Big Data Regional Innovation Hubs

Accelerating Innovation & Building Compelling Public-Private Partnerships

Academia

Industry



Non-profit

Government



THE FEDERAL BIG DATA RESEARCH AND DEVELOPMENT STRATEGIC PLAN

THE NETWORKING AND INFORMATION TECHNOLOGY RESEARCH AND DEVELOPMENT PROGRAM



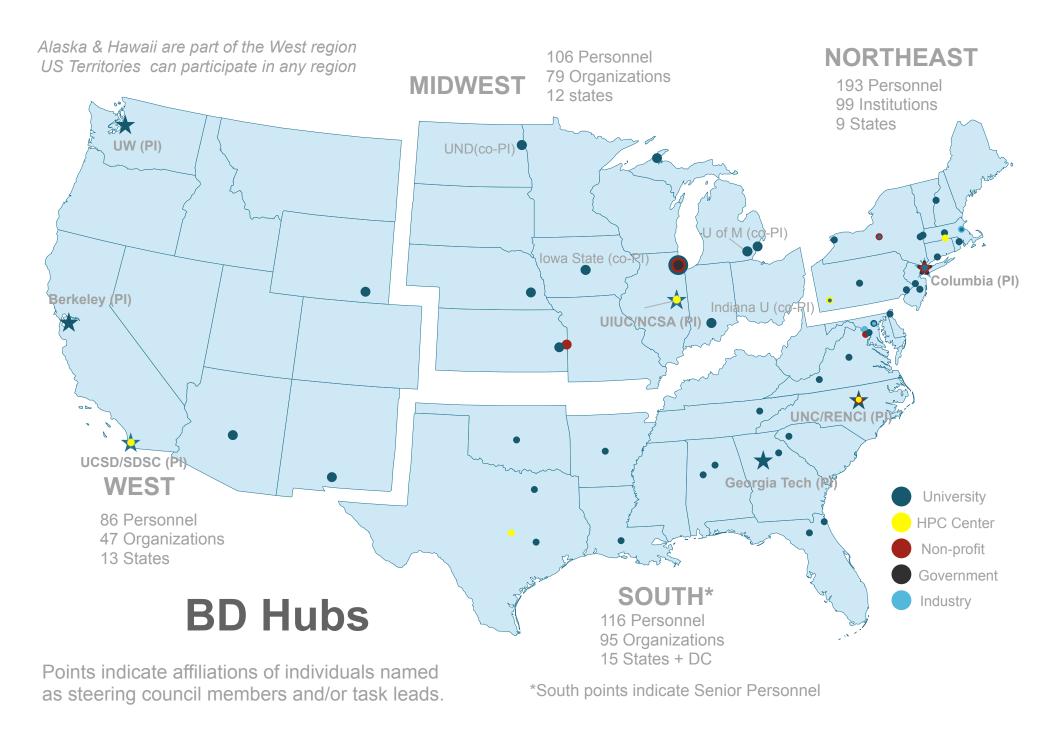
MAY 2016

Coordination

2012: White House Big Data Research and Development Initiative \$200M

2015: NSF launches 4 hubs with \$6M seed funding (3 years) and another \$10M to the Spokes in 2016

2016: Federal Big Data Strategic Plan



NSF Big Data Regional Innovation Hubs

- PROGRAM SOLICITATION NSF 15-562 https://www.nsf.gov/pubs/2015/nsf15562/nsf15562.htm
- Big Data Regional Innovation Hubs (BD Hubs); Accelerating the Big Data Innovation Ecosystem
- NSF Program Manager: Fen Zhao
- NORTHEAST: René Bastón, Executive Director
 - PI from Columbia University
 - Executive Committee from Penn State, Northeastern, UMass-Amherst, CMU
 - Governance committee from Academia, Foundations, Government, Industry, Non-Profit
 - http://nebigdatahub.org/leadership/
- SOUTH: Co-ED's: Renata Rawlings-Goss, Georgia Tech & Lea Shanley, UNC Chapel Hill
 - PI's from Georgia Tech and UNC Chapel Hill
 - Steering Committee from Academia, Industry, Government, Non-Profit
 - http://southbdhub.org/
- MIDWEST: Melissa Cragin, Executive Director
 - PI from: UIUC/NCSA; Co-PI's from IU, Iowa State, UMichigan, UNorth Dakota
 - Steering committee from Academia, Industry, National Lab, Regional Network, Foundation
 - http://midwestbigdatahub.org/steering-committee/
- WEST: Meredith Lee, Executive Director Pl's from: UCBerkeley, UW, UCSD/SDSC
 - Steering Committee from Academia, Industry, Government
 - http://westbigdatahub.org/people/

The NSF Big Data Regional Innovation Hub Spoke awards to date

NORTHEAST:

- A Licensing Model and Ecosystem for Data Sharing (MIT, Brown University, Drexel University)
- Grand Challenges for Data-Driven Education (<u>University of Massachusetts, Amherst</u>; <u>Worcester Polytechnic Institute</u>; <u>University of Pennsylvania</u>)
- Integration of environmental factors and causal reasoning approaches for large-scale observational health research (<u>Harvard University</u>, <u>Columbia University</u>, <u>University of Pittsburgh</u>, <u>Penn State</u>)

SOUTH:

- Large-scale Medical Informatics for Patient Care Coordination and Engagement (<u>Emory University</u>)
- Smart Grids Big Data (<u>Texas A&M</u>, <u>Temple University</u>, <u>Georgia Institute of Technology</u>)
- Using Big Data for Environmental Sustainability: Big Data + AI Technology = Accessible, Usable, Useful Knowledge (Georgia Institute of Technology, Smithsonian Institution)

MIDWEST:

- Advanced Computational Neuroscience Network (<u>University of Michigan</u>, <u>Indiana University</u>, <u>Ohio State</u> <u>University</u>, <u>Case Western University</u>)
- Digital Agriculture Unmanned Aircraft Systems, Plant Sciences&Education (<u>University of North Dakota</u>)

WEST:

- Accelerating and Catalyzing Reproducibility in Scientific Computation and Data Synthesis (ASU)
- MetroInsight: Knowledge Discovery and Real-time Interventions from Sensory Data Flows in Urban Spaces (University of California, San Diego; University of California, Los Angeles; ASU)

The NSF Big Data Regional Innovation Hub Planning Grants and EAGER awards to date

NORTHEAST:

- Big Data Literacy: Building Capacity for Regional Collaboration in Closing the Big Data Divide (<u>New York Hall of Science</u>)
- Cross-Organization Big Data Cyber Attack Awareness (<u>Pennsylvania State University</u>)
- Planning for Privacy and Security in Big Data (<u>Rutgers University</u>, <u>Penn State</u>)
- Partnerships for Energy Cycle Innovation through Big Data (PPEID) (<u>University of Buffalo</u>)

SOUTH:

- Rare Disease Observatory (<u>North Carolina State University</u>, <u>University of North Carolina</u>)
- EAGER: Data Bridge for Neuroscience (University of North Carolina at Chapel Hill)

MIDWEST

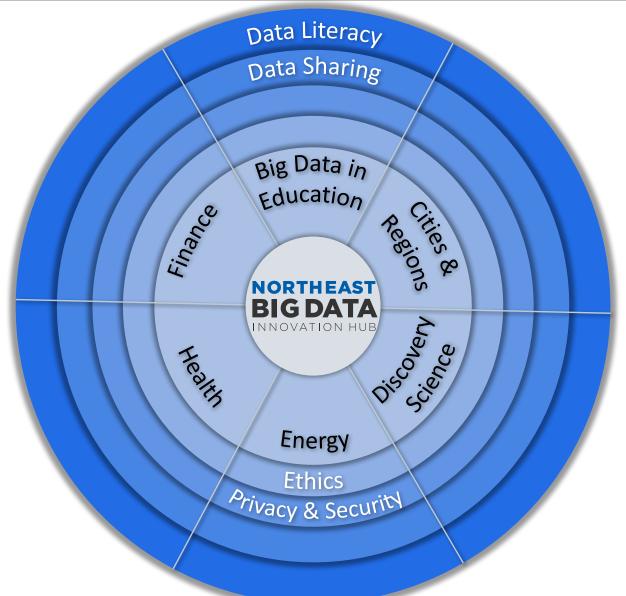
- Big Data Innovations for Bridge Health (University of Nebraska)
- Cyberinfrastructure to Enhance Data Quality and Support Reproducible Results in Sensor Originated Big Data (<u>Purdue University</u>)
- Networked Resilience of Communities Facing Natural and Social Emergencies (<u>University of Illinois at Urbana-Champaign</u>)
- EAGER: Incorporating Citizen Science into Real-Time Sensor-Based Estimates of Traffic-Related Air Pollution Exposure (Ohio State University)

WEST

- Big Data for Policing in the Western United States (<u>Boise State University</u>)
- Increasing collaborations in proteogenomics applications of genetic data (<u>Institute for Systems</u> Biology; <u>University of California</u>, San Diego)

Priority Areas





7

Northeast Big Data Hub Awards to date



Spoke Awards

- Integration of Environmental Factors and Causal Reasoning Approaches for Large-Scale
 Observational Health Research
 - Harvard University; Award Amount:\$393,871
 - Columbia University: \$372,880
 - University of Pittsburgh: \$111,075
 - Pennsylvania State University University Park: \$95,367
- Grand Challenges for Data-Driven Education
 - University of Massachusetts, Amherst: \$324,988
 - Worcester Polytechnic Institute: \$429,110
 - University of Pennsylvania: \$224,999
- A Licensing Model and Ecosystem for Data Sharing
 - Massachusetts Institute of Technology: \$444,000
 - Brown University: \$322,552
 - Drexel University: \$209,793

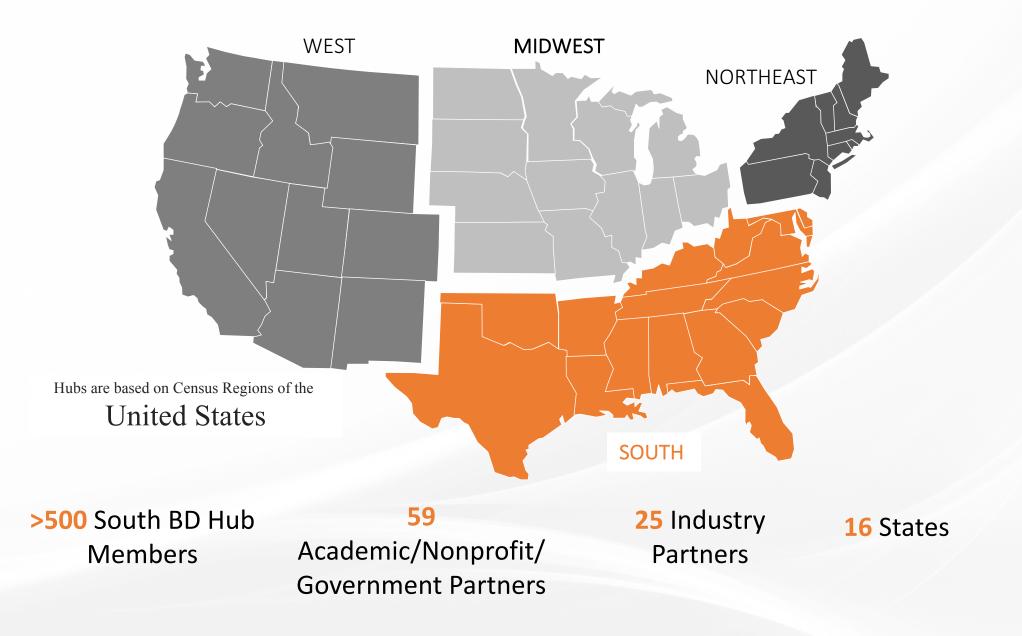
Planning grants:

- Cross-Organization Big Data Cyber Attack Awareness: Penn State- \$99,609
- Partnerships for Energy Cycle Innovation through Big Data (PPEID): SUNY Buffalo \$99,533
- Planning for Privacy and Security in Big Data: Rutgers, Penn State \$89,963
- Big Data Literacy: Building Capacity for Regional Collaboration in Closing the Big Data Divide

- New York Hall of Science: \$84,748

South Big Data Regional Innovation Hubs





South Big Data Hub Awards to date



Spoke Awards

- Smart Grids Big Data
 - Texas A&M Engineering Experiment Station: \$666,999
 - Georgia Institute of Technology, Georgia Tech Research Corporation: \$250,001
 - Temple University; Award Amount: \$100,000
- Using Big Data for Environmental Sustainability: Big Data + Al Technology = Accessible,
 Usable, Useful Knowledge! And Data!
 - Georgia Institute of Technology, Georgia Tech Research Corporation: \$754,600
 - Smithsonian Institution: \$245,267
- Large-Scale Medical Informatics for Patient Care Coordination and Engagement
 - Emory University: \$1,000,000

Planning grants:

- Collaborative: Rare Disease Observatory
 - North Carolina State University: \$71,143
 - University of North Carolina at Chapel Hill: \$28,857
- EAGER: Data Bridge for Neuroscience: A novel way of discovery for Neuroscience Data
 - University of North Carolina at Chapel Hill: \$89,892





IOWA STATE UNIVERSITY





Sarah Nusser PI (Illinois) Co-PI (Indiana) Co-PI (Iowa State) Co-PI (Michigan) Co-PI, (UND)

Brian Athev

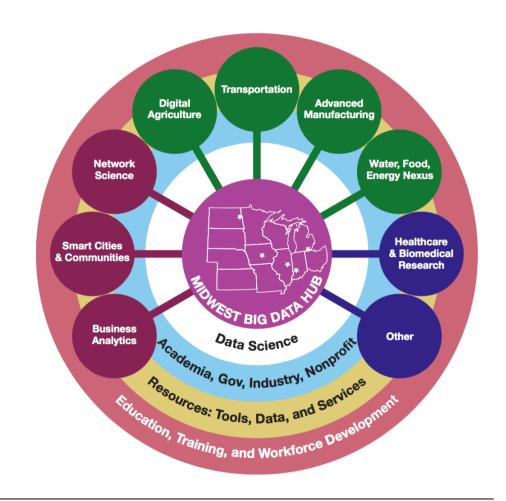
Josh Riedy





Thematic Spoke communities underway

- **Digital Agriculture**
- **Metropolitan Sciences**
- **Transportation**
- Healthcare & Biomedical Research
- **Business Analytics**
- **Advanced Manufacturing**
- Food, Energy, Water
- **Network Science**
- Others Proposed: Materials & Cybersecure energy systems











Midwest Big Data Hub Awards to date



Spoke Awards

- Advanced Computational Neuroscience Network (ACNN)
 - University of Michigan Ann Arbor: \$343,263
 - Indiana University: \$332,869
 - Ohio State University: \$166,454
 - Case Western Reserve University: \$164,912
- Digital Agriculture Unmanned Aircraft Systems, Plant Sciences and Education
 - University of North Dakota Main Campus: \$995,739

Planning Grants:

- Networked Resilience of Communities Facing Natural and Social Emergencies
 - University of Illinois at Urbana-Champaign: \$100,000
- Big Data Innovations for Bridge Health
 - University of Nebraska at Omaha: \$99,959
- Cyberinfrastructure to Enhance Data Quality and Support Reproducible Results in Sensor
 Originated Big Data
 - Purdue University: \$99,938
- EAGER: Incorporating Citizen Science into Real-Time Sensor-Based Estimates of Traffic-Related Air Pollution Exposure
 - Ohio State University: \$99,613

West Themes + Working/Interest Groups

West Big Data Innovation Hub

13

Many engagement options



- Metro Data Science
- Precision Medicine
- Natural Resources + Hazards
- Scientific Discovery + Learning
- Big Data Technology

- Governance
- Cross-Sector Collaboration (metrics)
- Infrastructure/Shared services
- Privacy/Security
- #datascistories (Data-Driven Stories)

West Big Data Hub Awards to date



Spoke Awards

- Accelerating & Catalyzing Reproducibility in Scientific Computation and Data Synthesis
 - Arizona State University; Award Amount: \$1,014,593
- MetroInsight: Knowledge Discovery and Real-Time Interventions from Sensory Data Flows in Urban Spaces
 - University of California-San Diego: \$517,817
 - University of California-Los Angeles: \$303,063
 - Arizona State University: \$179,117

Planning grants:

- Big Data and Criminal Justice in the Western United States
 - Boise State University: \$97,298
- Increasing collaborations in proteogenomics applications of genetic data
 - Institute for Systems Biology: \$71,000
 - University of California-San Diego: \$29,000

14

Internet2 Participation in NSF Big Data Hubs

- Internet2 Distributed Big Data & Analytics Innovation Working Group requested involvement with NSF BDHubs to align our efforts for Internet2 community and NSF
- Began NSF BDHub collaboration at Northeast BDHub kickoff Dec 2015 at Columbia University - CINO and NS engaging with infrastructure teams in Northeast and South hubs – key is to find ways for them to leverage the Internet2 infrastructure
- Internet2 on Organizing Committee for Data Sharing workshop led by Northeast BDHub Sept 2016 with South, Midwest, West plus NSF Program Manager Fen Zhao
- Smart Grid data sharing chosen as use case for NE BDHub data sharing spoke –
 Northeast BDHub engaging with South BDHub and "Smart grids big data" spoke
- Internet2 invited to present at "Federal Data in Action Summit" Dec 15, 2016 to discuss data science, Trust/Identity/Privacy/Protection/Safety/Security of IoT data
- NS and CE engaging Midwest hub to support genomics big data transfer

Back-up



Governance Committee



Academia	
Nabil Adam	University of Rutgers- Newark (NJ)
Mark Gerstein	Yale University (CT)
Tim Kraska	Brown University (RI)
Theresa Pardo	SUNY Albany (NY)
John Storey	Princeton University (NJ)
Lorenzo Torresani	Dartmouth College (NH)
Foundations	
Josh Greenberg	Alfred P. Sloan Foundation
Government	
Amen Ra Mashariki	Mayor's Office of Data Analytics (NY)
Industry	
Micah Adler	Fiksu (MA)
Jennifer Chayes	Microsoft Research (MA)
Yann LeCun	Facebook and New York University (NY)
Non-Profits	
Cheryl Begandy	Pittsburgh Supercomputing Center (PA)
John Goodhue	Massachusetts Green HPC Center (MA)
Dave Mankiewicz	CenterState Corporation for Economic Opportunity (NY)

Spoke and Ring Leads



18

Ryan Baker	Penn State (Education)
Jane Greenberg	Drexel University (Data Sharing)
James Hendler	Rensselaer Polytechnic Institute (Data Literacy)
Chris Hill	Massachusetts Institute of Technology (Discovery Science)
Chirag Patel	Harvard University Medical School (Health)
Michael Kearns	UPENN (Finance)
Constantine Kontokosta	New York University (Cities & Regions)
Mark Latonero	Data & Society (Ethics)
Karen Levy	Data & Society (Ethics)
Sam Madden	Massachusetts Institute of Technology (Data Sharing)
Manish Parashar	Rutgers University (Discovery Science)
Abani Patra	SUNY Buffalo (Energy)
Adam Smith	Penn State University (Privacy and Security)
Jennifer Stromer-Galley	Syracuse University (Ethics & Policy)
Beverly Woolf	UMASS Amherst (Big Data in Education)

South BDHub Steering Committee



19

Name	Title	Organization
Academia:		
Dan Stanzione	Executive Director of the Texas Advanced Computing	University of Texas, Austin/TACC
	Center (TACC)	
Anthony Mezzacappa	Director of the Joint Institute for Computational Sciences	JICS/University of Tennessee, Knoxville
	(JICS)	
Joseph JaJa	Professor, Laboratory for Parallel Computation	University of Maryland, Institute for
		Advanced Computer Studies
Jim Bottum	Chief Information Officer	Clemson University
David Norton	Vice President for Research	University of Florida
Donald Brown	Director, The Data Science Institute	University of Virginia
Mladen Vouk	Department Head and Distinguished Professor of	North Carolina State University
	Computer Science, and Director of the NC State Data	
	Science Initiative	
Carolina Cruz-Neira	Professor, Department of Information Science	University of Arkansas – Little Rock
Michael Zwick	Assistant Dean for Research, Emory U SOM; Assistant	Emory University
	Vice President for Research, Robert W. Woodruff Health	
	Sciences Center	
Industry:		
Colin Parris	Vice President, GE Software Research	General Electric
Armistead Sapp	Executive Vice President & Chief Technology Officer	SAS
Pradeep Dubey	Director, Parallel Computing Lab (PCL)	Intel
Dennis O. Little	Vice President of Production	Lockheed Martin Corporation
Kevin W. Kolander		McKesson
Government:		
Tom Frieden	Director	CDC
Joe McClosky		DOD
Non-Profit:		
Chad Evans	Senior Vice President	Council of Competitiveness

Midwest Bid Data Hub – Working Groups



"Spoke" is a thematic community addressing priority areas identified for our region "Ring" is a working group that will address priorities that cross all spokes.

- Spoke 1: Food-Water-Energy; Lead: Klara Nahrstedt (Illinois)
- Spoke 2: Health Sciences, Life Sciences, Bioinformatics, Genomics; Lead: Brian Athey (Michigan)
- Spoke 3: Metro Sciences, Lead: Charlie Catlett (Argonne National Lab/U. Chicago)
- Spoke 4: Digital Agriculture (precision farming, sustainability, etc.); Lead: Joe Colletti (lowa State)
- Spoke 5: Advanced Manufacturing; Lead: Jim Barkley (UI Labs)
- Spoke 6: Network Science; Lead: Bernice A. Pescosolido (Indiana)
- Spoke 7: Transportation; Lead: H.V. Jagadish (Michigan)
- Spoke 8: Business Analytics; Lead: Ratna Babu Chinnam (Wayne State)
- Ring 1: Data, Tools and Services; Lead: Carol Flannagan (Michigan)
- Ring 2: Data Science; Lead: Beth Plale (Indiana)
- Ring 3: Education; Lead: Wolfgang Kliemann (Iowa State)





Midwest BDHub Steering Committee



Name	Title	Organization
	Specific and Regular Seats	
Ed Seidel	PI, Interim Steering Committee Chair	NCSA/Univ of Illinois Urbana-Champaign
Brian Athey	Co-PI, Interim Healthcare & Biomedical Spoke Lead	University of Michigan
Sarah Nusser	Co-PI	Iowa State University
Beth Plale	Co-PI, Interim Data Sciences Spoke Lead	Indiana University
Grant McGimpsey	Co-PI	University of North Dakota
Charlie Catlett	Metro Sciences Spoke Lead	Argonne National Labs & Univ of Chicago
R. Babu Chinnam	Business Analytics Spoke Lead	Wayne State
Klara Nahrstedt	Food, Energy, Water Spoke Lead (Interim)	University of Illinois Urbana-Champaign
Jim Barkley	Digital Manufacturing Spoke Lead	UI Labs
Bernice Pescosolido	Social Networks Spoke Lead	Indiana University
Joe Colletti	Digital Agriculture Spoke	Iowa State University
H. V. Jagadish	Transportation Spoke Lead	University of Michigan
Lior Shamir	Education Ring Lead	Lawrence Technological University
Carol Flannagan	Data, Tools, and Services Ring Lead (Interim)	University of Michigan
	At-Large Members - Elected	
Vetria Byrd	2-year term, 2016-2018	Purdue Polytechnic Institute, Purdue Univ
Jennifer Clark	2-year term, 2016-2018	University of Nebraska-Lincoln
Aneetha Gopalan	2-year term, 2016-2018	Conservis Corp, Minneapolis
Hridesh Rajan	2-year term, 2016-2018	Iowa State University
Dan Reed	2-year term, 2016-2018	University of Iowa
	At-Large Members - Appointed by Executive Committee	
Greg Monaco	2-year term, 2016-2018	Great Plains Network
Keith Elliston	2-year term, 2016-2018	tranSMART Foundation

West Big Data Innovation Hub People



	West dig data
Leadership Team	Innovation Hub
MEREDITH LEE	Executive Director
CHRISTINE KIRKPATRICK	Deputy Director
SARAH STONE	Deputy Director
MICHAEL FRANKLIN	PI, Professor, UCBerkeley
ED LAZOWSKA	PI, University of Washington, eScience Institute Director
MICHAEL NORMAN	PI, Director, SDSC, Distinguished Professor of Physics, UCSD
BILL HOWE	Co-PI, University of Washington, eScience Institute
ERIN ROBINSON	Co-PI, Executive Director, Foundation for Earth Science
ARIEL ROKEM	Co-PI, University of Washington, eScience Institute
Steering Committee	
NICK BAGULEY	President, Big Data Utah - Evolve STEAM, Utah
TITUS BROWN	Associate Professor, UC Davis, California
STEVE CUTCHIN	Associate Professor, Boise State University, Idaho
ALEXANDER "SAM" FERNALD	Director, Water Resources Research Institute, New Mexico State Univ, New Mexico
THOMAS HAUSER	Director of Research Computing, University of Colorado, Boulder, Colorado
ALICIA JOHNSON	San Francisco Emergency Management, California
DANE SKOW	Director of Research Services, UWIT, University of Wyoming, Wyoming
VIDYA SPANDANA	Principal, New Knowledge Data Science, Oregon
IAN SWANSON	Founder & CEO, DataScience, California
KRISTIN TUFTE	Portland State Univ, U.S. Department of Transportation, Oregon
JACKIE WIRZ	Assistant Professor, Oregon Health & Science University, Oregon
JULIE MEIER WRIGHT	Strategic Advisor,, Collaborative Economics

Back-up - NSF BDHub and Spoke Awards

BD Spokes: SPOKE: MIDWEST: Collaborative: Advanced Computational Neuroscience Network (ACNN)

Award Number:1636846; Principal Investigator:Dhabaleswar Panda; Co-Principal Investigator:Xiaoyi Lu, Hari Subramoni; Organization:Ohio State University;NSF Organization:IIS Start Date:09/01/2016; Award Amount:\$166,454.00; Relevance:48.0;

BD Hubs: NORTHEAST: The Northeast Big Data Innovation Hub

Award Number:1550284; Principal Investigator:Kathleen McKeown; Co-Principal Investigator:; Organization:Columbia University;NSF Organization:IIS Star Date:10/01/2015; Award Amount:\$1,579,387.00; Relevance:48.0;

BD Spokes: SPOKE: WEST: Accelerating and Catalyzing Reproducibility in Scientific Computation and Data Synthesis

Award Number:1636796; Principal Investigator: C Michael Barton; Co-Principal Investigator: Kenneth Buetow, Marco Janssen, Allen Lee; Organization: Arizona State University; NSF Organization: IIS Start Date: 09/01/2016; Award Amount: \$1,014,593.00; Relevance: 48.0;

BD Spokes: SPOKE: MIDWEST: Digital Agriculture - Unmanned Aircraft Systems, Plant Sciences and Education

Award Number:1636865; Principal Investigator:William McGimpsey; Co-Principal Investigator:Joe Colletti, Gregory Monaco, Jennifer Clarke, Travis Desell; Organization:University of North Dakota Main Campus;NSF Organization:IIS Start Date:10/01/2016; Award Amount:\$995,739.00; Relevance:48.0;

BD Spokes: SPOKE: WEST: Collaborative: MetroInsight: Knowledge Discovery and Real-Time Interventions from Sensory Data Flows in Urbai Spaces

Award Number:1636879; Principal Investigator:Rajesh Gupta; Co-Principal Investigator:Ilkay Altintas, Julian McAuley; Organization:University of California-San Diego;NSF Organization:IIS Start Date:09/01/2016; Award Amount:\$517,817.00; Relevance:48.0;

BD Hubs: Collaborative Proposal: SOUTH: A Big Data Innovation Hub for the South Region

Award Number:1550305; Principal Investigator:Srinivas Aluru; Co-Principal Investigator:; Organization:Georgia Tech Research Corporation;NSF Organization:IIS Start Date:10/01/2015; Award Amount:\$856,682.00; Relevance:48.0;

BD Spokes:PLANNING: MIDWEST: Networked Resilience of Communities Facing Natural and Social Emergencies

Award Number:1636942; Principal Investigator:Marshall Poole; Co-Principal Investigator:Alex Yahja; Organization:University of Illinois at Urbana-Champaign;NSF Organization:IIS Start Date:09/01/2016; Award Amount:\$100,000.00; Relevance:48.0;

BD Spokes: SPOKE: SOUTH: Collaborative: Smart Grids Big Data

Award Number:1636770; Principal Investigator:Zoran Obradovic; Co-Principal Investigator:; Organization:Temple University;NSF Organization:IIS Start Date:09/01/2016; Award Amount:\$100,000.00; Relevance:48.0;

BD Spokes: SPOKE: SOUTH: Collaborative: Smart Grids Big Data

Award Number:1636772; Principal Investigator:Mladen Kezunovic; Co-Principal Investigator:Le Xie, Dilma Da Silva, Panganamala Kumar;

Organization:Texas A&M Engineering Experiment Station;NSF Organization:IIS Start Date:09/01/2016; Award Amount:\$666,999.00; Relevance:48.0;

BD Spokes: SPOKE: SOUTH: Collaborative: Smart Grids Big Data

Award Number:1636783; Principal Investigator:Santiago Grijalva; Co-Principal Investigator:; Organization:Georgia Tech Research Corporation;NSF Organization:IIS Start Date:09/01/2016; Award Amount:\$250,001.00; Relevance:48.0;

BD Spokes: PLANNING: NORTHEAST: Partnerships for Energy cycle Innovation through Big Data (PPEID)

Award Number:1636818; Principal Investigator:Abani Patra; Co-Principal Investigator:; Organization:SUNY at Buffalo;NSF Organization:IIS Start Date:09/01/2016; Award Amount:\$99,533.00; Relevance:48.0;

EAGER: DBfN: Data Bridge for Neuroscience: A novel way of discovery for Neuroscience Data

Award Number:1649397; Principal Investigator:Arcot Rajasekar; Co-Principal Investigator:Howard Lander; Organization:University of North Carolina at Chapel Hill;NSF Organization:IIS Start Date:10/01/2016; Award Amount:\$89,892.00; Relevance:48.0;

BD Spokes: SPOKE: MIDWEST: Collaborative: Advanced Computational Neuroscience Network (ACNN)

Award Number:1636840; Principal Investigator:Richard Gonzalez; Co-Principal Investigator:George Alter, Ivo Dinov; Organization:University of Michigan Ann Arbor;NSF Organization:IIS Start Date:09/01/2016; Award Amount:\$343,263.00; Relevance:48.0;

BD Spokes: SPOKE: MIDWEST: Collaborative: Advanced Computational Neuroscience Network (ACNN)

Award Number:1636893; Principal Investigator:Franco Pestilli; Co-Principal Investigator:Lei Wang, Andrew Saykin, Olaf Sporns; Organization:Indiana University;NSF Organization:IIS Start Date:09/01/2016; Award Amount:\$332,869.00; Relevance:48.0;