BIG DATA REGIONAL INNOVATION HUBS & SPOKES

Accelerating the Innovation Ecosystem

QUILT Winter Member Meeting 2016

Fen Zhao
Staff Associate, Strategic Innovation
CISE Directorate, Office of the Assistant Director
WHAT IS THE BDHUBS PROGRAM?
An Agenda for The Discussion Today

01 THE HISTORY
BDHubs continue and scale up the innovation activities initiated by White House Data2Action event

02 THE STRATEGY
The multiphase BDHubs program aims to build regionally focused consortia around the country that will ideate, plan, and support Big Data partnerships and collaborative activities

03 THE SPOKES
NSF has released a solicitation (16-510) to kick off the second phase of the program
THE HISTORY BEHIND BD HUBS
The National Big Data R&D Initiative & Data to Knowledge to Action (Data2Action)

**Launch**
MAR 2012
OSTP and NITRD Agencies kick off National Big Data R&D Initiative with new federal programs totaling $200M

**Data2Action**
NOV 2013
90 organizations announce 29 new Big Data partnerships supported by $100M in non-federal funds

**Big Data Partnerships Workshop**
MAY 2013
Industry, academia, and government representatives gathered to learn about current Big Data partnership and brainstorm new ideas

**Partnerships Bear Fruit**
JUN 2014
Partnerships update NITRD on midterm outcomes from announced projects

**BDHubs**
MAR 2015
NSF initiates BDHubs effort to sustain and scale up collaborative Big Data innovation activities
THE HISTORY BEHIND BD SPOKES

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

- **JUN 2015** Hubs Proposals Submitted
  - Large collaborative proposals submitted to NSF

- **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)
CURRENT ACTIVITIES FOR THE PROGRAM

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

**BD Spokes Launched**
NOV 2015
BD Spokes Solicitation launched as 16-510 and Hubs meet in DC workshop

**Spokes Letter of Intent Submitted**
JAN 2016
NSF received ~100 letters of intent as a combination of planning grants and spoke proposals on a number of topics

**Spokes Full Proposals Due**
FEB 2016
Teams to submit full proposals on Feb 24, 2016

**Hubs Organize Spoke Drafting**
DEC 2015
Each Hub organizes workshops and draft Letter of Intent submission process to select project they would like to support

**Spokes Funded**
Summer 2016
BD Spokes and planning grants will be funded in FY16
“America is rich with institutions that are expert at generating data, but as a Nation we have not fulfilled our potential to make the most of these data by merging pre-competitive resources, partnering on analytics, and sharing lessons learned.

Today’s announcements show that we are maturing in this respect, finding synergies and collaborative opportunities that will accelerate progress in a wide range of scientific, social, and economic domains.”
WHAT ARE THE BENEFITS OF PARTNERING?

Achieve collectively what is impossible individually

**INITIATE PARTNERSHIPS**
Hubs will bring together academia, industry, non-profits, and government to initiate new partnerships. By collectively ideating and bringing together resources from across sectors, partnerships can drive faster innovation and more novel ideas.

**COMMON RESOURCES**
Participants can leverage the resources contributed by partners to Hub partnerships. Hubs can help develop “plug and play” infrastructure resources for partners. Resource providers can find users that will develop novel applications for their infrastructure.

**ACCESS TO TOP TALENT**
In a world where demand for Big Data talent far exceeds supply, Hubs will connect partners with students in academia. Projects with academia will train those students in projects of interest to partners before they even leave school.

**SHARED BEST PRACTICES**
Big Data practices, especially in a socio-technical context, are increasingly complex. Partners can develop and share best practices in areas such as privacy, discrimination, and ethics to ensure adoption while minimizing unwanted consequences.

**REDUCED COORDINATION COSTS**
Partnerships always come with a logistical cost. With BDHubs, NSF will fund the staff and logistics support necessary for more complex collaborations, reducing overhead and maximizing benefits for participants.
### Healthcare
- Novartis, Pfizer, and Eli Lilly partner to improve access to information about clinical trials.
- New platform builds on clinicaltrials.gov data to provide more detailed and patient-friendly information, including a machine-readable "target health profile" to improve the ability of healthcare software to match individual health profiles to applicable clinical trials.

### Foundational Research
- Berkeley AMPLab is funded by NSF, DARPA, DOE and a large number of private sector companies such as AWS, Google, and SAP.
- AMPLab creates Apache open source software platform (BDAS) for the whole community, including Spark/Shark, Mesos, Tachyon.
- Sponsors are able to interact with researchers and students at meetings, hearing about progress in cutting edge research.

### Education
- Funded by the Schmidt Family Foundation, University of Chicago runs the Data Science for Social Good summer program.
- Fellows work to solve and create apps to solve data science challenges defined by DSSG partners.
- Partners include City of Chicago, Cook County Land Bank, Cook County Sheriff, Ushahidi, Qatar Computing Research Institute, Lawrence Berkeley National Laboratory, Environmental Defense Fund and many others.
BASIC HUB STRUCTURE

NSF has set a broader structure for the Hub, with details to be determined by Hub participants as appropriate.

Steering Council
- Tasked with making key decisions (i.e. governance and agenda setting) for the regional consortium as a whole
- Consist of unpaid representatives from a subset of participating organizations
- Encouraged to be representative BD Hub’s membership, while also considering participation from underrepresented groups

Executive Staff
- The proposing organization should provide fiscal and implementation oversight to the BD Hub
- The proposing institution will establish a full-time, paid executive director and associated staff
- Will implement the decisions of the steering council and oversee day-to-day operations of the BD Hub

Partner Organizations
- Need not be members of the steering council
- Can join the Hub at the inception or during the period of the BD Hub award
- Need not be located within a region to be engaged in the corresponding consortium, given that many organizations have a national scope and will therefore span multiple regions.
Within the broader NSF portfolio, BDHubs focuses on building partnerships around Big Data.
HOW IS THE BDHUBS PROGRAM DIFFERENT?

BDHubs is not your typical NSF research program

NETWORKING
NSF is funding the staff & networking activities between partners, not research.

DYNAMIC
Hubs will be dynamic and grow over time to accommodate more interested participants.

COLLABORATION NOT COMPETITION
NSF asked for one proposal per region that describes the general consensus around Hub activities.

MULTIPHASE
Partners can use networking activities to determine what future priority areas to take on. Activities around these areas will be funded in later phases.
Hubs based on Census Regions of the
United States

Alaska & Hawaii are part of the West Region
US Territories can participate in any region
Throughout April 2015, NSF is sponsoring a series of Regional Charrettes.
Alaska & Hawaii are part of the West region
US Territories can participate in any region

BD Hubs
Points indicate affiliations of individuals named as steering council members and/or task leads.

MIDWEST
106 Personnel
79 Organizations
12 states

UIUC/NCSA (PI)
Indiana (co-PI)
Iowa State (co-PI)
U of M (co-PI)
UND (co-PI)

NORTHEAST
193 Personnel
99 Institutions
9 States

UCSD/SDSC (PI)
Berkeley (PI)
UW (PI)
Georgia Tech (PI)
UNC/RENCI (PI)

SOUTH*
116 Personnel
95 Organizations
15 States + DC

*South points indicate Senior Personnel

Columbia (PI)

WEST
86 Personnel
47 Organizations
13 States

UCSD/SDSC (PI)
Berkeley (PI)

Points indicate affiliations of individuals named as steering council members and/or task leads.
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

SPOKES TO DIRECTLY ADDRESS

AUTOMATION

ONE OR MORE THEMES

SHARING ASSETS

GRAND CHALLENGES

SPOKES MAJOR THEMES

Three different ways of slicing the Big Data Innovation problem
Accelerating progress towards societal grand challenges relevant to regional and national priority areas.

Due to the pervasiveness of Big Data in virtually all national priority areas, the BD Spokes have the opportunity to bring rapid change in application areas, by facilitating the creation of interdisciplinary and multidisciplinary data-intensive teams.
Steps in the **data lifecycle** include: ingestion, validation, curation, quality assessment, anonymization, publication, active data management, and analysis (including information extraction, visualization, and annotation).

**Automated (or, semi-automated) techniques** are needed in order to keep up with the rapid data rates, large volumes, and immense heterogeneity of Big Data. Automation may also aid the reproducibility of data processing and analysis workflows.
Enabling access to and increasing the use of valuable, available data assets, also including international data sets, where relevant.

One of the desirable roles for a BD Spoke is as a catalyst for organizing and sharing datasets and related data services among a larger set of stakeholders, across disciplinary areas, within the geographic region, or across the national community.
AREAS OF EMPHASIS

Some NSF priority areas include

- Neuroscience
- Replicability & Reproducibility in Data Science
- Smart & Connected Communities
- Data Privacy
- Data Intensive Research in the Social, Behavioral, & Economic Sciences
- Education
FOR FURTHER QUESTIONS CONTACT

Fen Zhao, fzhao@nsf.gov 703 292 7344