



BD Hubs: Midwest: “SEEDCorn: Sustainable Enabling Environment for Data Collaboration”

## Midwest Big Data Hub

Accelerating the Big Data Innovation Ecosystem

One of four Big Data Regional Innovation Hubs (BD Hubs) funded by the National Science Foundation through award #1550320

---

Edward Seidel

Director, National Center for Supercomputing Applications

Founder Prof. of Physics, U of Illinois

---

**Midwest Big Data Hub**

Accelerating the Big Data Innovation Ecosystem



# Welcome to our First All Hands Meeting!

- Introductions
  - PIs: Sarah Nusser, Beth Plale, Brian Athey, Josh Riedy, Ed S
  - Exec Dir: Melissa Cragin; Project Manager: Kandace Turner
  - NSF: Fen Zhao, Chaitan Baru
  - Other Hub Leaders
- What is MBDH all about?
  - SEEDCorn: **S**ustainable **E**nabling **E**nvironment for **D**ata **C**ollab**O**ration
  - A partnership of academia, government, industry, nonprofits
  - Over 100 partners already
    - Colleges, Universities, Medical Centers, of all types
    - Industry, Non-profits, NGOs
    - States, cities, communities

*“Creating communities that effectively harness the growing power of data to solve societal and economic problems of relevance in the Midwest”*

**Midwest Big Data Hub**

Accelerating the Big Data Innovation Ecosystem



# Most Important Messages and Goals

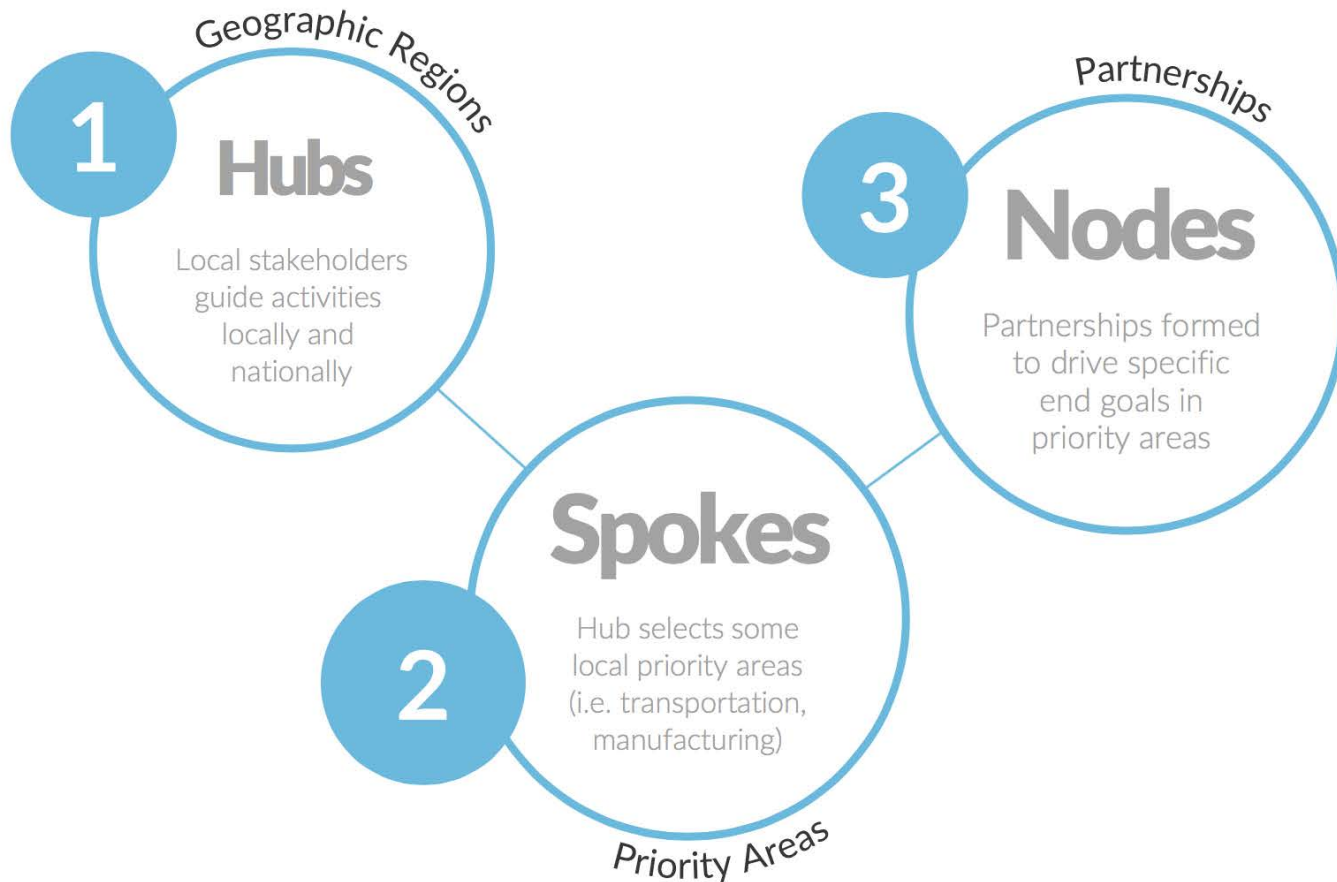
- Getting to know who we are
- Getting spoke organized and energized
  - What is a spoke?
  - Plenty of room for more projects in a spoke, and more spokes in the hub
  - What do spokes need from us
- Hearing from Industry and non-academic partners
  - How do we best partner?
- Begin to assess Data assets and needs
  - Assignments to Spoke sessions; discussion tonight and tomorrow
  - Feed into national meetings in future
- Refining governance of MBDH

**IN MORE DEPTH...**

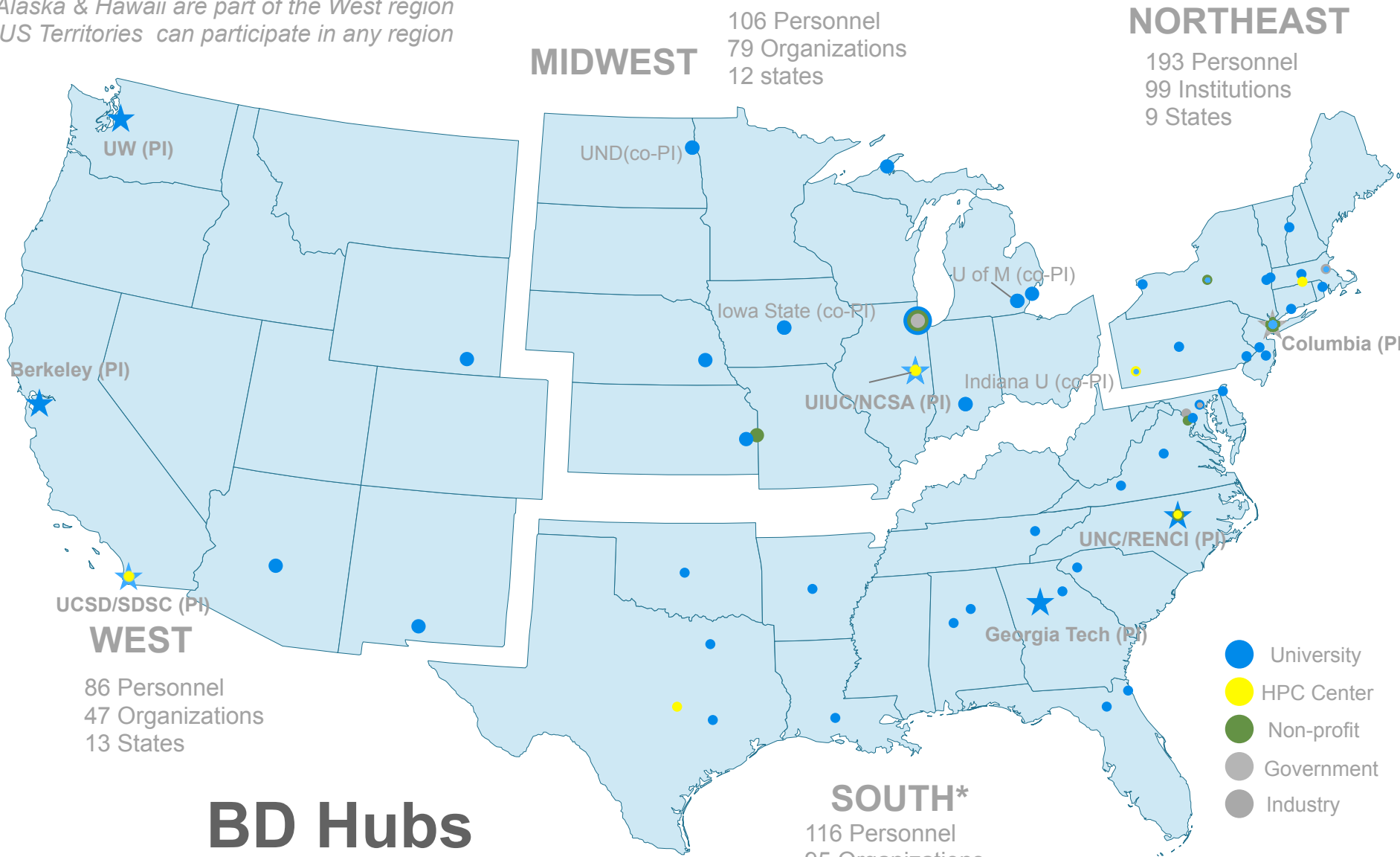
# NSF Big Data Hubs

## WHAT IS THE **BDHUBS NETWORK**?

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



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.

\*South points indicate Senior Personnel

## MIDWEST

106 Personnel  
 79 Organizations  
 12 states

## NORTHEAST

193 Personnel  
 99 Institutions  
 9 States

UW (PI)

UND(co-PI)

Berkeley (PI)

Iowa State (co-PI)

U of M (co-PI)

Columbia (PI)

UCSD/SDSC (PI)

UIUC/NCSA (PI)

Indiana U (co-PI)

UNC/RENCI (PI)

Georgia Tech (PI)

- University
- HPC Center
- Non-profit
- Government
- Industry

## SOUTH\*

116 Personnel  
 95 Organizations  
 15 States + DC

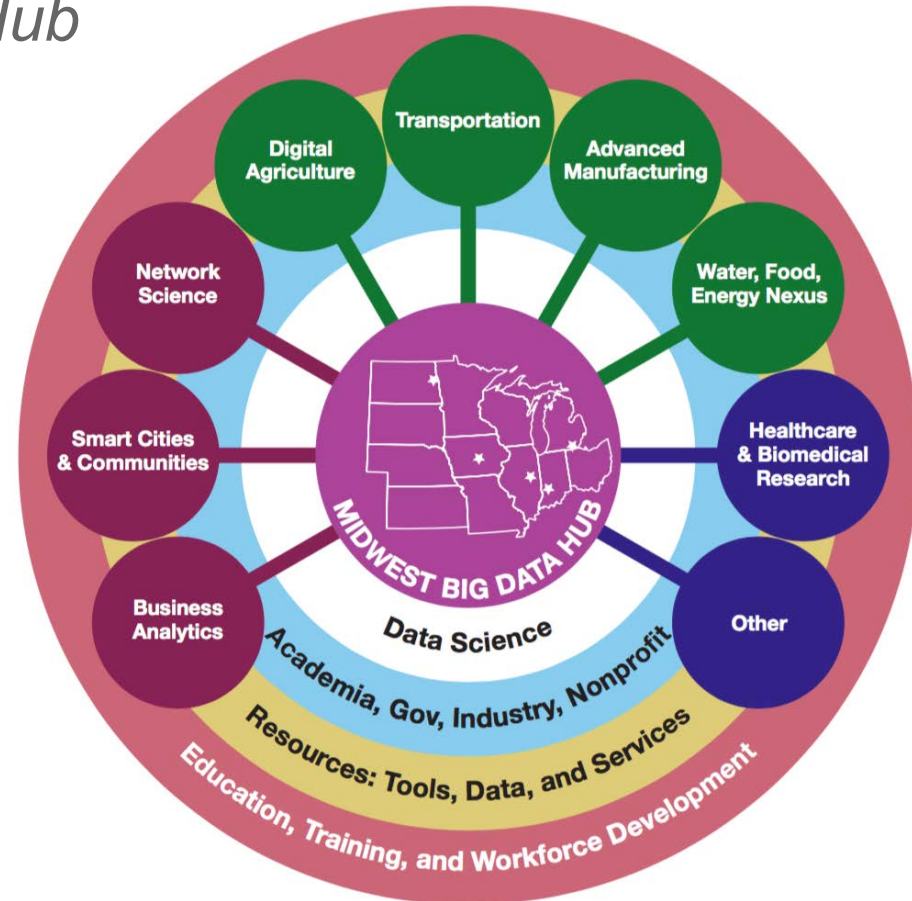
## WEST

86 Personnel  
 47 Organizations  
 13 States

# Spokes *Currently* Identified by MBDH

- Network Science
- Urban Science/Connected Communities
- Business Analytics
- Digital Agriculture
- Transportation
- Advanced Manufacturing
- Food, Energy, Water
- Healthcare & Biomedical Research
- Others as proposed...
  - *Materials*

*Spokes are supported by the Hub*





# Crosscutting Rings Supported by MBDH

- **Data Science**

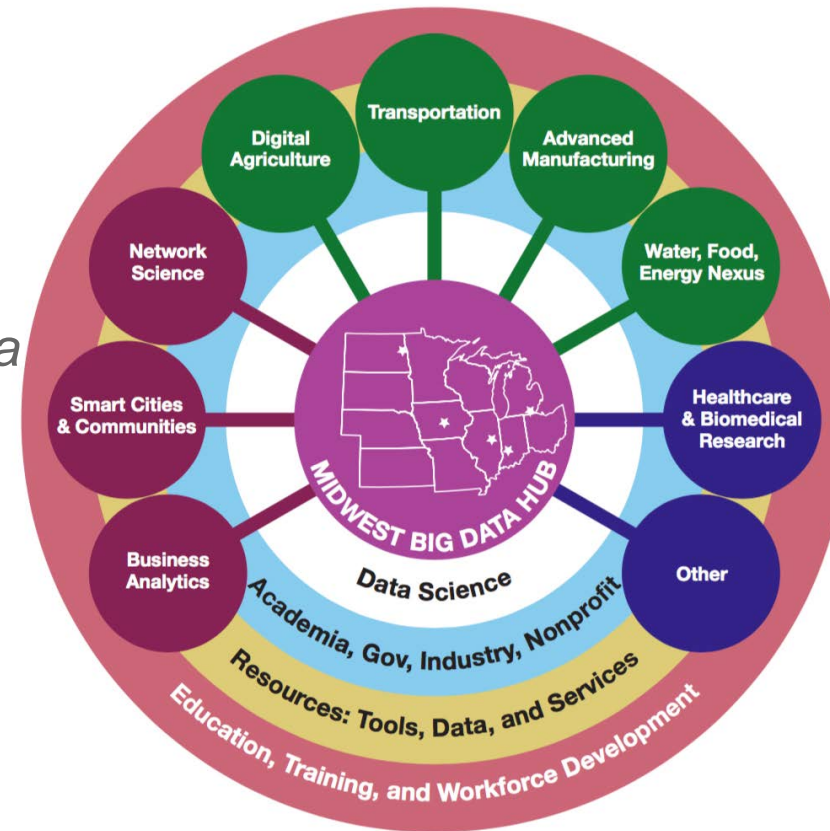
- *Including Data Intensive Research in the Social, Behavioral, and Economic Sciences...*
- *Replicability and Reproducibility in Data Science*

- **Education**

- *Including new approaches to STEM learning environments...*

- **Data Tools and Services**

- Data sets
- Services to share data and tools to act on them



*Rings are cross-cutting, supporting all spokes*

# Goals and Outcomes/Impacts Expected

*(Not necessarily today!)*

- Creating, strengthening, and securing funding for numerous new public-private partnerships
  - Additional funding from agencies (NSF, NIH, DOE, NIST, USDA...), NGOs, governments, industry will be sought
- Accelerating technology transfer projects
- Introducing new Big Data educational activities into universities, industry and government
  - Data policies, management, and best practices with real data for real impact
  - Will involve, train many young data scientists

# Goals and Outcomes/Impacts Expected

*(Not necessarily today!)*

- Starting *pilots* in data environments (SEEDCorn!)
  - Collaborations will come together to develop and test new approaches to data sharing, policies, algorithms
  - Will work with various organizations to test pilots with real data
    - For example: helping farmers balance productivity and sustainability with detailed data on crop growth, soil conditions & environment
    - Research Data Alliance (RDA), National Data Service (NDS), other orgs. HPC centers supporting pilots
    - **Some personnel support for communities** (will grow over time)
- Developing and implementing new sustainability models
  - Models for long term data stewardship, private-public partnerships, educational practice
  - Different approaches will be needed!

# Midwest Big Data Hub

## Potential Tools, Services, and Resources

This page collects information about potential tools, services and resources in the MidWest region (Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin) with the potential to support MBDH spokes and other activities. This page is being updated as information is collected. To add or amend entries please write to [admin@midwestbigdatahub.org](mailto:admin@midwestbigdatahub.org).

These resources are in addition to the [National Data Service \(NDS\)](#) capabilities which are available for spokes and supported by the MBDH as discussed on the hub web pages. The resources listed here were identified as assets available within our region. Spoke PIs should contact the organizations and individuals listed below as point of contacts to discuss availability of services.

[Networks](#)

[HPC Centers and Compute Resources](#)

[Cloud Resources](#)

[Data Curation and Publishing Tools](#)

[Storage and Repositories](#)

[Data Facilities and Infrastructure](#)

[Services](#)

[Software Services](#)

[Specialists](#)

[Educational Resources](#)

How can we bring these commitments to bear on our grand challenge problems?

# Back to Goals and Actions for this Meeting

- Spoke Updates and Breakouts
  - Where are we?
  - What do spokes need?
    - Grand challenge problems, communities, services from Hub
    - Data sets, access policy and need, services (see instructions)
  - Breakout session reports
- Industry Panel
  - How do we really engage industry effectively?
- Lightning Talks
  - Who are you and what would you like to do?
- Reception
  - Parallel session to coordinate across Hubs

# Back to Goals and Actions for this Meeting

- Spoke Updates
- Panel Session across Hubs
- Tools and Services discussions
  - Summary of needs and assets of spoke breakouts
  - What do we bring to the hub already?
    - MDF, NDS, others...
  - Outcomes
    - Pilots
    - Working Groups
      - Workshop planning for services for MBDH
      - Whitepaper to describe
      - As needed: you tell us!
- Final Discussion

# Upcoming Related Events



## 5th National Data Service Consortium Workshop

📅 April 4-6, 2016 📍 RENCI, Chapel Hill, North Carolina

Our theme will be *The National Data Service (NDS) and the Science Data Ecosystem*. The organizing committee is now contacting leaders in the research data community for plenary presentations. As speakers are confirmed we will update the meeting website.

- April NDS meeting: many partners across the US
  - Key sites in each region: SDSC, TACC, NCSA, PSC
  - Many NSF DIBBS/DataNet projects, publishers, and more
  - Tomorrow we will hear about Materials Data Facility as template for services for the Hubs
  - We have a great opportunity to create symbiotic relationship to serve MBDH and other hubs
- Other Hub meetings will help assess needed → Future events

**Midwest Big Data Hub**

Accelerating the Big Data Innovation Ecosystem

