Academic Program Development in Data Science
Office of the Provost and Executive Vice President
Mississippi State University

(Submitted: 12/10/2020 by Cindy Bethel and Jamie Dyer)

I. Mission and Charge Statement

Mississippi State University is a leading institution in research related to data science, and there currently exists a unique window of opportunity to leverage this research capacity into innovative transdisciplinary academic programs. At the direction of Vice Presidents David Shaw, Julie Jordan, and Reuben Moore, this task force was charged with exploring how MSU can best fulfill its land-grant mission of teaching, research, and service through the development and implementation of comprehensive transdisciplinary programs in the field of data science.

The task force was comprised of the following individuals representing a cross section of expertise related to data science and academic programs at MSU: Cindy Bethel (Chair); Mimmo Parisi; Wes Burger; Keith Coble; Shahram Rahimi; Seth Oppenheimer; David May; and Dan Peterson. Additionally, Jamie Dyer, Professor of Geosciences and administrative intern in the Office of the Provost and Executive Vice President, participated in and supported the task force from the Provost's office.

The primary objective of the task force was to outline the requirements, potential structure(s), and known/expected obstacles associated with the development of programs focused on data science. Driven by the expertise of the task force, and under the guidance of the Office of the Provost, the general objectives and goals included the following:

- A description of potential academic programs focused on data science that aligns with the strengths of MSU faculty, resources, and research foci.
- Potential curriculum structure and content based on existing and potential MSU course offerings.
- An outline of potential program structures (e.g., certificate, degree, etc.) and their individual advantages/disadvantages.
- An overview of existing and anticipated obstacles in the development and/or success of a program in terms of administrative, economic, political, or other factors.

Given MSU's existing research and academic expertise in Data Science, as well as available resources (i.e., the High Performance Computing Collaboratory (HPC^2), Bagley College of Engineering, Data Science for Social Science (DS3) Lab, etc.), the University is in a strong and unique position to become a leader in Data Science education by providing students with a strong foundation in data science *ethics, literacy*, *competency*, and *applications*. Program development with a focus on these four tenets will lead to the recruitment of students who will become competitive graduates with marketable skills.

II. Program Recommendations

Summaries of the task force recommendations are provided below in order of preference and/or time to implement, as defined by the task force. Additionally, the associated justifications for each recommendation are provided to further define the expected impact and potential structure of associated academic programs.

- **Survey course in Data Science** - *Existing Development:* There is an existing class in the Honors College that can be used as a foundation for this recommendation. The existing class has no prerequisites and was designed to allow students from any major to be exposed to the topic of data science. The class was also designed to help students interested in Data Science to pursue an academic path (Masters or Ph.D.) or a professional path (industry recognized credentials, certification).

  The proposed class is about the basic knowledge needed to become a wise consumer of data science, a well-rounded, data-driven individual with a deep appreciation for the value of data and how data may change the way we live our lives. Students will be exposed to methods that allow the use of data with a special focus on data and predictive analytics, machine learning and artificial intelligence, high performance computing, the cloud, data governance, and data security. The ethical issues raised by these new technologies will be addressed as an important aspect of the course. The overall goal is to provide students with an overview of how data are turned into assets for improving the delivery of goods and services and for unlocking trends and knowledge necessary to make strategic decisions. Students will have the opportunity to develop the ability to identify the appropriate questions, the data necessary to address those questions, and the methods used to produce insights from these data. They will also gain the basic leadership skills that are required to bring all the necessary resources together in the Data Science lifecycle (how data are captured, processed, maintained, analyzed, and communicated.) In the course, students interested in pursuing a career in data science will have the opportunity to understand the fundamental skill sets sought by employers and the required general competencies in the current economy.

  This course could be taught by faculty from a variety of disciplines and could be cross listed across several departments. Ideally, faculty experts would serve as guest lecturers based on their experience and expertise. Another option would be a hybrid format with sets of expert lectures available on Canvas, with the instructor running class discussions on the topics raised by the videos in a required face-to-face class. This approach, combined with sets of common required readings, would allow for a broad class of disciplinary faculty to participate, bringing their own expertise and experience to the course while building off common material.

- **MS in Data Science** – *Existing development:* The current proposal for a MS in Data Science involves a core of 15 credit hours in Statistics and Computer Science courses offered by both the Math and Statistics (ST) and Computer Science and Engineering (CSE) departments. There will be an additional 15

hours of course work and a three-hour capstone project that would be offered by different specialization tracks to allow students to follow either a "general" track with additional ST and CSE courses or a specialty track such as Agriculture or Business. Other specialization tracks such as Social Science, Education, Electrical, and Computer Engineering can be developed to leverage the core training. Students in specialization tracks would do a capstone project related to the specialization. The proposal includes both on-campus and online delivery modes. The online version of the course is a particularly good opportunity to expand the academic capabilities of Data Science at MSU.

- **Minor/Concentration/Academic Certificate Program in Data Science** – *New Development:* The task force is united in its belief that MSU needs to provide either a minor/concentration at the undergraduate level and/or an academic certificate (MCAC) in data science at the undergraduate and/or graduate level. We believe that this MCAC program would be attractive to students from a wide array of disciplines in the natural and social sciences, business, agriculture, and education as well as others. Though not an exhaustive list, the task force has identified several faculty members across the university who are actively participating in some form of data science (see Appendix A); thus, the greatest challenge in a minor is likely not in the development of the courses or encouraging faculty to teach them, but in organizing the minor in a comprehensive, logical way that appeals to all undergraduate disciplines. This recommendation would require further details to be developed. For a minor/concentration in Data Science at the undergraduate level, students would be required to complete one basic course (e.g., the survey course proposed in the first recommendation) and several specified courses. The course credit hours would need to be determined, but based on similar programs, range between an additional 15-20 hours in this specialization with 9+ credit hours at the 3000-level and above. This would be the same for either a minor or concentration.

  For the academic certificate program, the certificate could be at either the undergraduate or graduate level. This would also require a foundational course like the proposed survey course in the first recommendation and then additional coursework at the 3000-level or above for an undergraduate certificate or at the graduate level for a graduate certificate in data science. The number of hours needs to be determined but typically ranges from 15-20 credit hours of specialized courses in data science.

- **BAS in Data Science** – *New Development:* The BAS in Data Science would provide an applied pathway for students transferring from community colleges to obtain a Bachelor of Applied Science (BAS) building from an Associates in Applied Science (AAS) in Data Analytics or Data Science. Currently, some community colleges (e.g., Mississippi Gulf Coast Community College [MGCCC]) has an AAS in Data Analytics and is considering the addition of an AAS in Data Science. They have stated that they are willing to adjust curriculum to meet the necessary requirements for students to be prepared from a mathematics perspective to complete a BAS in Data Science. The

challenge is ensuring that the BAS classes have sufficient applied and practical problem-solving assignments to meet the criteria for a BAS in Data Science. Some concerns were expressed by the task force on whether this recommendation would be feasible and whether the students could be properly prepared, but it is expected that this challenge can be addressed. The development of the curriculum and the identification of specific prerequisite courses and knowledge would need to be investigated and then implemented. It may also be important to include an internship or similar activity for practical hands-on experiences with industry as part of this program. We feel this will be an attractive degree program for industries in need of applied data scientists.

- **Data Science Professional Certificate Program** – *New Development:* The task force has identified the need for industry professionals to perform ongoing professional development through coursework. It is believed that providing professional development courses and a certificate program in Data Science for industry professionals would be beneficial to both the university and industry in Mississippi and across the United States and abroad. The courses could be the same as those taken by MSU students, especially those in the BAS program mentioned in the previous recommendation. The course credit hours and the identification of the specific applicable courses would need to be determined. The credit hours would likely range from 15-20 credit hours, with some type of experiential learning course included as part of the curriculum.

- **Community Outreach and Education Programs** - *New Development:* Data science efforts can extend beyond the on-campus classroom through the MSU County Extension Offices. This initiative can leverage on-campus instruction to engage with non-traditional learners through Extension programs. Examples could include non-credit certificate programs, outreach to K-12 schools, learning opportunities for rural communities, and engagement with industry in the state. Training related to data science applications in forestry, healthcare, agriculture, and small business represents some obvious opportunities associated with this recommendation.

  New Development: Through the Research and Curriculum Unit (RCU), outreach activities can be provided to the community through educational programs at the K-12 levels with different community schools. Certification programs can be made available to the public who may be interested in how to analyze data associated with their small businesses or other needs. A part of this may include the establishment of a Data Science Academy, similar to the Mississippi Coding Academy for workforce development and retraining. There is a shortage of trained professionals in data science and this could be an opportunity to provide members of different communities with these skills. Providing knowledge and information associated with data literacy and interpretation is critical, especially in terms of data presented in media and other related venues. This program could provide entry-level college credit that could be applied later to a degree program for life-long learning or related educational opportunities.

New Development: (Consultancy) Beyond educational programs, MSU could provide assistance and consulting to local governments and Mississippi businesses. The task force views this as leveraging the expertise of Mississippi State in the service of the economic development of Mississippi as well as in making local government more efficient and effective. How this will be funded is yet to be determined, and is a point of future discussion.

- **Public Policy –** *New Development:* An emphasis on data science can expand the services that MSU provides to the public sector and it's vision in the area of public policy. More than ever, public officials are making data-driven decisions about policy and resource allocations, linking budget decisions to measurable outcomes (performance-based budgeting). Economic competitiveness is now linked to a state's ability to use data in demonstrating strength in economic and workforce development to prospective businesses. Data science also plays a role in personalizing education and health services. Mississippi State has developed internationally recognized public-sector programs such as NSPARC, the Research and Curriculum Unit, the Social Science Research Center, and the Stennis Institute, that make contributions in this area.

- **International Partnerships –** *New Development:* A focus on data science can expand and strengthen the global outreach of the University. MSU presently enjoys an international reputation for its focus on food security, and has already established a strong partnership with the UN Food Agricultural Organization (FAO) and the World Food Organization (WFO). Drawing on Data Science, both organizations are investing in the development and implementation of techniques and methods to improve service delivery, evaluation of programs, and research initiatives. To be sure, MSU can become a key player in helping FAO and WFO achieve their mission.

  Mississippi State University (MSU) has also established student exchange programs with several European universities where students gain experience in fields such as business, agriculture, and engineering. With a focus on data science, European students might see MSU as an attractive place to come and expand their knowledge and experience in data science. European universities have been called to have a much greater role in filling the growing demand for data scientists in Europe. MSU can be a key player supporting many other universities, thereby expanding its footprint in Europe. Under the "Next Generation" program in Europe there is a stronger emphasis on lifelong learning in data literacy. The general idea is to offer online courses at three levels: (1) basic knowledge and methods of data science, (2) mathematics, and (3) specialized areas of application such as health and education. MSU's distance education programs can become a resource to European Universities in their effort to offer online courses in the fields of Data Science. More recently, MSU has signed a Memorandum of Understanding with the International Data Science Foundation in Bologna, Italy to expand the use of high-performance computing in the areas of smart cities, health, and economic development. MSU, with its long-time investment in high-performance computing (HPC) and autonomous vehicle research

(CAVS), can play a key role in the application of high-performance computing in Europe.

- **Academic Institute** - *New Development:* It is known that university-level research institutes and centers are catalysts for economic growth, research productivity, and intellectual creativity. It is the opinion of the task force that MSU would benefit from the eventual formation of a campus-wide data science institute charged with the mission of fostering MSU's growing expertise in data science research, education, and outreach. In terms of general organization, we suggest a scheme based on the operation of MSU's Institute for Genomics, Biocomputing & Biotechnology (IGBB) (see www.igbb.msstate.edu). Of note:

  a. The data science institute would be led by a director with considerable theoretical and applied expertise in data science. The director would be responsible for supervising the data science institute staff (see below) and would be skilled at putting together multidisciplinary research teams to pursue data science funding opportunities.

  b. The institute would employ three or more highly trained core associates (CAs). As in the IGBB model, each CA would have expertise in a slightly different area, with the CA team collectively having expertise in all areas of applied data science. The data science CAs would receive roughly 50% of their pay from high-level university stakeholders, e.g., VPs and colleges. Such hard funding would allow the CAs to serve as "free" consultants charged with helping any MSU faculty/staff/student members needing advice on how to achieve a particular data science goal. The remainder of the salaries of the CAs could come from one of two models (or a blend of both).

     1. In the first model, MSU faculty who wish to hire one or more of the CAs to conduct data science research on funded projects would be able to buy a portion of a CA's "time and effort." This model is utilized by the IGBB.
     2. In the second model, each CA would be hired as a tenure track faculty in an academic department with a joint appointment with the institution. This model will follow similar models used in the Honors College and African American Studies. Each CA would teach one course per semester in their respective data science area. Such course offerings would help strengthen the minor and/or certificate programs in data science when they become available.

     Using the CA approach would allow MSU employees to benefit from the institute whether they are just in the initial stages of considering a project, are applying for funding for a project, or have funding in hand and need data science assistance.

  c. Any faculty interested in being associated with the data science institute would be accepted as institute affiliates. Financial investment in the data science institute through grant procurement, employment of data science CAs, etc., could merit higher levels of institute association (e.g., institute fellows).

d. The data science director would answer to those VPs, deans, etc., that contribute to the salaries of the CAs.

e. Affiliation of the data science institute with the High Performance Computing Collaboratory (HPC^2) would be advantageous. It is our understanding that the HPC^2 is developing a way in which MSU employees can utilize certain supercomputing resources without being members of an HPC^2-affiliated

f. center. It is expected that this will be developed through a "pay-to-play" service center. Whether the proposed data science institute becomes an HPC^2-affiliated center (like ICRES/CAVS, GRI, NGI, IGBB, ASSURE, CCI, and CCS) or simply a "pay-to-play" user should be a topic of future discussions.

III. Potential Challenges

The task force identified a common set of challenges associated with the Data Science program developments discussed previously. These include:

Defining credit/ownership.
- Where will the program be housed? Who will oversee administrative tasks associated with the program (i.e., application and approval processes)?
- How will resources be allocated given faculty involvement across multiple academic units?

Funding and resource allocation.
- What will be the funding stream for the programs, and how will funds be collected/allocated among academic units?
- Who will provide the initial resources for program development?
- If required, who will provide the physical space to house the program?

Faculty recruitment and recognition.
- How will faculty involvement in potential transdisciplinary programs be recognized for promotion and/or tenure?
- How will faculty be recruited into potential programs, either as new hires or as associates from existing academic units across campus?
- Faculty implications associated with teaching responsibility in Data Science programs may be an issue, especially if there are limitations to class size and teaching resources.

Defining niche program focus areas based on MSU expertise.
- What existing courses could be used to develop academic programs in Data Science? What courses could be developed with existing faculty expertise and availability?
- As programs in Data Science are already in place at other peer and peer-plus institutions[1], the task force recognized that competition with other universities could be an issue.

[1] https://datascience.uark.edu/

Identifying and developing external partnerships.

- How can industry, government, and academic partners be recruited and utilized in Data Science program development?
- How can current and future graduates connect with external partners for internships, capstone projects, and job placement?

IV. Paths Forward

Due to the wide-ranging breadth and complexity of the program recommendations set forth by the task force, it is expected that the time to implementation will vary for each program. Additionally, many of the programs can be developed through a tiered approach, allowing for developments in one area to be shared with another program. Finally, some of the goals that are desired - such as being a resource for industry, state and federal agencies, and other partners - will develop over time. What has been proposed will not happen until we begin training people and start doing coordinated science, demonstrating broad relevance, and proving applicability. At that point, academic efforts will allow for technology transfer beyond MSU to partners and stakeholders.

Short-Term Goals: (< 1 year)

1. Inventory of existing courses that could be utilized for a Data Science program.
2. Moving forward with the initial undergraduate survey course already developed.
3. Moving forward with the M.S. in Data Science Program that is already developed and needs to be finalized and processed through UCCC and IHL.
4. Create an internal to MSU Academic Institute for Data Science with Provost Approval. (We recommend the establishment of a task force for investigating all issues and concerns associated with the academic institute.)

Medium-Term Goals: (1-4 years)

1. Investigate the allocation of responsibility and credit for different academic departments participating in this program.
2. Investigate how promotion and tenure would be addressed across disciplines.
3. Investigate financial and other resources that would need to be invested for these efforts to be realized.
4. The creation of a university-wide IHL-approved Academic Institute for Data Science including the support funding. The Academic Institute would enable and oversee the other proposed recommendations discussed in this document.

Long-Term Goals: (4+ years)

1. Integration of teaching, service, and outreach components associated with the Academic Institute for Data Science that would have an integration of academic and research efforts.