MBDH Initiative: Data Science in Small Colleges
Lior Shamir, Small College Data Science Collaboration
Kansas State University

**Synopsis**

While large R1 institutions receive the vast majority of the available federal funding, these institutions train less than 15% of the national IT workforce. The majority of the students are trained at community colleges (40%), Carnegie Master’s universities (23%), liberal arts colleges (7%), and others. That is, according to the existing models of higher education support, the vast majority of the students do not have access to the resources provided by the federal government.

Moreover, non-R1 institutions serve many of the least privileged student populations: the financially disadvantaged student who takes courses at their local community college, the working student who can only take evening classes, the student who has childcare responsibilities and takes classes at the institution nearby his/her house, or the first-generation college student and the underrepresented minority student who often do not see an exclusive elite school as their natural next step after graduating from high school. These students need the intervention of the federal government more than other institution student populations, yet the institutions attend receive disproportionately low federal support.

The initiative of Big Data and Data Science at non-R1 institutions aims at changing the existing reality to provide equal access to all Data Science students. With the Midwest Big Data Hub, we aim at identifying ways of sharing resources and services between small and large institutions, including the needs of non-R1 institutions in decision processes, and designing a new paradigm of federal and academic partnerships. As a new discipline, we have the unique opportunity to design Data Science in an inclusive manner, embracing equity for all student populations, and creating an inviting and supportive environment for all genders and ethnicities.

**Challenges**

- **Industry Partnership**
  - Funding, including NSF funding for teaching and learning scholarship.
  - Small-sized awards.
  - Funding from smaller projects (smaller than a typical NSF award).
  - Databases fromindustry.
  - Research opportunities for research topics that do not have research programs at the school.
  - Access to industry scientists, including data access and high-speed support.
  - Partnership with R1 and non-R1 institutions; Partnership with R1 researchers for grant proposals.

- **Research**
  - Non-R1s do not have an effective office to communicate with industry partners.
  - Non-R1s need smaller grants that are not provided by the NSF.
  - Non-R1 needs grants for teaching and learning scholarship in data science.

- **Resources**
  - Access to affordable and qualified people is often difficult and expensive.
  - Distributed computing.

- **Education and Curriculum Development**
  - Communication of the importance of data science.
  - Curriculum development including courses and labs.
  - Teaching big data.
  - Development of interdisciplinary curriculum.
  - Development of courses for students and faculty.

- **Inclusion and Diversity**
  - Better resources, at the Center, Development Office in non-R1 institutions in order to get professors, teaching for professional opportunities in data science.
  - Collaboration with industry, government, and academic.
  - Overcoming barriers for teaching and learning in data science.
  - Need for collaboration between data science and other disciplines.

**Institutional Readiness**

- **Vision and Activities**
  - Create an active resource network.
  - Diverse faculty, curriculum development.
  - Student resources, faculty resources.
  - Diverse student populations, and creating an inviting and supportive environment for all genders and ethnicities.

- **Data Access**
  - Better access to data for teaching and learning.
  - Databases from industry.
  - Research opportunities for research topics that do not have research programs at the school.

- **Faculty Development**
  - Better access to data for teaching and learning.
  - Development of interdisciplinary curriculum.
  - Development of courses for students and faculty.

**Vision and Activities**

- **Research**
  - Non-R1s do not have an effective office to communicate with industry partners.
  - Non-R1s need smaller grants that are not provided by the NSF.
  - Non-R1 needs grants for teaching and learning scholarship in data science.

- **Resources**
  - Access to affordable and qualified people is often difficult and expensive.
  - Distributed computing.

- **Education and Curriculum Development**
  - Communication of the importance of data science.
  - Curriculum development including courses and labs.
  - Teaching big data.
  - Development of interdisciplinary curriculum.
  - Development of courses for students and faculty.

- **Inclusion and Diversity**
  - Better resources, at the Center, Development Office in non-R1 institutions in order to get professors, teaching for professional opportunities in data science.
  - Collaboration with industry, government, and academic.
  - Overcoming barriers for teaching and learning in data science.
  - Need for collaboration between data science and other disciplines.

**Data Access**

- Better access to data for teaching and learning.
- Databases from industry.
- Research opportunities for research topics that do not have research programs at the school.