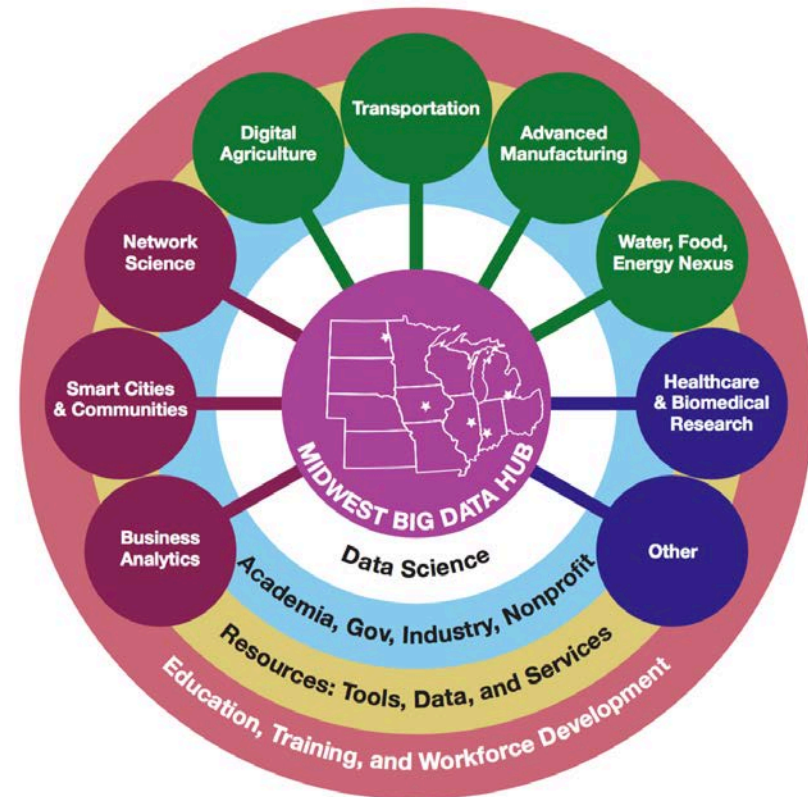# Tools and Services Ring

Ed Seidel

Kenton McHenry

Tools & Services Ring Goals

- Bolster/support Hub efforts
  - Use data collections from multiple sources to address grand challenges posed by the spokes
- Intersects many/all other Spokes
- Identify/leverage data tools and services from across the region
  - Address interoperability
  - Find gaps
- Address longer term sustainability

Midwest Big Data Hub
Accelerating the Big Data Innovation Ecosystem

Tools & Services Ring

- National Data Service Consortium
  - Search, Publish, Link, Reuse
  - Interoperability between data cyberinfastructure efforts
- Support pilot projects to aid MBDH spoke activities

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

Midwest Big Data Hub
Accelerating the Big Data Innovation Ecosystem

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.

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
Tech Transfer and/or Incubators
Experimental facilities that generate data

Networks

- BOREAS, Broadband Optical Research, Education and Sciences Network (Iowa, Minnesota, Wisconsin)
  - Collaboration of Iowa State University, University of Iowa, University of Minnesota, and University of Wisconsin-Madison.
- Citizen Science Grid (North Dakota), Travis Desell
  - Hosted by UND's Computational Research Center and Information Technology Systems and Services, dedicated to supporting a wide range of research and educational projects using volunteer computing and citizen science.
- Great Plains Network (Iowa, Kansas, Minnesota, Missouri, North Dakota, South Dakota, Wisconsin)
  - PoPs in Kansas City, Kansas, and Tulsa, Oklahoma; connects 9 state networks to the Internet2 and ESnet backbones:

89 regional resources

Midwest

Accelerating the Big Data Inno

Break Down Science Drivers

- Data Collections
  - Un-digitized
  - Offline
  - Restricted
  - Unstructured
  - Un-curated
  - Simulation outputs
  - Representations
  - Diverse data
  - Cross disciplinary collections
  - Large datasets
  - Many small datasets

- Data Handling Needs
  - Data hosting
  - Access control
  - Data collaboration
  - Data curation tools
  - Data analysis tools
  - Data visualizations
  - Long term archiving
  - Support for community tools
  - High performance computation
  - Cloud computination

Midwest Big Data Hub
Accelerating the Big Data Innovation Ecosystem

Break Down of Technology Components

- Infrastructure & Resources
  - Data
  - Networking
  - Storage
  - Security
  - High performance computing
  - Cloud computing resources

- Data Service Capabilities
  - Authentication
    - Identity management
    - Access control
  - Transfer
    - Large transfers
    - Robust transfer
  - Storage
    - Storage abstraction
    - Sharing/collaboration
    - Archiving
    - Replication
    - Data publishing
    - Code publishing
  - Curation
    - Ingestion
    - Description
    - Provenance
    - Transformation
    - Integration
  - Analysis
    - Data analysis
    - Event notification
  - Exploration
    - Federated search
    - API
    - User interface

Midwest Big Data Hub
Accelerating the Big Data Innovation Ecosystem

Break Down of Active Efforts

| | Data Service Capabilities | | | | | | | | | | | | |
| Storage | | | | Curation | | | | | Analysis | | Exploration | | |
Data Archiving	Data Replication	Data Publishing	Code Publishing	Data Ingestion/ Curation	Data Description	Provenance	Data Transformation	Data Integration	Data Analysis	Event Notification	Federated Search	Data Exploration API	Data Exploration User Interface
								◎					
			◎			◎	◎		◎			◎	
								◎	◎				
		◎		◎									
		◎		◎		◎						◎	◎
		◎		◎									
			◎						◎				
◎	◎												
				◎	◎								
								◎	◎				

Midwest Big Data Hub

Accelerating the Big Data Innovation Ecosystem

MBDH Tools & Services

- Contained broadly reusable components addressing some specific data need
  - Data Transfer Service
  - Identity Service
  - Data Storage Service
  - HPC/Cloud Computation Service
  - Data Sharing/Publishing Services
  - Data Archiving Service
  - Data Transformation Service
  - Code/Tool Publishing Service
- Interoperability with other such components

Midwest Big Data Hub

Accelerating the Big Data Innovation Ecosystem

All Hands Workshop '16

- Identify needs specific to potential regional spokes
  - A. Datasets that are needed or can be made more broadly available that others might make use of
  - B. Terms and policies for sharing data
  - C. What tools and services would help users in using the data

Midwest Big Data Hub
Accelerating the Big Data Innovation Ecosystem