Midwest Big Data Hub Supplement Plan
Please see the MBDH Supplement Summary for background on the Midwest Big Data Hub and additional rationale for selection of these activities.

## 1. Organizational Development

a. *Governance*: In spring, 2017, the Steering Committee will charge subcommittees to develop and formalize various aspects of the MBDH organization, which will include sustainability planning. We anticipate that these working groups will require coordination and technical support, as these activities will occur throughout this next year and beyond. The MBDH will be adding a part-time Academic Professional to its local staffing, to serve in a program management role; part of their work will be to support the working groups.

b. *Sustainability, Strategic Planning and Evaluation:* As we continue our strategic planning activities next year, it is essential that we develop an evaluation strategy that will address both programmatic and long-term decision-making. The MBDH will add a Graduate Hourly to the local team, who will work with the Executive Director through the next academic year to develop a knowledgebase and network maps of the Midwest region, which will be used to support outreach and sustainability strategy development.

We will also bring on an expert evaluator (new sub-award) with proven experience in evaluating large cyberinfrastructure and big data projects, who will work with us on strategy as well as direct evaluation activities. In additional, she will meet with the Executive Directors of all four Big Data Hubs, to facilitate a framework for identifying shared high-value metrics that address the common aspects of our respective missions.

One goal for the MDBH is to employ a robust independent evaluation designed to provide formative information to guide program improvement and summative assessment of program quality, effectiveness and impact. Much of the first-year activities will focus on developing an evaluation plan with accompanying methods and metrics.  The evaluation is designed to address four concepts:
1. **Implementation**: Are MDBH activities being implemented on schedule and as planned?
2. **Effectiveness:** Are key components of MDBH operating effectively? How might they be improved?
3. **Impact:** What near and longer term outcomes are associated with MDBH?  What is the value-added of MDBH to key stakeholders, the region and the nation?
4. **Sustainability:** How and to what extent are elements of MDBH becoming institutionalized? To what extent will aspects of MDBH be sustained beyond NSF funding? What opportunities and barriers exist?

Lizanne DeStefano, Professor of Psychology and Executive Director, Center for Education Integrating Science, Mathematics and Computing (CEISMC) at the Georgia Institute of Technology, will lead the evaluation activities which will include design of evaluation processes aligned with the MDBH strategic plan, focus groups and interviews with key stakeholders to develop metrics, benchmarks and targets, administration of a stakeholder satisfaction survey and event evaluations. The evaluator will work closely with MDBH leadership to ensure that evaluation findings are integrated into planning and decision-making.  An evaluation plan, interim report and final report will serve as deliverables, and a timeline of actions and milestones will be developed in summer, 2017.

2. **Extending the reach of the MBDH Network and Building Regional Big Data and Data Science Capacity**

The Great Plains Network (GPN) initiated the ENCITE (ENabling CyberInfrastructure via Training and Engagement) project to help broaden participation in cyberinfrastructure activities on campuses throughout the Midwest. The project was funded by Award No. 1440774 from the National Science Foundation to Kansas State University (PI: Dr. Greg Monaco). Two main aspects of this project will be utilized in this effort:

First, GPN partnered with six state networks to increase the project reach to smaller campuses across a six state region. This partnership increased the reach of the project from 28 GPN higher education members to 179 higher education members, including 9 tribal colleges and 4 HBCUs. The project demonstrates that it is possible to achieve increased reach through strategic partnerships with state networking organizations.

Second, GPN staff have learned that personal visits to smaller campuses to meet with faculty interested in advanced cyberinfrastructure are more likely to engage significant numbers of students and faculty than are centrally hosted meetings that require time and travel on the part of the same faculty to attend. GPN has also learned through these visits that, contrary to initial expectations, the number of faculty at smaller campuses who require data storage and data access for their research exceed the number of faculty who require access to computational resources. The MBDH will support new activities that will extend big data and data science networks and expertise in the region by targeting geographic areas and constituencies where there is a need for much greater engagement.

a. *Great Plains Network Expanding Carpentry Training Across the Midwest Big Data Hub*

The Great Plains Network, a regional partner of the Midwest Big Data Hub, has conducted a pilot project to expand regional access to Software Carpentry and Data Carpentry training. GPN has leveraged its Carpentry Foundation membership to have individuals from regional higher education institutions become certified Carpentry instructors. These individuals are then able to conduct Software and Data Carpentry training at their home institutions and, more importantly, at smaller and less well-resourced campuses that will benefit from Carpentry instruction. Dr. Greg Monaco, the GPN Director for Research and Cyberinfrastructure Initiatives, has coordinated this effort. Dr. Monaco is also a MBDH steering committee member. The Midwest Big Data Hub will be providing direct travel support for Carpentry trainers to attend and lead training workshops.

b. *Lawrence Technical University (LTU) will launch a planning project: "MBDH Resource Center for Big Data at Small Colleges"*

**Overview**
While most big data research activities take place at large research-intensive universities, smaller higher education institutions fulfill the duty of training the majority of the IT and data science workforce. The Midwest Big Data Hub (MDBH) provides a unique opportunity for smaller universities to be involved with big data education and research, and expose students majoring in computer

science, statistics, mathematics, and related fields to big data training, aiming at developing practical solutions to real-world problems. That unique and critically important experience will be pivotal in the preparation of trainees for becoming big data experts in the present and future job market. Predominantly undergraduate universities and two-year colleges of higher education normally have a dominant teaching agenda, but often lack access to real-world data, as well as access to a variety of domain experts and research labs. In many cases, smaller colleges and universities also lack appropriate computing facilities allowing effective analysis of very large databases. In order to provide effective data science education and train professionals for solving real-world data science problems, these gaps should be closed by providing smaller institutions of higher education with access to domain experts, real-world data science problems, education modules, data science tools, data infrastructure, and computing facilities.

We propose to use the Midwest Big Data Hub as a platform to identify pathways and model approaches to close these gaps, by establishing the Big Data Small College Resource Center (BDSCRC). The goal of BDSCRC is to provide essential resources that are not available normally to small colleges – and to leverage findings from the research conducted by the Great Plains Network and others to develop effective and durable access and use of resource networks and data infrastructures.

**Goals:**
The proposed work will focus strictly on network development and collaborative planning of the Big Data Small College Resource Center. These activities will include the participation of faculty and administrators from small colleges (two and four year), to ensure that the current and near future needs of these institutions are met. In the end of the period, we plan to submit a full grant proposal that will support the activities of the research center. The resource center will become part of the MBDH and will be used as a bridge for small and predominantly undergraduate institutions who wish to participate in the MBDH. Preliminary goals of the resource center include the planning of the following:

a. Data and computing resources: Small colleges normally do not have substantial computing clusters, as the research bandwidth of predominantly undergraduate institutions do not justify the cost of the installation and maintenance of powerful computing facilities. Therefore, we propose to plan collaborative approaches to access and partner with existing regional and national computing resources, as well as exploration of computing needs particular to small colleges. Such a resource center could optimize the use of research resources, providing all member institutions with their data science needs yet without the requirement of highly expensive facilities.

b. Doman experts: In addition to computing facilities, small colleges are often limited by the availability of domain experts in different disciplines with which they can collaborate. We propose to plan effective practices of communication between research institutions and predominantly undergraduate institutions, allowing the resource center to serve as a hub where methodology researchers and domain researchers can meet and establish collaborations.

c. Industry connections: The MBDH includes many industry partners, and the resource center will allow small universities to utilize Hub mechanisms and activities to connect with these partners to learn about industry needs, and for information and access to internships and long-term employment for their graduates. Our planning activities will include developing effective connections between small universities and industry partners of MBDH.

d. Collaboration with large, research intensive universities: Through communication and events with the MBDH, small universities will be able to learn about REU opportunities, educational programs,

and potential research collaboration that fits the needs of small colleges. This will include access to data science education modules developed at institutions across the MBDH network and beyond.

e. Real-world data: To practice real-world problems, the resource center will also facilitate access to real-world data available through regional universities, government, and industry members. Data sets being brought to the attention of the Hub such as transportation data provided by automakers, or smart city data, can allow students from small universities to design and implement big data solutions using real-world data on actual problems.

f. Inclusiveness in data science education: As predominantly undergraduate institutions train the vast majority of college students in the United States, inclusiveness and diversity in data science is a matter of high priority. Inclusiveness and diversity in data science will be a priority in the efforts of the resource center. (See Broader Impacts)

**Broader impacts:**
The proposed activities are of a clear inclusive nature, aiming at providing access to students who do not normally have access to substantial Big Data resources. Unlike the REU mode, which provides access to a selected number of students through participation in research activities, the resource center will be designed in a fashion that provides access to a broad population. It will also aim at providing professional and career development opportunities for faculty at a large number of institutions, who may have very limited options to develop their career in basic or applied research.

Planning and development activities going forward will also include the participation of the TIDES (Teaching to Increase Diversity and Equity in STEM) network, which aims at increasing diversity in the field of computer science. The TIDES participants will give seminars at the meeting to present their work and experience in increasing diversity and equity in computer science. Dr. Shamir, who is leading this MBDH activity at LTU is a Principal Investigator on a TIDES award and an active participant in the TIDES network. Resource Center planning will include participants from community colleges and two-year institutions which serve under-represented and minority student populations.

**Timeline:**
In the summer of 2017, we will hold an open meeting of representatives from a variety of institutions of higher education. That will include public and private four-year predominantly undergraduate institutions, liberal art colleges, two-year institutions, minority-serving institutions, as well as representatives from industry and research universities. In that meeting we will identify shared needs and goals related to data infrastructure and data science capacity, as well as potential paths to forming a sustainable network.

Planning activities will continue via on-line meetings to refine the network's mission, goals, and practices. A smaller meeting is planned during the MBDH All Hands Meeting, and will involve participants from all other spokes and hub members, and a presentation is planned on efforts to date. Another meeting will take place in the spring of 2018 on the campus of one of the participating institutions, preferably in the south west part of the MBDH region. That meeting will be used to finalize the mission and plans, for the preparation of a full spoke proposal that will take place during the summer.

**Milestones:**

| Summer 2017 | First meeting with broad participation |
|---|---|

| Summer 2017 – Fall 2017 | On-line meetings and definition of the mission and goal of the resource center. |
|---|---|
| Fall 2017 | A smaller meeting at the MBDH AHM. |
| Spring 2018 | A second meeting of the resource center participants. |
| Spring 2018-Summer 2018 | Finalizing the definition of the network/resource center, its goals, governance, resource management and sharing practices, etc.; and a proposal for an appropriate solicitation will be generated |
| End of summer 2018 | Final report submitted to the MBDH |

## c. *University of North Dakota: Big Data Partnership Summit and Software/Data Carpentry Supplemental Funding*

**Overview of Activities**
The University of North Dakota will host a series of Big Data events from June 2017 to Sept 2018. The primary activity is to co-host a regional Big Data Partnership Summit (BDPS) with the University of Minnesota – Twin Cities (UoM). This big data partnership summit arose out of an on-going conversation between the North Dakota research universities and UoM, which seeks to build collaborative research activities across the northern part of the Midwest region. The Summit will focus on building regional industry, academic, government partnerships around the data lifecycle in regards to unmanned aircraft systems, remote sensing, and other shared areas of research that align with MBDH priority areas (e.g. Food-energy-water; data science). Industry partners and government agencies who regularly work with UND and UoM will be invited to attend and present.

In addition, attached to this larger endeavor we will launch a Software/Data Carpentry instructor training program, sponsored by UND for regional tribal college and community college faculty and a series of coordinated carpentry workshops. The reach of the Summit will be enhanced by a co-located, in-person Software/Data Carpentry instructor training program.

**Roles and Responsibilities:**
The Big Data Partnership Summit will be held twice, once in the late fall of 2017, and again in the late summer of 2018. Each time it will be hosted on the UoM campus, with participant support each year for up to 17 attendees of this event provided by UND through the MBDH SEEDCorn award and this supplement. UND staff (including Grant McGimpsey, UND SEEDCorn Co-PI, Aaron Bergstrom, UND MDBH Project Coordinator, and Evan Boucher, UND Project Specialist) will serve on the organizing committee for the summit. UoM staff from the Minnesota Supercomputing Institute (MSI) and the Institute for Mathematics and its Applications (IMA) will also serve on the organizing committee. As the Summit will be hosted on the UoM campus, IMA will take on the role of primary Summit organizer.

As part of this supplemental funding request, UND will purchase a Gold-Level membership in the Software/Data Carpentry organization. UND MBDH Project Coordinator (A. Bergstrom) will serve as the main UND contact to the Carpentry organization, managing the training of carpentry instructors and scheduling of the coordinated carpentry workshops. An Advisory Committee will be established, with Bergstrom as Chair and Boucher as assistant chair. Faculty and staff from the University of South Dakota, University of Nebraska – Omaha, and the University of Minnesota have already committed to serve, and we will seek representatives from at least two of the Tribal Colleges in the region. The

Advisory Committee will assist in the recruitment of carpentry instructor trainees from regional tribal and community colleges, as well as other small universities. The Committee will establish review criteria for selection, which will include a letter from their college approving on-going participation in future training activities. The Committee will review applications and make selection for the trainee program. The committee will also establish a proposal process for distributing three of UND's gold-level carpentry organized workshops for delivery by trainee instructors.

**Broader Impacts**
The Summit and Carpentry activities will target the states of North Dakota, South Dakota, Nebraska, Minnesota, Wisconsin, and Iowa. It will bring together regional research institutions, tribal colleges, community colleges, government agencies, and private industry for a discussion around the application of big data technologies as they related to unmanned aircraft systems, remote sensing, robotics, and education. The carpentry training and workshop activities will bring a discussion of advanced cyberinfrastructure to rural and tribal communities, and including them in the larger regional collaborative discussion.

**Outcomes and Evaluation**
The direct outcomes of the project will be to foster a regional awareness for the need to develop collaborations in the area of big data as it relates to unmanned aircraft systems, remote sensing, and robotics. This regional collaborative environment will be based on the growing collaborative relationship between the North Dakota University System and University of Minnesota. In addition to a locally housed instructor, UND will sponsor the training of 14 tribal and community college Software/Data Carpentry certified instructors, who will be to deliver carpentry workshops across ND, SD, NE, MN, WI, and IA.

Project Coordinator Bergstrom will work with IMA to develop a post-summit survey to be distributed to all Summit attendees. The results of the 2017 survey will be used to determine the structure of the follow-on Summit in 2018. Bergstrom will work with the Carpentry organization to collect carpentry workshop assessments, which will provide feedback to the newly trained instructors. All of this information will be included in the final SEEDCorn report to NSF.

**Timeline and Milestones:**

| Summer 2017 | • Purchase Gold Level Institutional Membership in Software/Data Carpentry<br>• Launch Advisory Committee<br>• Big Data Partnership Summit (BDPS) dates set for Summit in Minneapolis.<br>• UND disseminates Carpentry Instructor Training Solicitation to Tribal Colleges and Community Colleges within the states of North Dakota, South Dakota, Minnesota, Wisconsin, Nebraska, and Iowa |
| --- | --- |
| Fall 2017 | • Host North Dakota Data Carpentry event<br>• Host Nebraska Software Carpentry event<br>• BDPS hosted on Univ. of Minnesota campus.<br>• Co-located Trainer Gathering for in-person carpentry instructor training held |

| Winter 2018 | • Advisory Committee meets to select Coordinated Workshop awardees |
|---|---|
| Spring 2018 | • Host South Dakota Data Carpentry event<br>• Communicate with Carpentry Organizing entity to collect assessment results for workshops held so far<br>• Reconvene BDPS organizing committee to choose late summer date for 2018 event |
| Summer 2018 | • BDPS hosted on Univ. of Minnesota campus |
| End of summer 2018 | • Final report submitted to the MBDH |

### d. *Midwest Big Data Summer School (Iowa State University - Summer, 2017)*

This summer school will be designed as a one-week, intensive curriculum aimed at early career researchers to get them started in data-driven research. The school will include full-day lectures from Iowa State University (ISU) faculty and faculty from other institutions, on topics ranging from: data acquisition, data preprocessing, exploratory data analysis, descriptive data analysis, data analysis tools and techniques, visualization and communication, ethical issues in data science, reproducibility and repeatability, and understanding of domain/context. The summer school will also include hands-on data analysis experience.  Post-event, program curriculum and training materials will be made available online.

**Benefits for the Early Career Researchers**
The proposed summer school will have an immediate positive impact on the skill-sets of our summer school participants. Getting started with data-driven research is difficult due to expertise required in a large number of data science related areas, and this summer school is designed to directly address this difficulty. We will advertise the activity broadly, especially to smaller schools in the Midwest, encouraging early career researchers to attend the summer school, and with special focus on encouraging minority and underrepresented researchers to attend. The skills acquired during the summer school will have a direct impact on data-driven research and educational projects being conducted by all of the participants and their collaborators.  Last but not least, the intense weeklong participation in team-based learning activity is likely to start some collaboration among early career researchers within ISU and also among Midwest PIs.

**2016 as a Successful Model**
The summer school builds on our success organizing a similar school in the summer of 2016 that was broadly attended. The first edition of this summer school was organized in June 2016. 144 individuals attended the school. Participants were from across 19 universities and 10 organizations. Based on the self-reported data, 32% reported no prior experience with data science, 40% reported that they were beginners in the area, and 24% reported that they had intermediate-level expertise. After the summer school, a survey was distributed; of survey participants, 88% reported they were very satisfied or satisfied with the event and training.

**Roles and Responsibilities**

The Midwest Big Data Summer School will be coordinated by an experienced ISU faculty member as the overall Program Lead. There will be a Curriculum Development Committee, and a Program Planning Committee. The Program Planning Committee will also address how to sustain the program and identify resources for ongoing programming after the 2017 session.

**Broader Impacts**
The Midwest Big Data Summer School provides unique training for early career researchers on data acquisition, data reprocessing, exploratory data analysis, descriptive data analysis, data analysis tools and techniques, visualization and communication, ethical issues in data science, reproducibility and repeatability, and understanding of domain/context. These activities are designed to enable students and researchers to begin to use these concepts and techniques to get familiar and comfortable with data-driven research. It includes a hands-on, in-person experience that is of core value to participants.

In making training and select curriculum from the Midwest Big Data Summer School available online, this tangible reference can also be accessed across the regional Hubs. This training will benefit many, and can be referred to both by those in attendance and those who wish to access remote training post-event. The broader impact includes availability of training to build and collaborate across the data science community, and provide and expand quality education in this area.

ISU will encourage from a pedagogical perspective, broadening participation of minority and under-represented students and early career researchers. In selecting participants for funding, priority will be given to the US citizens/resident than other US-based people in that order, with attention to underrepresented groups and attention to need (e.g., applicants not supported by a grant).

**Outcomes and Evaluation:**
Following the Midwest Big Data Summer School, reporting will be provided related to the list of attendees, their affiliation, and demographics. An evaluation of the program will be conducted by sending out a post-event electronic survey. The curriculum and training development, and program planning committees will come together prior to the event, to discuss a process for making online training and materials available, as well as to plan evaluation for online components. The teams will also meet post-event to discuss best practices and lessons learned from this program.

**Timeline & Milestones:**

| In progress | Program Planning (all elements) and Curriculum/Training Development |
| --- | --- |
| July 10-14, 2017 | Summer School Session |
| Fall 2017 | Post-event dissemination of materials Outcomes report to MBDH |

## 3. Program Development

The MBDH will continue to plan and lead annual "All Hands" meetings for the region. These events are highly valued by the range of constituents that attend. We have held two in Chicago thus far, and have had growing participation from academia, government, and industry. In 2017, we will move

west to Omaha, and partner with the University of Nebraska-Lincoln and UN-Omaha, Creighton University, and the Kiewit Corporation, which is providing space in their new Kiewit Innovation and Learning Center. In 2018, we plan to hold our "All Hands" meeting on the Eastern side of the region. The MBDH has budgeted participant costs for both meetings to support student and early career participants, as well as speakers. The Hub will continue to live stream these events, and provide edited video of the talks and proceedings, which provide a valuable resource and reference for Hub participants. ISU will provide these services for the 2017 meeting, which will include audio/video services to stream and record the October 2017 MBDH All-Hands meeting, including pre-production, production/equipment, post-production and video encoding/compression. Similar to previous MBDH events that have been coordinated via Iowa State University, the Brenton Center (http://www.brenton.iastate.edu/technology.html) will coordinate the A/V component in terms of supplying the equipment, coordinate recordings and post-event availability online, and staffing necessary to capture and record the All-Hands meeting. The video will be closed-captioned to allow for broad use.

Assessment processes around these All Hands events have provided valuable feedback from regional members and newcomers on data infrastructure needs, ideas about MBDH services, and particularly for emergent areas of shared interest for research and data science network development. We anticipate that All Hands meetings will continue to be a significant location for gathering stakeholder feedback. For example, we now have a growing community of researchers with state and local government partners who are spurring new MBDH activity in Data-driven Community Health. In addition, some of the many "spoke" communities identified in the original SEEDCorn grant are transforming as the conceptual framing of problems and priorities in the R&D space evolves. For example, there is great interest across our academic, industry, and government constituencies in the "Internet of Things," which will bring together our unmanned systems and sensors communities, the business analytics community, as well as transportation and smart cities groups.

In the large urban centers of the Midwest, many smart and "data-driven" activities and networks already exist, and as we engage with them we will continue to facilitate the participation of new people or organizations. In addition to cultivating relationships with organizations in Chicago and Kansas City, there are current plans for additional direct outreach to a number of data-related smart city groups in Detroit, Columbus, and Omaha. However, it is not surprising that there is strong regional interest to facilitate the development of networks and collaborations to address these same themes and priorities for the rural areas of the Midwest, where socio-eco-physical systems are different, and contexts for science and data-driven approaches may also be different. In addition to the community health area, we have included additional participant support funds that we will use for a large community-engagement event to bring together cross-sector and cross-disciplinary stakeholders with shared interests in addressing smart & connected problems for rural communities.

Finally, some funds will be allocated to participant support and travel that are as yet unallocated, to allow the Hub continued flexibility to support new opportunities and activities that continue to emerge as we develop new partnerships and collaborations.

Justification for Supplement and Overview of the Midwest Big Data

The benefit of this supplement is that it will allow us to accelerate organizational development with new activities specific to evaluation and sustainability; support targeted outreach and expand our data science education network to four year institutions; and to support strategic community engagement activities. Sustainability, regional "reach," and impact are intentionally addressed in the design of these activities.

The Midwest Big Data Hub (MBDH) is one of four Big Data (BD) Innovation Hubs launched in 2015 with support from the National Science Foundation. The BD Innovation Hubs are intended to strengthen the data ecosystem, and develop effective academic-industry-government-NGO networks to address scientific and social issues of regional and national interest. Based at the National Center for Supercomputing Applications (NCSA) at the University of Illinois in Urbana-Champaign, the MBDH has four academic associate sites at [Indiana University](), [Iowa State University](), the [University of Michigan](), and the [University of North Dakota](), and a growing network of academic and industry partners. The MBDH serves twelve states: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin.

The MBDH was launched to enhance the national data ecosystem through community engagement and activities that bring together diverse expertise, projects, and resources surrounding big data applications important in the Midwest. Our outreach and program efforts are focused in thematic and priority areas identified by our regional stakeholders, including: Food-Energy-Water; digital agriculture; transportation; metro sciences & smart cities; advanced manufacturing; materials sciences, health & biomedical sciences and neuroscience; network sciences; and business analytics. The Hub's mission is to, "Develop effective cross-sector, data-enabled networks to solve shared problems of regional and societal interest." To accomplish this mission, we focus our expertise and resources in four ways:

- We cultivate communities and resource networks within the region by convening physical and virtual meetings and attending diverse events; "match-making" direct connections amongst people, resources, and organizations; and we provide participant support and presenters' travel. These activities have served to accelerate partnerships and catalyze new research, and to generate many new collaborations for data-related proposals and (funded) projects.
- Developing and supporting programs and activities that build capacity in Data Science and Big Data. We work with partners to improve access to data tools & services; produce and host data-related training events, and support participation at events hosted by others. We engage with industry and government regularly on workforce development needs, and we are now engaging directly with data-oriented education programs at universities in the region.
- We also work on activities intended to reduce friction in R+D and Data-to-Decision systems. We are collaborating on pilot projects designed to enable access and use of data previously unavailable or difficult to move; and, we collaborate with projects and communities to find solutions for barriers to data sharing and access.

- We <u>collaborate with the other three Big Data Hubs</u> in several ways that benefit our respective Hub organizations and our regional constituents. The Executive Directors work together on projects and partnerships that have cross-Hub impact or national scale, such as our partnership with Microsoft Research to provide Azure credits to Hub-affiliated projects in our respective regions; or the upcoming National Data Challenge on reducing traffic fatalities. We also share expertise and best practices; we regularly make regional training activities and data science expertise available to members of the other three Hubs; and we connect researchers and organizations within our own regions to potential collaborators in the other Hubs when their interests are best aligned with a specific project or priority area in those other regions.

During the first 15 months of funding, our approach to outreach and partnership development has been flexible and opportunistic, allowing slow but systematic cross-sector network development, and the ability to take advantage of emergent and potential opportunities. We have engaged broadly across communities through regional and national workshops, campus visits and invited talks. These activities have led to several collaborations with pilot projects now underway (e.g. with Internet[2] and GLOBUS; and, with the Open Commons Consortium and the Center for Advanced CyberGIS), and newly planned events driven by community interest and assessment.

The MBDH has led and managed four events – two "All Hands" meetings, the national Data Science in Agriculture Summit (funded by NIFA), and Machine Learning: Farm-to-Table. (The latter event will be held April 18-19, 2017, is a collaboration with International Food Security at Illinois, which will bring together domain scientists from Agriculture, Bioinformatics, Food-Energy-Water, Food Security, and computational communities. The objective of this workshop is to stimulate new data-driven R+D activity at the intersections of these communities.)

Additionally, in the last year the MBDH has provided with matching funds and participant support to launch the "Midwest Big Data Summer School" short-course, and ten other workshop and hackathon events across the Midwest. These spanned Big Data (2), Food-Energy-Water (2), health sciences (2), community health (1), neuroscience (1), digital agriculture (1), industry engagement and partnering (1), and data quality (1).

In the coming 18 months, we will be focused on 3 priorities:
- Accelerating the development of a strategically-driven and sustainable organization. We will begin work with an external evaluator with deep experience in assessing cyberinfrastructure and big data-related projects, who will help us to develop an evaluation framework tied to our strategic and sustainability plans. We will also add a part-time staff member to work with the Executive Director on supporting the work of the MBDH Steering Committee.
- Extending the reach of the Midwest Big Data Hub. We will leverage conferences and invited meetings to engage with new groups in the region, and begin targeted outreach to organizations that are aligned with the MBDH thematic priorities but not yet involved.

Programmatic and planning activities will develop our network across four-year and minority serving institutions.

- Continuing program activities that build capacity in regional data science and facilitate cross-sector engagement around regional priorities. Workshops and other events will be designed to facilitate new cross-sector and multidiscipline collaborations, and to identify potential services or programs to address gaps in regional and national data ecosystem. The Midwest Big Data Summer School short-course will be held again; and, we will leverage memberships in Software Carpentry to develop a cohort of software and data trainers to reach constituencies where data science capacity is limited. Finally, we will also work with our academic community leads and industry partners to identify, develop and test scalable models for consortia or exchange-like services that will address shared needs. We see this as an important role for the regional Big Data Hubs, and aligned with our efforts to develop sustainable a business model in the Midwest.