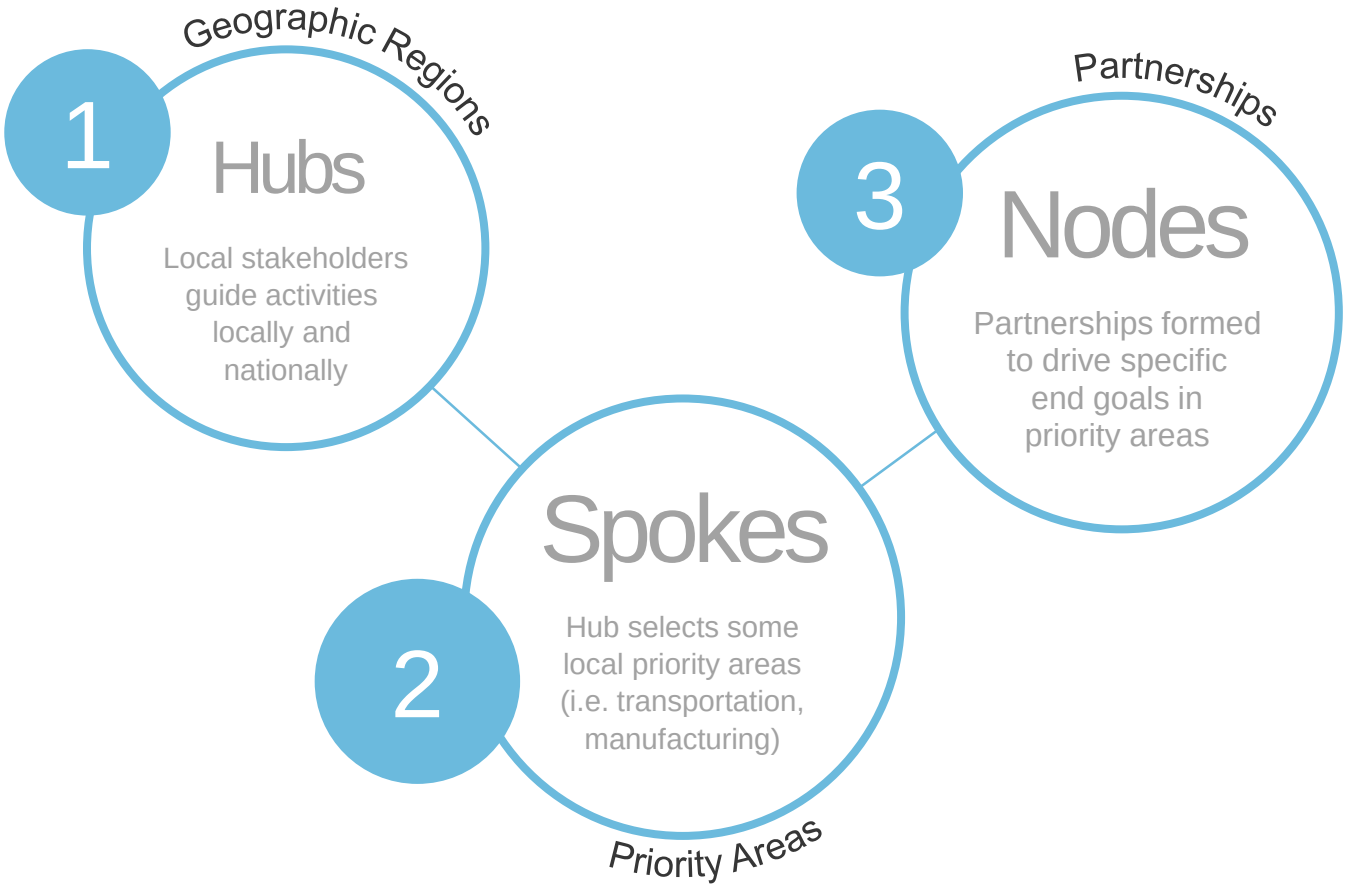
SEPT 2016

BD Spokes Awarded

10 (+1) Spokes and 10 planning grants awarded

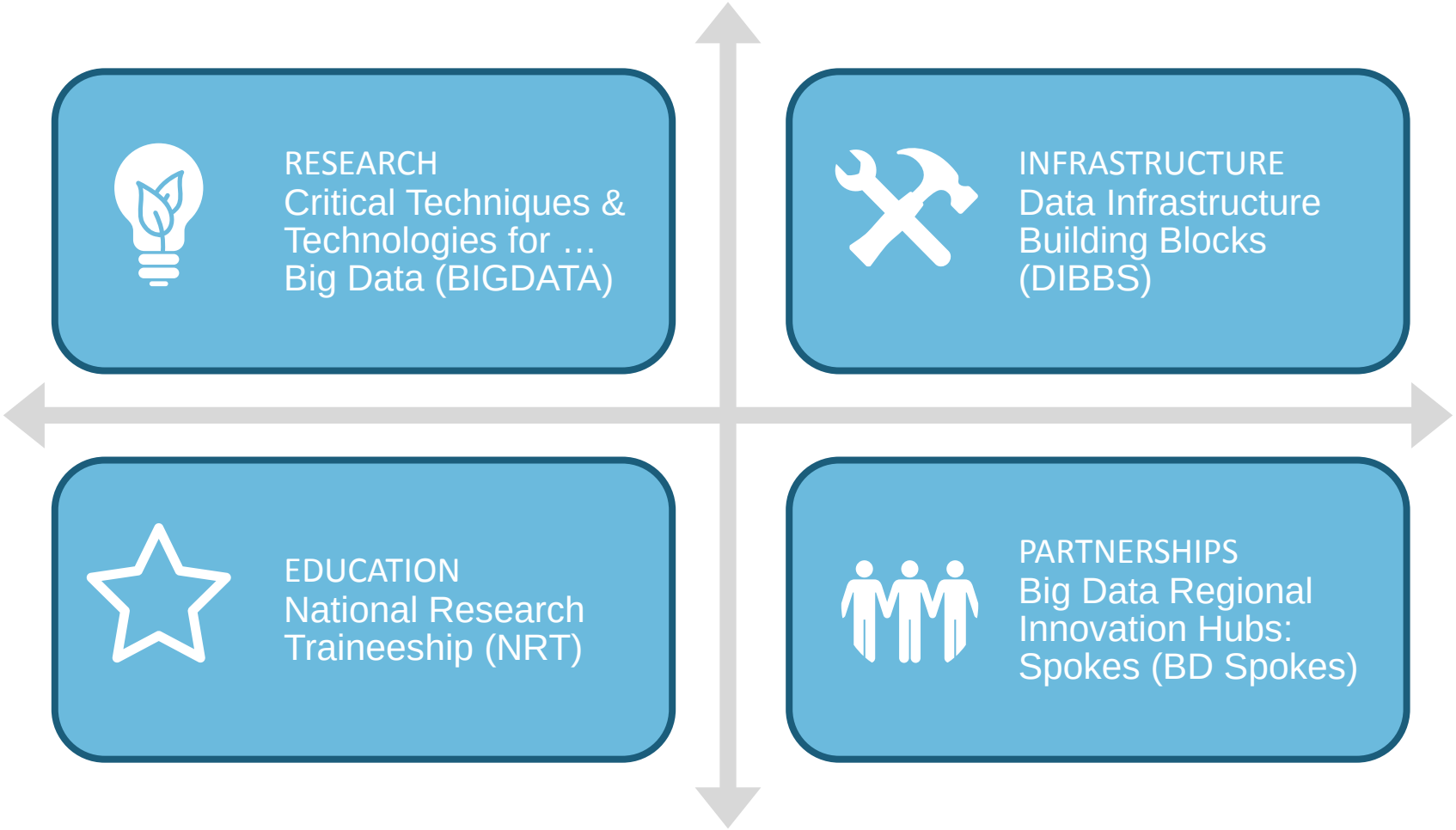
WHAT IS THE BDHUBS NETWORK?

"Hub and Spoke"– A Nation-Wide Network for Data Innovation

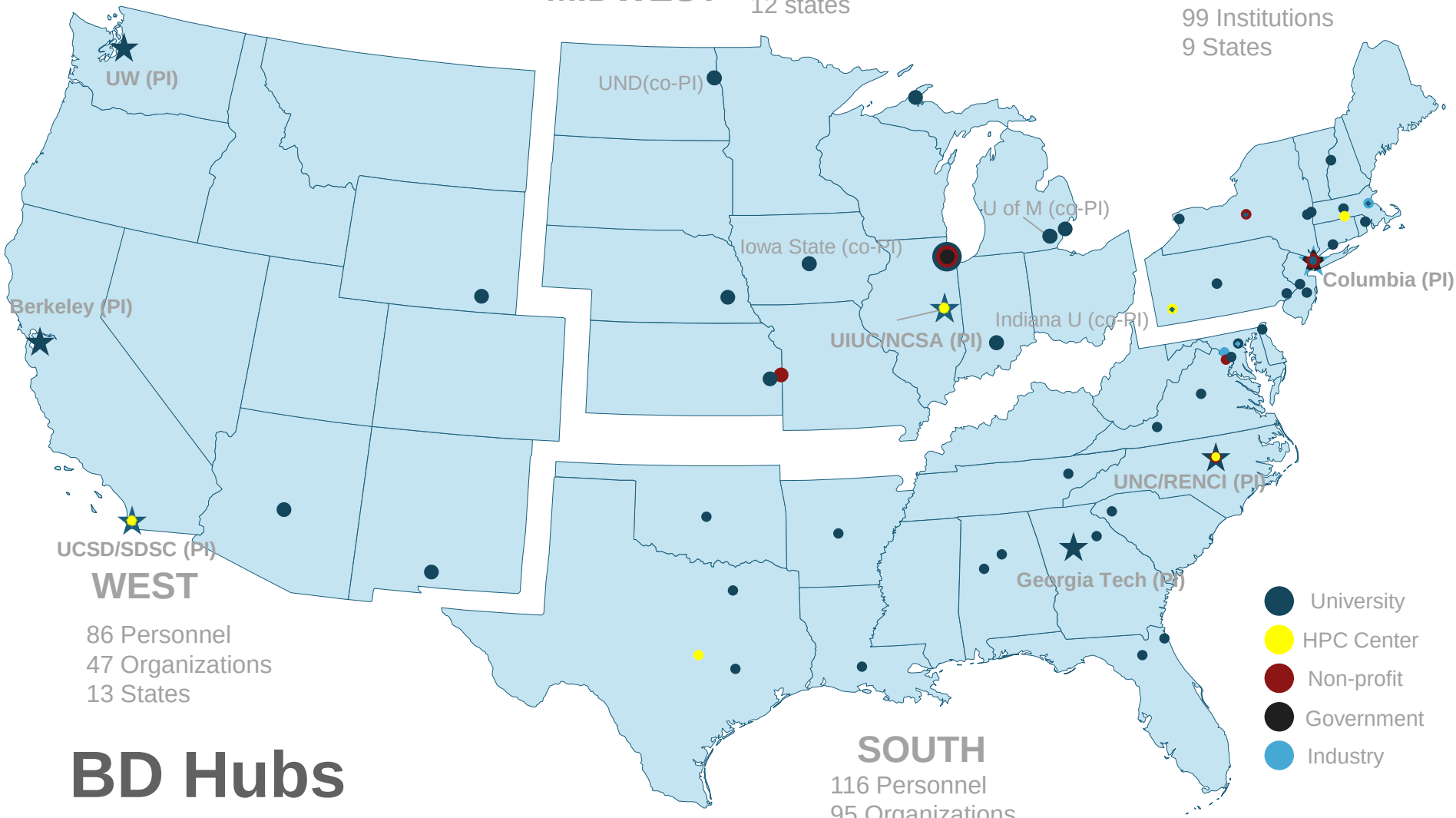


WITHIN THE BIG DATA PORTFOLIO OF PROGRAMS

Within the broader portfolio, BD Hubs and BD Spokes focuses on building partnerships around Big Data



Alaska & Hawaii are part of the West region
US Territories can participate in any region

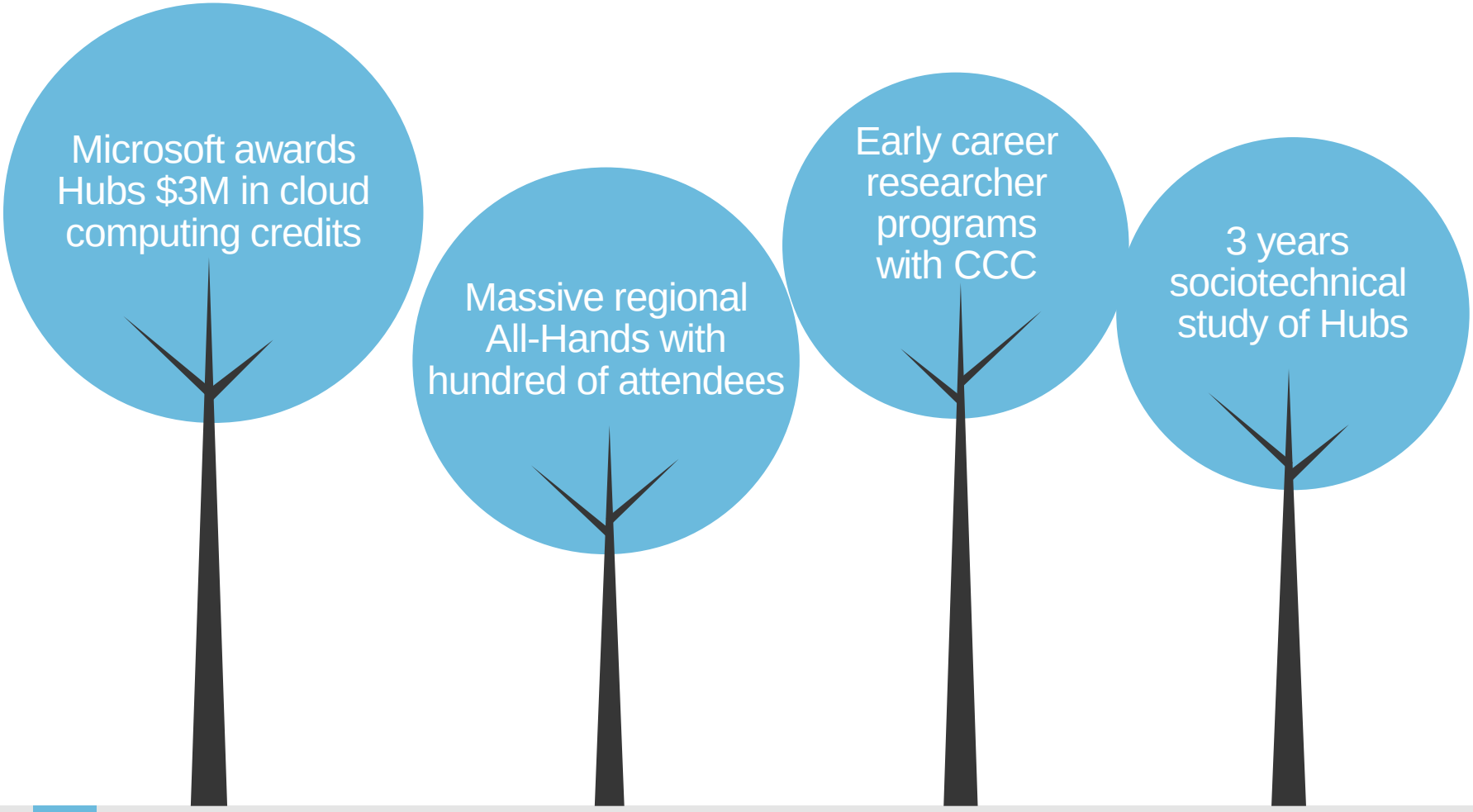


BD Hubs

Founding organizations for BDHubs in 2015
Points indicate affiliations of individuals named as steering council members and/or task leads or senior personnel.

HUB ACTIVITIES

Hubs ideate and coordinate Spokes, but also host a variety of activities for the community



The strategy behind BD SPOKES

BD Spokes are
not your typical R&D project
nor are they mini Hubs

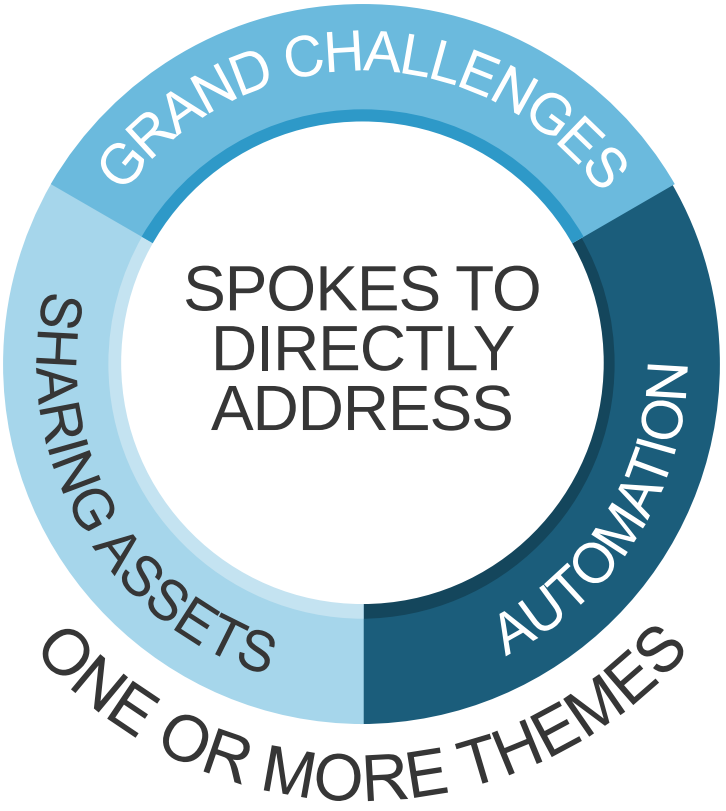
MISSION DRIVEN SPOKES

BD Spokes proposals must articulate a clear focus within a specific Big Data topic or application area, while highlighting their Big Data Innovation theme.

All BD Spokes must have clearly defined mission statements with goals and corresponding metrics of success.

SPOKES MAJOR THEMES

Three different ways of slicing the Big Data Innovation problem



AREAS OF EMPHASIS

Some NSF priority areas include

NEUROSCIENCE

REPLICABILITY & REPRODUCABILITY IN DATA SCIENCE

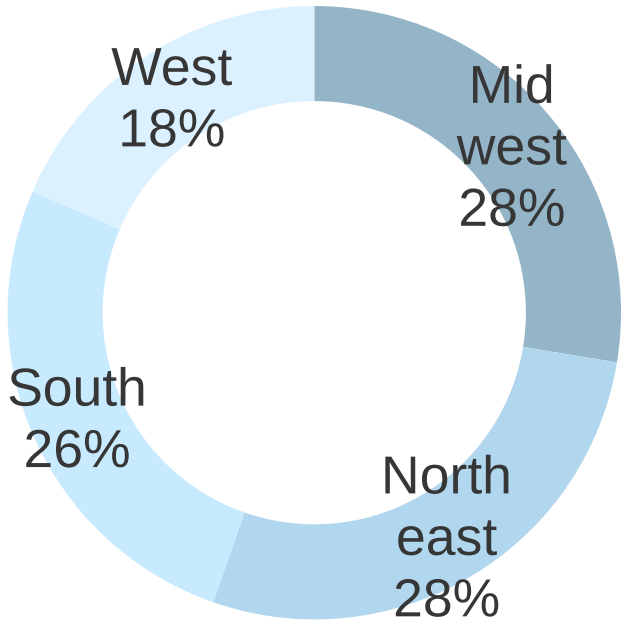
SMART & CONNECTED COMMUNITIES

DATA PRIVACY

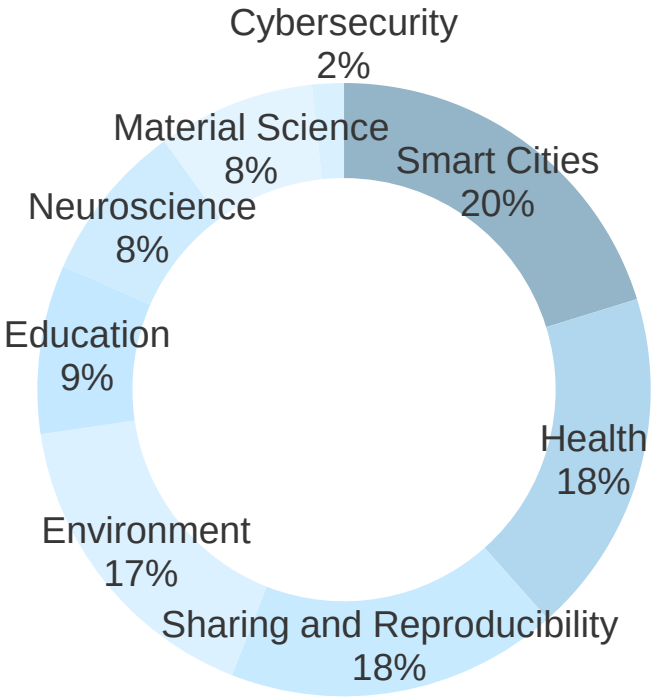
DATA INTENSIVE RESEARCH IN THE SOCIAL, BEHAVIORAL, & ECONOMIC SCIENCES

EDUCATION

Percent funding per region

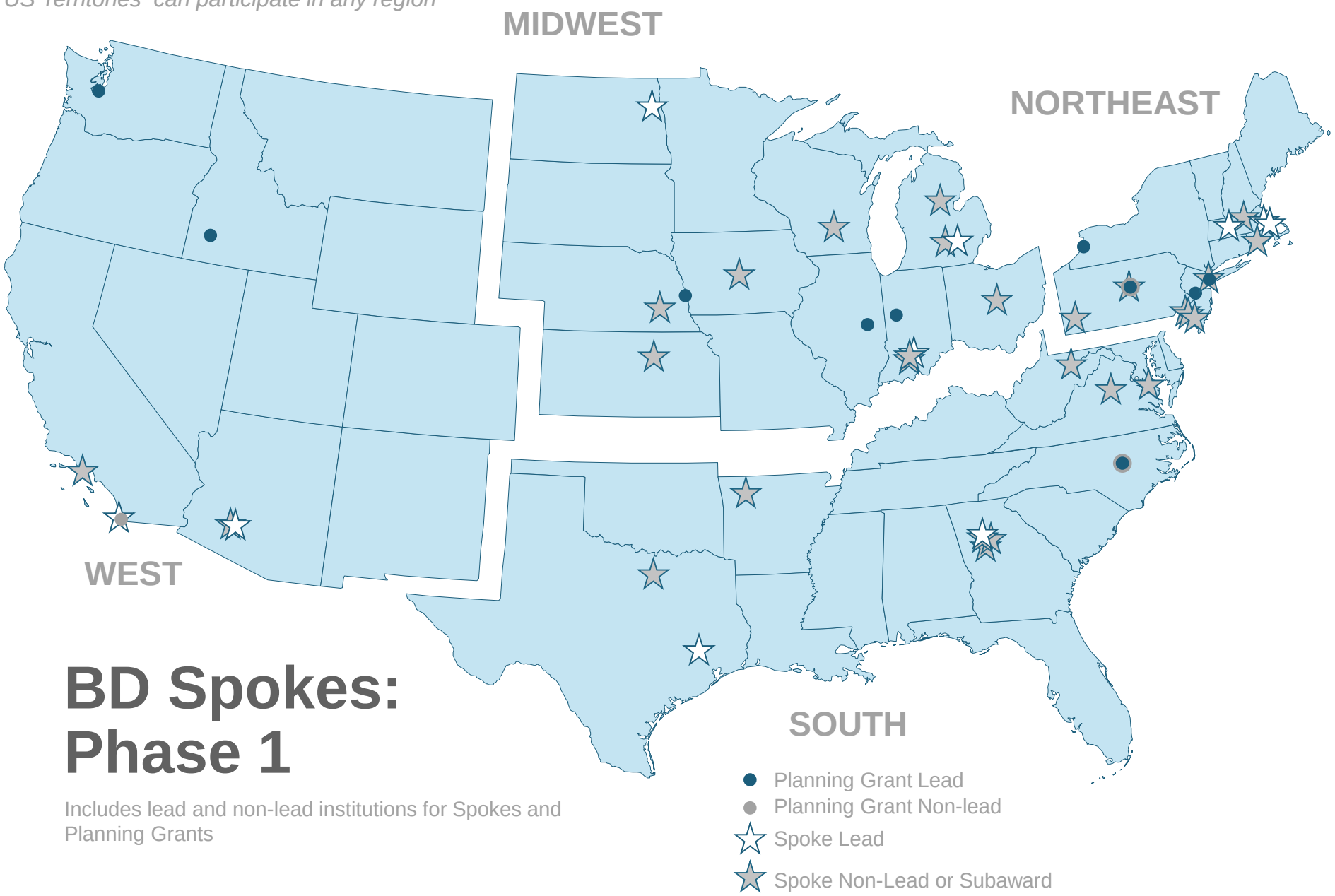


Percent funding per topic area



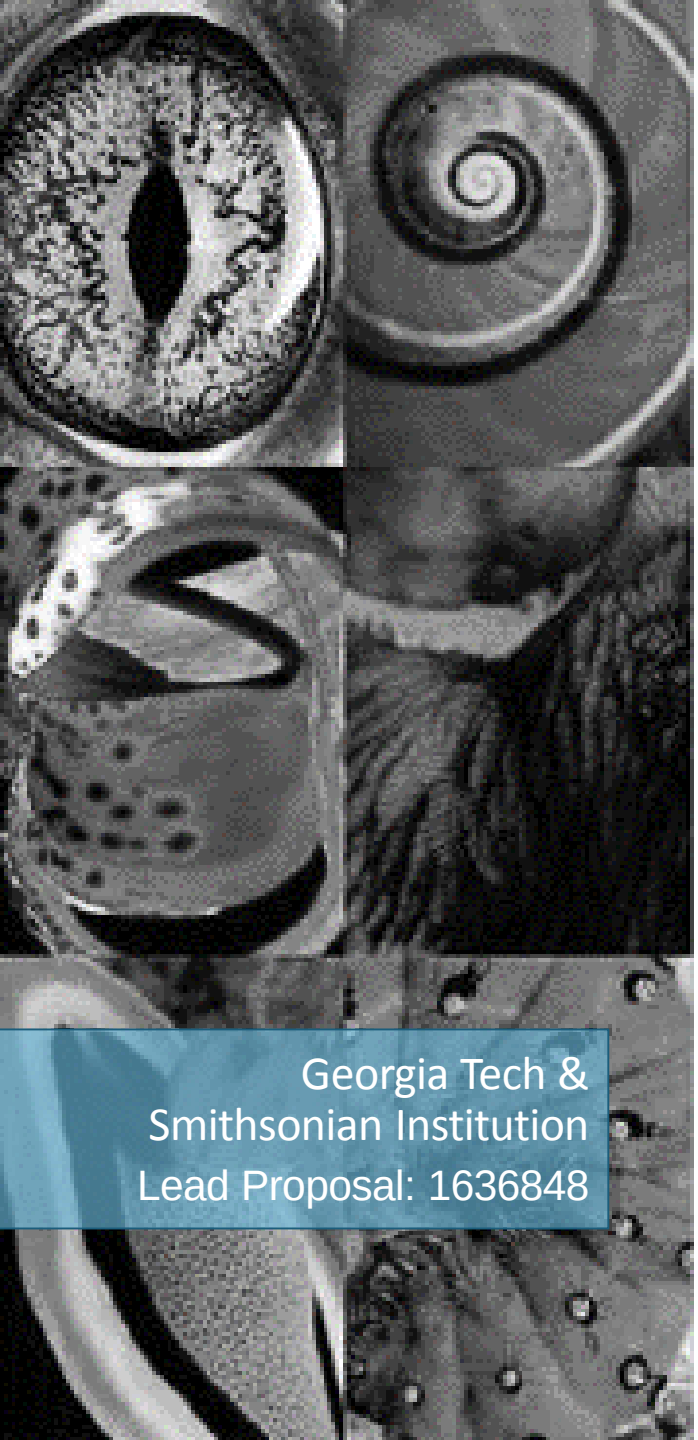
Total Spokes ~$12M in first round

Alaska & Hawaii are part of the West region
US Territories can participate in any region



BD Spokes: Phase 1

Includes lead and non-lead institutions for Spokes and Planning Grants



Georgia Tech &
Smithsonian Institution
Lead Proposal: 1636848

IBM WATSON + ENCYCLOPEDIA OF LIFE

“Using Big Data for Environmental Sustainability: Big Data + AI Technology = Accessible, Usable, Useful Knowledge!”

Encyclopedia of Life (EOL) is the world's largest database of biological species and other biodiversity information. EOL also works closely with scores of other biodiversity datasets such as BISON, GBIF, and OBIS.

This project seeks to make EOL and related biodiversity data sources accessible, usable, and useful, by integrating extant artificial intelligence tools for information extraction, modeling and simulation, and question answering.

(1) **Cognopsi**: semantically annotate documents in EOL through controlled vocabularies for specific domains within ecological and environmental science

(2) **MILA-S**: constructs conceptual models of ecological phenomena and automatically spawns simulation models; use with EOL TraitBank, to generate and test explanatory hypotheses as well as make predictions about ecosystems

(3) **Watson+:** adds semantic processing to Watson to act as a virtual research assistant; will train Watson+ for answering questions about biological species using EOL.

SMART GRID DATA SHARING

“Smart Grids Big Data”

Will create an organization that brings together a cross disciplinary capability from academia, industry, and government. The goal of the project is to ideate from Smart Grid Data new knowledge and solutions offering major improvements in smart grid operation (e.g., power generation and distribution; renewable energy) and smart grid user necessities (critical infrastructures, smart cities, transportation, etc.)

Over 67 organizations submitted letters of collaboration.

Will be building an open data and software exchange. Initial data committed:

- data provided by over 50 utility companies and 30 utility industry solution vendors
- National Lightning Detection Network Data from Vaisala
- Lawrence Livermore National Lab (LLNL) data coming from local sensor network including several PMU’s and weather monitoring devices
- International partners: Brazilian power system project MedFasee; demand side management studies University of Manchester, renewable generation data collection activities - University of Cyprus
- And many, many more

DIGITAL AGRICULTURE

"Unmanned Aircraft Systems (UAS), Plant Sciences and Education"

Will organize academic, industrial, and governmental sectors around the development of policies and best practices for data science and Big Data applications in agriculture

Main focus on automating the Big Data lifecycle:

- automation of transport, storage, dissemination, and analysis of UAS imagery and ground characterizations
- automation of Big Data pipelines and the integration, interoperability and re-use of databases across plant and cropping systems – from farm management and remote sensing to high throughput plant phenomics and crop genomics

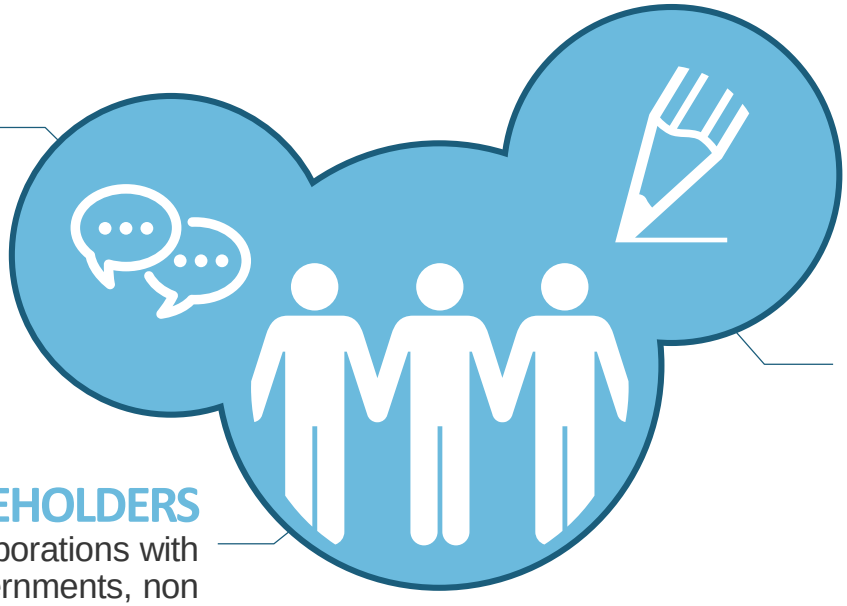
Activities focus on workshop series, hackathons, challenges, for example:

- Will develop a set of webinars on ontology, analytics, data management, data sharing, data standards and conventions, and data instrumentation to be used as a blueprint for a graduate level seminar on data science in agriculture
- Runs a competition for "mini proposals" in data annotation and interoperability for ag-genomics

KEY TAKEAWAYS

01 THE PROGRAM

Brings together domain scientists, computer scientists, and end users to use data to solve challenges



02 THE STAKEHOLDERS

Encourages collaborations with industry, state & local governments, non profits, and others that are not typical NSF participants

03 PARTICIPATION

Opportunity for NASA and your communities to get involved!

FOR FURTHER QUESTIONS CONTACT

Fen Zhao, fzhao@nsf.gov 703 292 7344

NSF Headquarters, Arlington VA