BIG DATA REGIONAL INNOVATION HUBS & SPOKES

An Update

Fen Zhao
June 9, 2017
WHAT IS THE BDHUBS NETWORK?

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

1. Hubs
   Local stakeholders guide activities locally and nationally

2. Spokes
   Hub selects some local priority areas (i.e. transportation, manufacturing)
WITHIN THE BIG DATA PORTFOLIO OF PROGRAMS

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

- **RESEARCH**
  Critical Techniques & Technologies for … Big Data (BIGDATA)

- **INFRASTRUCTURE**
  Data Infrastructure Building Blocks (DIBBS)

- **EDUCATION**
  National Research Traineeship (NRT)

- **PARTNERSHIPS**
  Big Data Regional Innovation Hubs: Spokes (BD Spokes)
THE BD HUBS & SPOKES TIMELINE

Currently

- **SEPT 2015**
  - Hubs Awards Made
  - Awards made to coordinating institutions

- **SEPT 2016**
  - BD Spokes Awarded
  - 10 (+1) Spokes and 10 planning grants awarded

- **MAR 2017**
  - 2nd Spokes Solicitation Released
  - $1M and <$500K award sizes

- **Summer 2017**
  - Hubs work with BD Spoke proposals
  - Hubs have internal deadlines to coordinate collaboration with BD Spoke proposals

- **SEPT 2017**
  - 2nd round BD Spokes Submitted
  - $10M in awards expected
Alaska & Hawaii are part of the West region
US Territories can participate in any region

**MIDWEST**
- 106 Personnel
- 79 Organizations
- 12 states

**NORTHEAST**
- 193 Personnel
- 99 Institutions
- 9 States

**WEST**
- 86 Personnel
- 47 Organizations
- 13 States

**SOUTH**
- 116 Personnel
- 95 Organizations
- 15 States + DC

**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.
Big Data Hubs and Spokes Ecosystem

Each BD Hub works with multi-institutional and multi-sector teams, or BD Spokes, to organize and coordinate projects to use data to address regional priorities.

$12M Awarded in FY2016-17:
11 Spokes Grants ($1M over 3 years); 10 Planning Grants ($100K for 1 year)

Themes:
- Improve data sharing,
- Automate the big data lifecycle, and
- Use data to address societal grand challenges

<table>
<thead>
<tr>
<th>Midwest</th>
<th>Northeast</th>
<th>South</th>
<th>West</th>
</tr>
</thead>
<tbody>
<tr>
<td>midwestbigdatahub.org</td>
<td>nebighub.org</td>
<td>southbigdatahub.org</td>
<td>westbigdatahub.org</td>
</tr>
<tr>
<td>Neuroscience Network</td>
<td>Licensing Model and Ecosystem for Data Sharing</td>
<td>Medical Informatics</td>
<td>Reproducibility in Scientific Computation</td>
</tr>
<tr>
<td>Digital Agriculture</td>
<td>Data-Driven Education</td>
<td>Smart Grids</td>
<td>MetroInsight</td>
</tr>
<tr>
<td>Integrative Materials Design</td>
<td>Observational Health Research</td>
<td>Big Data for Environmental Sustainability</td>
<td>Policing in the Western US</td>
</tr>
<tr>
<td>Bridge Health</td>
<td>Big Data Literacy</td>
<td>Rare Disease Observatory</td>
<td>Poteogenomics</td>
</tr>
<tr>
<td>Cyberinfrastructure for Sensor Originated Big Data</td>
<td>Big Data Cyber Attack Awareness</td>
<td>Applications of Genetic Data</td>
<td></td>
</tr>
<tr>
<td>Networked Resilience of Communities</td>
<td>Privacy and Security in Big Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Energy Cycle Innovation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Spoke Grant
Planning Grant
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 TO DIRECTLY ADDRESS

GRAND CHALLENGES

SPOKES TO DIRECTLY ADDRESS

SHARING ASSETS

ONE OR MORE THEMES

AUTOMATION

SPOKES MAJOR THEMES

Three different ways of slicing the Big Data Innovation problem
Percent funding per topic area

- Smart Cities: 20%
- Health: 18%
- Sharing and Reproducibility: 18%
- Environment: 17%
- Education: 9%
- Neuroscience: 8%
- Material Science: 8%
- Cybersecurity: 2%

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.
Big Spokes Solicitation (17-546)

Same 3 Themes: Data Sharing, Automation, Grand challenges

10 to 20 awards total $10M anticipated

MEDIUM Grants
$500,000 to $1,000,000 for up to 3 years

SMALL Grants
$100,000 to $500,000 for up to 3 years

Proposal submission deadline: 9/18/2017
No Letters of Intent
TANGIBLE OUTCOMES

Medium proposals are expected to articulate tangible outcomes:

• **Explicit results** from data-enabled or data-facilitated inquiry in a scientific or engineering field or other domain area, including social science and education research; must explicitly define the field- or domain-specific questions addressed and which specific researchers would be conducting the research to answer these questions.

• A **prototype** or proof of concept for a technology platform, data product, data standards, or other data infrastructure.

• An innovative **education or workforce development program** with a plan for evaluating the effectiveness of the program. The program could be virtual-, classroom-, or workplace-based and should engage underrepresented groups.
AREAS OF EMPHASIS FY 17

Some NSF priority areas include

- Neuroscience
- Replicability & Reproducibility in Data Science
- Data Driven Research in Chemistry
- Education
- Data Intensive Research in the Social, Behavioral, & Economic Sciences
- Data Analytics for Security
LIMITED COMPETITION

An organization may participate as submitting organization (or, in the case of collaborative proposals, as the lead organization) for at most one proposal responsive to this solicitation.

No limit is placed on an institution submitting a non-lead proposal in a collaborative proposal.

Proposal submissions are limited to maintain a balanced geographic representation of the Regional Hubs and Spokes program and to increase diversity of participating institutions.
FOR FURTHER QUESTIONS CONTACT

Fen Zhao, fzhao@nsf.gov 703 292 7344

NSF Headquarters, Arlington VA
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

- Foundational Research
- Cyberinfrastructure
- Education & Workforce Development
- Collaborations & Partnerships
- Policy
Harnessing the Data Revolution: five themes
Harnessing the Data Revolution: five themes

**Research** across all NSF Directorates

- **Theoretical foundations**
  - mathematics, statistics, computer & computational science

- **Systems, algorithms**
  - data-centric algorithms, systems

- **Data-intensive research**
  - in all areas of science and engineering

**Science domains**

- **Foundations**
- **Cyber infrastructure**
- **Systems, algorithms**
- **Education, Workforce**

**Educational pathways**

- Innovations grounded in an education-research-based framework
- **Advanced cyberinfrastructure**
  - Accelerating data-intensive research
Harnessing the Data Revolution: Domains

Research domains
science, engineering, education, ...

SBE  BIO  CISE
MPS  GEO  EHR
ENG
Harnessing the Data Revolution: Systems

Science domains
- Systems, algorithms
  - Predictive analytics
  - Data mining
  - Machine learning
  - Benchmark data sets
- Integrity and accessibility
- Privacy and protection
- Human-data interface

Fair, interpretable, transparent, trustworthy, auditable, ...
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

NSF TRIPODS: Transdisciplinary Research In Principles Of Data Science

Science domains
Systems, algorithms
Foundations
Theoretical foundations for data-driven discovery and decision making: analysis and modeling of complex heterogeneous data
Harnessing the Data Revolution: Cyberinfrastructure

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

Robust, open, science-driven, integrated research CI ecosystem, with data as a "first-class object"
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*

Workshop on Translational Data Science
University of Chicago, June 26-27, 2017

NSF is uniquely positioned to execute on the convergent, full-breadth of HDR activities
Data Corps
Envisioned as an effort at the local, state, national, and international levels to help unleash the power of data in the service of science and society.

- Enable American data scientists and data science students to obtain practical experience with data-intensive applications. Promote the use of data; help make data more usable; help make data interoperable;

- Promote a better understanding of the power of data and the state of the art in Data Science on the part of the communities served;

- Help promote a better understanding of the role that data can play in addressing issues at all levels—local, regional, national, international;

- Help develop capacity in data science in communities, organizations, and institutions at the local, state, national, and international levels
Data Corps

Data Corps Volunteers / Volunteer Organizations

Data Corps Projects / Project Organizations

Professionals from Industry, NGOs

Students from Academic Institutions

Edu Institutions

Undergrad / Grad Universities;
4-year Colleges;
Community Colleges;
Online Programs;
etc.

Basic Research

Smart & Connected Communities

Data Science for Social Good

Universities

Health

International Organizations

Criminal justice

Government

Transportation, Energy, …
Thank You!