**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

---

# Spoke Overview: Water, Food, Energy Nexus

**Shashi Shekhar**
McKnight Distinguished University Professor
University of Minnesota
www.cs.umn.edu/~shekhar

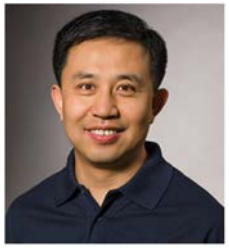


# Core Team: FEW Spoke

- Klara Nahrstedt (UIUC) – MBDH leadership council
- Shashi Shekhar, Jessica Hellman, David Haynes (UMN)
- Shaowen Wang, Anand Padmanabhan, Luis Rodriguez (UIUC)
- Carol Song (Purdue U)
- Aaron Packman (NWU)
- Adam Ward (Indiana U)





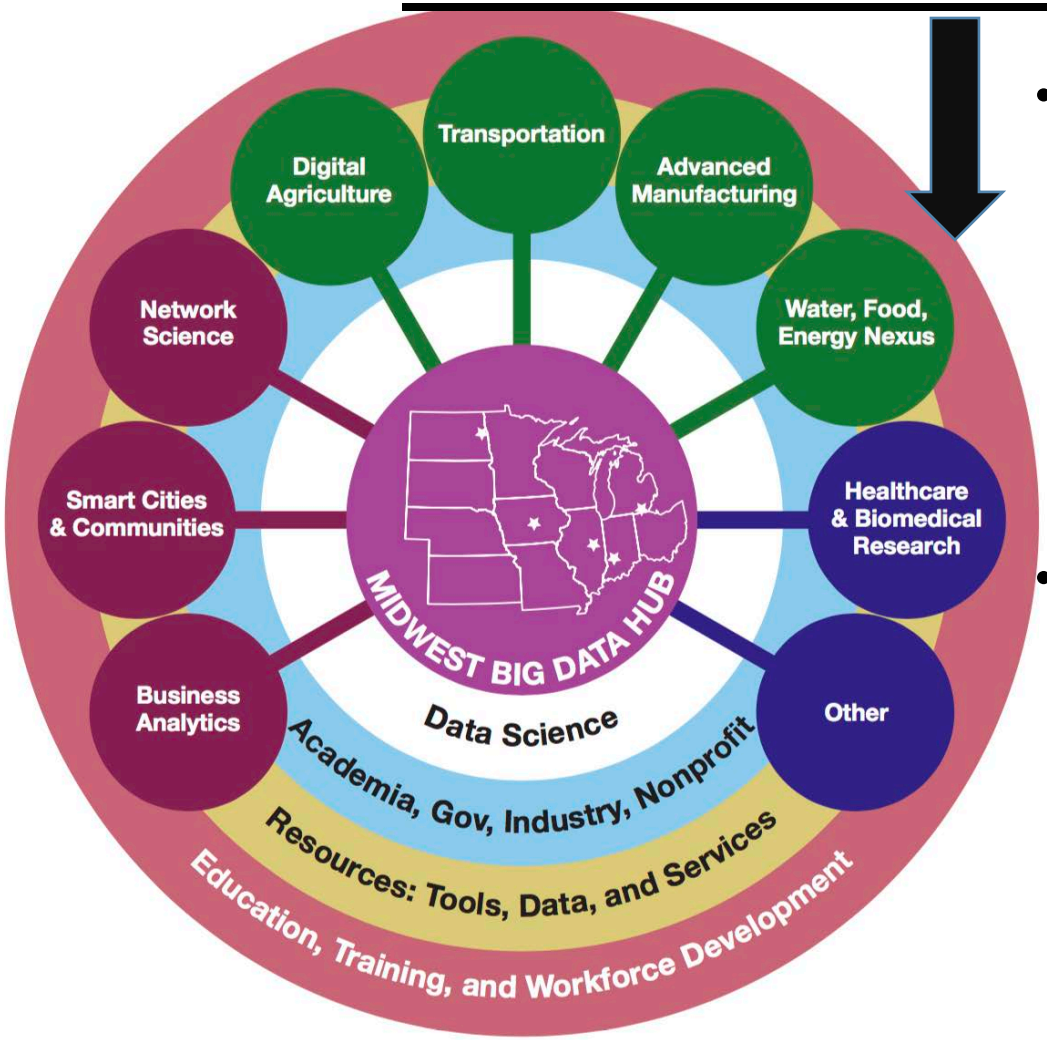





# **Collaborators** (Almost 40)

- Multi-sectoral
  - Industry: IBM, Climate Corp, …
  - Govt: Chicago Water District, USDOE:ANL, USGS, NCAR, …
  - NGO: Nature Conservancy, …
  - Academia: mid-west - Indiana U, NWU, Purdue, UIUC, UMN, NWU
  - Academia outside midwest: Texas A&M U, U of Glasgow
- Multi-Disciplinary
  - Big Data: TerraPop (Datanet), GABBS (DIBBS), CyberGIS
  - Water: NWU, Indiana, UIUC (Civil / Env. Eng.)
  - Food: AgMIP/GABBS (Purdue), UMN Food Protection & Defense Inst., …
  - Energy: NWU Inst. For Sustainability and Renewable Energy, …
  - Environment: Climate Corporation, NCAR, Minnesota Population Center, …
  - Education: UMN MS in Data Science, NWU STEM Education Partnership, …

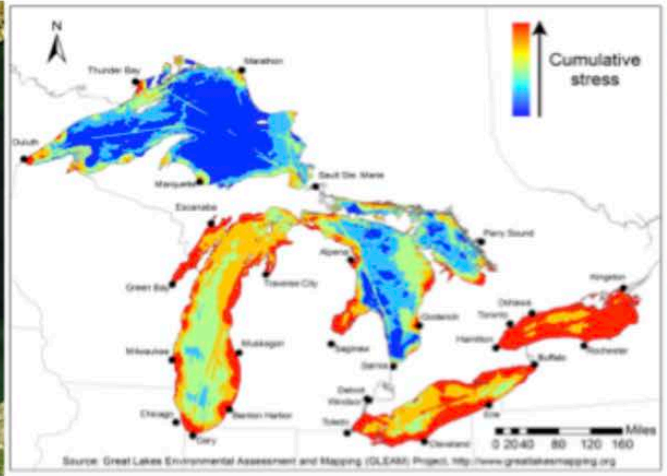
# Context: FEW Spoke

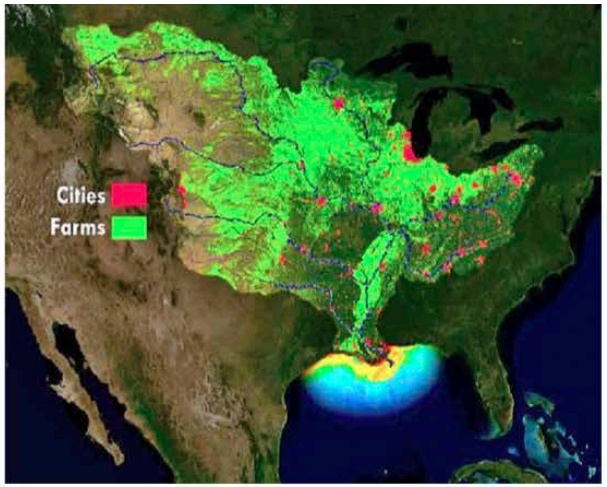



- Midwest
  - Water: Largest freshwater reserves, e.g., Great Lakes.
  - Food: Leader in agricultural production, processing, transportation, distribution
  - Energy: Dominant Biofuel Supplier
- NSF Cross-Directorate Initiative
  - Research: Innovations for Food, Energy, Water Nexus (INFEWS)
  - Education: NRT solicitation listed INFEWS as a priority
  - Infrastructure & Community Building: Our Spoke

# Motivation: FEW Spoke

- Piece-meal policies => unanticipated problems
  - Ex. Fertilizers affect Water quality (e.g., Great Lakes, Mississippi River)
  - Ex. Bio-fuel subsidy => Rise in food prices


- Crucial to understand interactions across Water, Food, Energy Systems
  - Not just for mid-west
  - National priority with initiatives from NSF, USDA USDOE, USGS, …
  - Global priority with initiatives from U.N., …

# NSF INFEWS Data Science Workshop

(@ USDA NIFA, Oct. 5th-6th, 2015; Shekhar, Mulla, & Schmoldt; www.spatial.cs.umn.edu/few)

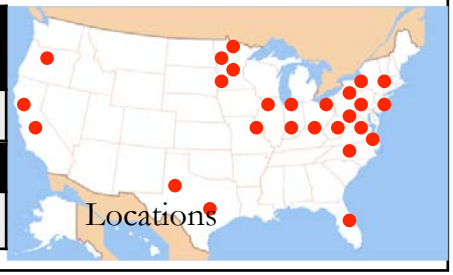
**Goals:**

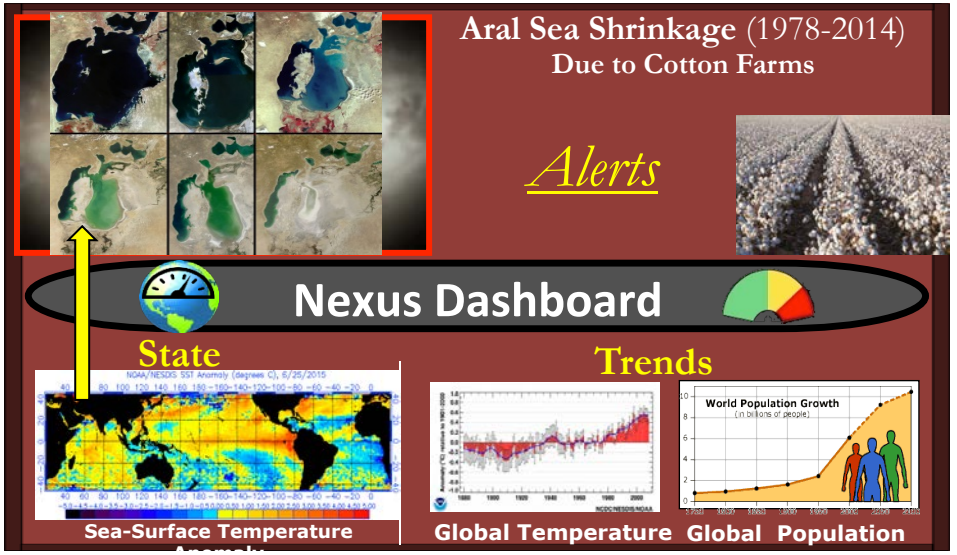
- Design compelling visions, Identify gaps
- Develop a research agenda

**55 Participants** (Data-driven FEW & Data Sciences)

| Food | Energy | Water | DataSc. |
|---|---|---|---|
| 14 | 10 | 11 | 20 |

| Gov. | Aca. | Industry |
|---|---|---|
| 26 | 24 | 5 |




**Finding 1: Data & Data Science are crucial!**

- Understand problems, connections, impacts
- Monitor FEW resources, and trends to detect risks
- Support decision and policy making
- Communicate with public and stakeholders

**Finding 2: However, there are show-stopper gaps.**

1. Data Gaps: No global water & energy census, Heterogeneous data formats & collection protocols
2. Data Science (DS) Gaps: Current DS methods are inadequate for spatio-temporal-network FEW data.

**Potentially Transformative Research Agenda:**

- National FEW Nexus Observatory & Dashboard for chokepoint monitoring, alerts, warnings
- Novel Physics-aware Data Science for mining nexus patterns in multi-scale spatio-temporal-network data despite non-stationarity, auto-correlation, uncertainty, etc.
- Scalable tools for consensus Geo-design via participative planning with nexus observations and policy projections
- An INFEWS data science community to address crucial gaps, and shape next-generation Data Science

Midwest Big Data Hub

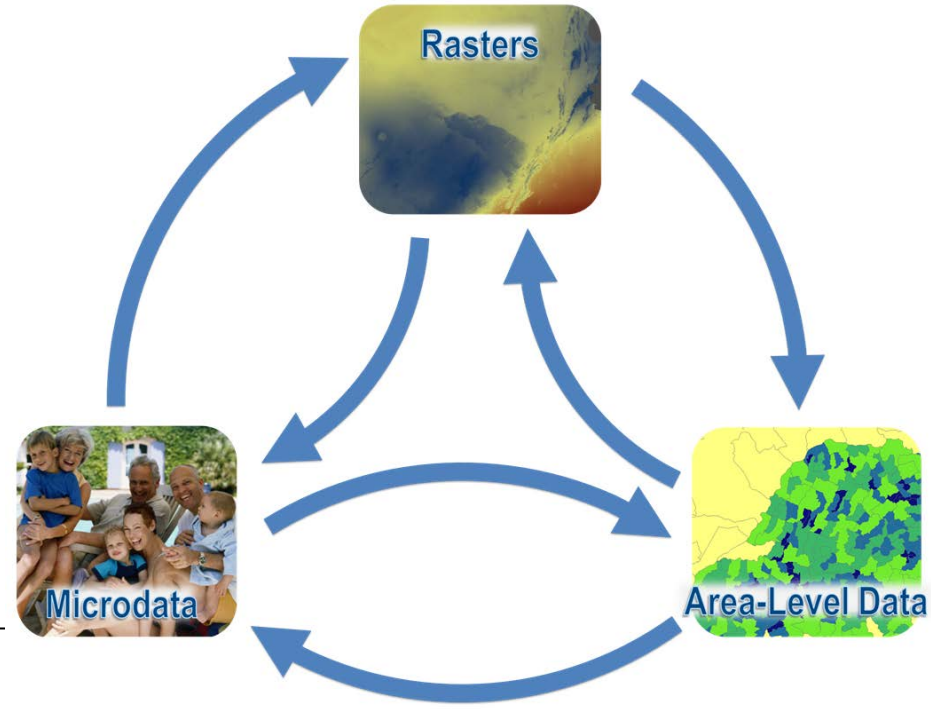
Accelerating the Big Data Innovation Ecosystem

# Related Work

- Challenge:
  - Spatial nature of FEW datasets.
  - Popular BD tools (e.g. MapReduce, Spark) inadequate for Spatial Data
- State of Spatial BD Tools
  - Siloed with small user communities
  - Ex. Minnesota Population Center: TerraPop
  - Ex.: GABBS for Agriculture Model Inter-comparison(AgMIP)
  - Ex. CyberGIS, SpatialHadoop
- Note:
  - Lightning talk on GABBS and CyberGIS
  - We welcome interest in spatial tools from all spokes

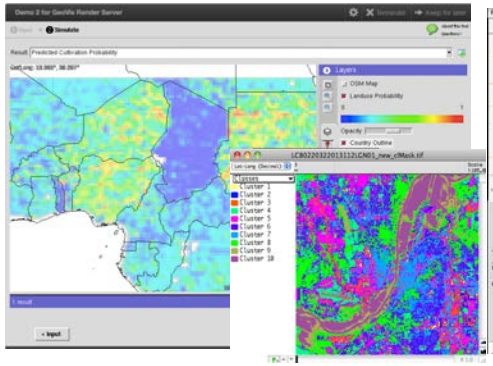
# Terra Populus: Integrated Data on Population & Environment

- Goal: Enable investigation of human impacts and vulnerability
- Integrated data related to
  - agriculture, land cover, climate, and population
- Curated collection
  - global population and environmental data
  - linked to locations
- Location-based integration
  - Raster
  - Vector
  - microdata

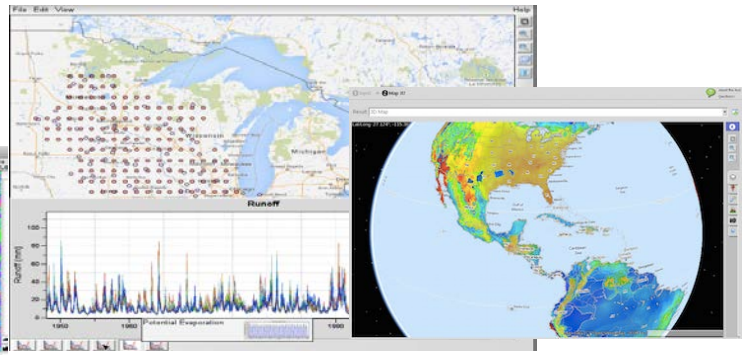


# G·A·B·B·S

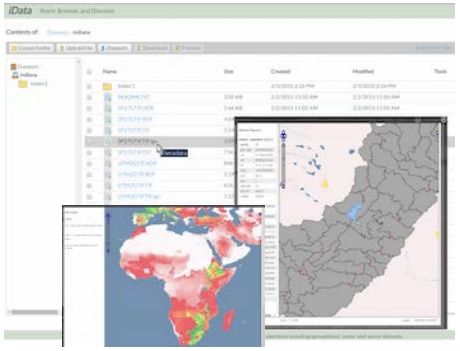
geospatial data analysis building blocks



Modeling & Analysis



Explore and visualize data

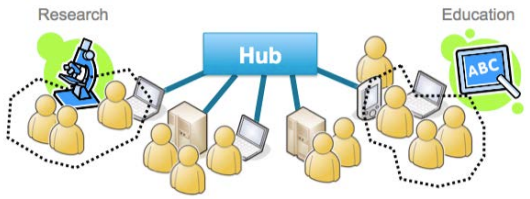


Share & Publish

## Geospatial data and computing building blocks – NSF DIBBS project

- Geospatial data processing, analysis and visualization support inside HUBzero
- Map library, Rapid Tool Development API (Rappture) with geospatial extension for developing online applications without web programming
- Online data management system linked to user tools
- DIY online interactive tool and data publishing (with DOI), publications linked to viewers and interactive tools

hubzero **Platform for Scientific Collaboration**




**Computing Resources**
High-performance
High-throughput
Cloud

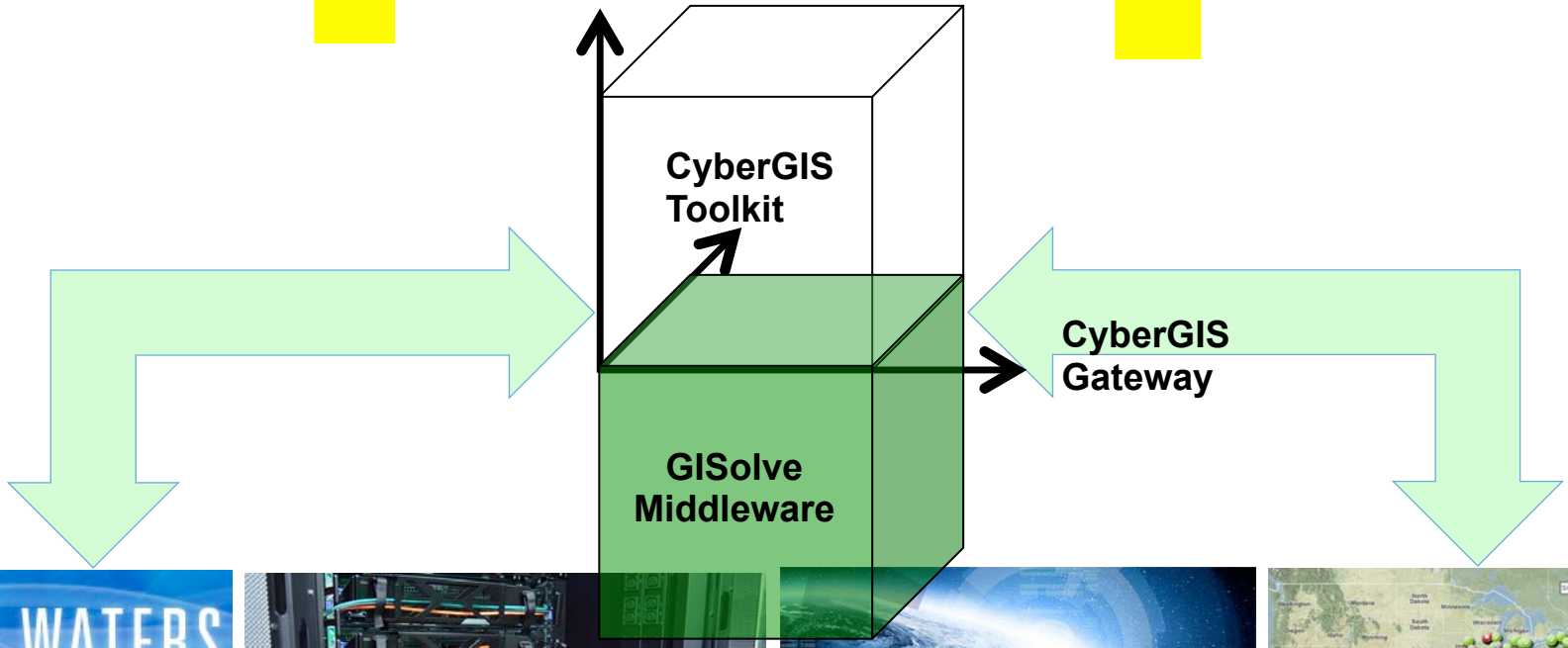
GeoVis
Renderer

Midw H

Accelerating the Big Data Innovation Ecosystem

# http://cybergis.org

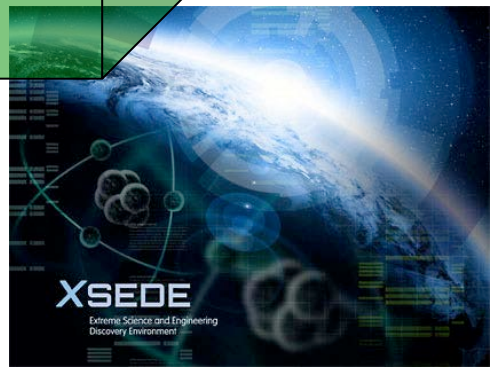
**Geospatial Discovery and Innovation**




BLUE WATERS
SUSTAINED PETASCALE COMPUTING

BlueWaters/


http://cybergis.illinois.edu

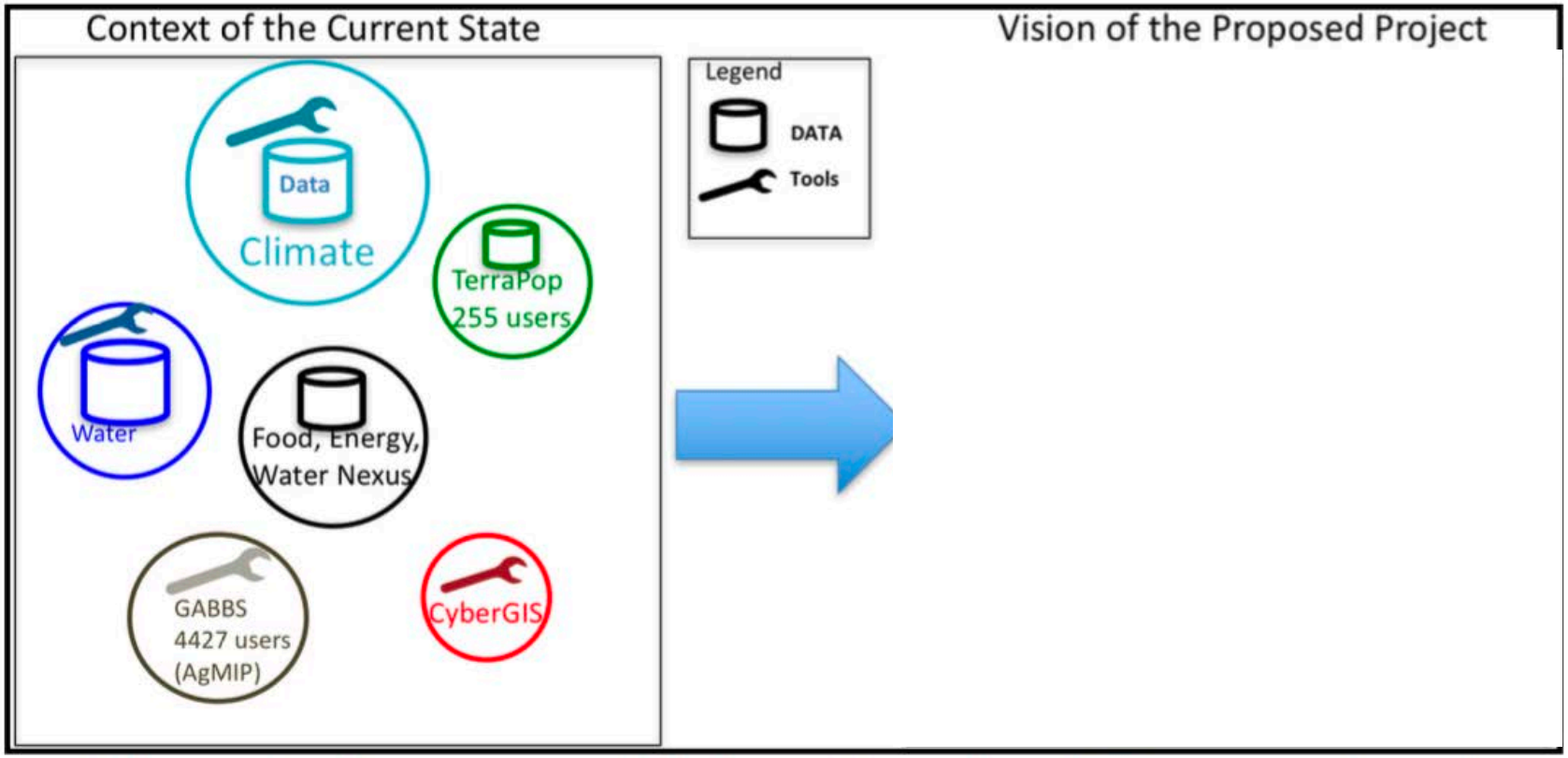



www.xsede.org

www.opensciencegrid.org

# Spoke Mission & Vision

- Grow and Connect Communities
  - Producers and Consumers of FEW Nexus Data, Tool, Services



# Data, Tools, Services

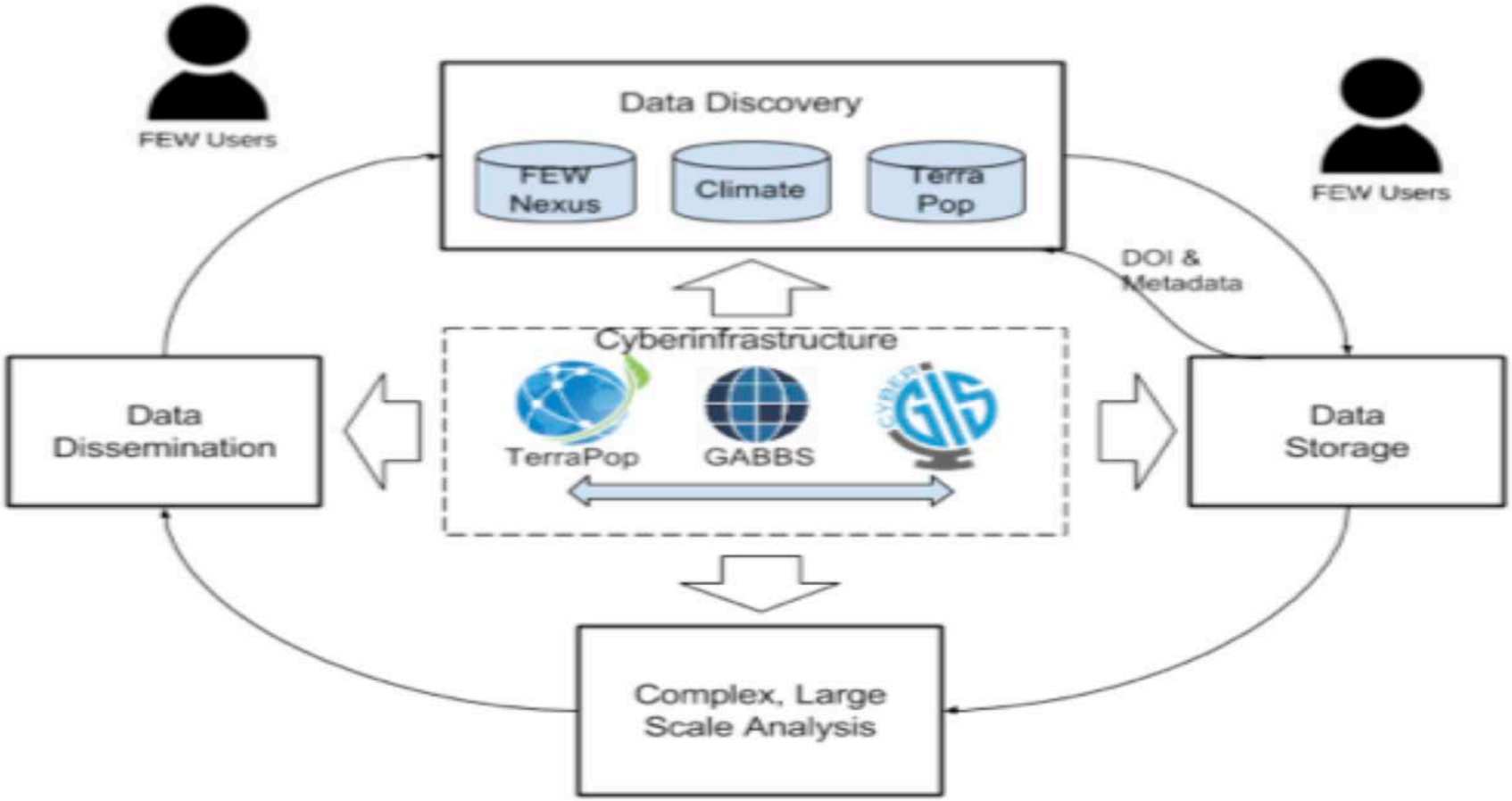



*Figure 4: Integration of TerraPop, GABBS, and CyberGIS (Best in color)*

# Activities: FEW Spoke

- **Past and Ongoing**
  - Capitol Hill Presentation, House Ag Committee Reception on “Deconstructing Precision Agriculture” (Shekhar, 3/15)
  - NSF Workshop to Identify Interdisciplinary Data Science Approaches and Challenges to Enhance Understanding of Interactions of Food Systems with Energy and Water Systems (Shekhar, 10/15)
  - Symposium S-E2 (Towards a Food-Energy-Water Nexus Data and Data Science Community”, NCSE 2016 Conference (Shekhar, Rodriguez, …, 1/16)
  - BDSpokes proposal Big Data Community for the Nexus of Food, Energy, and Water Systems (Core Team, 2/16)
  - MBDH/CCC Call for Proposal to engage Young Researchers (Nahrstedt, 3/16)
  - TerraPop, CyberGIS, GABBS education and outreach activities (Lightning Talks)
- **Upcoming**
  - CyberGIS Curriculum Workshop (Wang, 4/16)
  - Michigan State U Symposium on Climate-Food-Energy-Water Nexus (Shekhar, 4/16)
  - CCC/CRA Symposium on Computing Research: Addressing National Priorities and Societal Needs (Nahrstedt, Shekhar, 5/16)

| No | Expertise |
|---|---|
| 1 | BD: Industry |
| 2 | FEWS: Industry |
| 3 | FEWS: Government |
| 4 | FEWS: Government |
| 5 | FEWS: Local Governme |
| 6 | FEWS: NGO |
| 7 | FEWS: International |
| 8 | BD: Big Data Communi Building |
| 9 | FEWS: Industrial Affilia Program |
| 10 | BD: Educational Progra |
| 11 | BD: Educational Progr |
| 12 | FEWS: Climate Science |
| 13 | FEWS: Environmental Science |
| 14 | FEWS and Big Data Science |
| 15 | FEWS: Water Science |
| 16 | FEWS: Sustainability |
| 17 | FEWS: Water Science, Hydrology |
| 18 | FEWS: Sustainability ar Energy |

| | Expertise | Name |
|---|---|---|
| 19 | FEWS: Food, Water Science | Kaiyu Guan |
| 20 | FEWS: Industrial Affiliate Program | Brian Herm |
| 21 | FEWS: Industrial Affiliate Program | Thomas He |
| 22 | FEWS: Food, Environment Science | Matthew H |
| 23 | FEWS: International, Industrial Affiliate Program | James W. Jo |
| 24 | FEWS: Educational Program, Economics | Madhu Kha |
| 25 | FEWS: Food, Agricultural Science | Amy Kirche |
| 26 | FEWS: Water Science, Hydrology | Praveen Ku |
| 27 | BD: Climate Science | Vipin Kuma |
| 28 | FEWS: Food, Agricultural Science | Stephen P. |
| 29 | FEWS: Food, Water Science | Venkatesh |

| | Expertise | Name | Affiliation |
|---|---|---|---|
| 30 | BD: Big Data Computation | Claudia Neuhauser | Minnesota Supercomputing Institute, University of Minnesota |
| 31 | FEWS: Educational Program – STEM Study and Outreach | Amy D. Pratt | Office of STEM Education Partnerships, Northwestern University |
| 32 | FEWS: Educational Program | Stephen W. Searcy | Department of Biological and Agricultural Engineering, Texas A & M University |
| 33 | FEWS: Environmental Sustainability Industrial Affiliate Program | Timothy Smith | NorthStar Initiative for Sustainable Enterprise, Institute on Environment, UMN |
| 34 | FEWS: National Laboratory | Seth W. Snyder | Argonne National Laboratory |
| 35 | FEWS: Energy, Water Science | Ashlynn S. Stillwell | Department of Civil and Environmental Engineering, University of Illinois |
| 36 | BD: Educational Program | Goce Trajcevski | Department of Electrical Engineering and Computer Science, Northwestern University |
| 37 | BD: Geospatial Big data Analytics | Raju Vatsavai | Center for Geospatial Analytics, North Caroline State University |
| 38 | FEWS: Water Science, Sustainability | Peter W. Voorhees | Northwestern-Argonne Institute for Science and Engineering |
| 39 | FEWS: Social and Behavioral Science | Moira Zellner | College of Urban Planning and Public Affairs, University of Illinois |