DiSCuS

A proposal for the Baylor University Institute for Data Science and Cyber Security
Data science applies to all areas of research conducted at Baylor University. And our success is increasingly tied to our ability to solve large data intensive problems while ensuring the security of data information.

Data breaches now cost an estimated $475 Billion per year, increasing about 20% per year. Cybersecurity must move from afterthought to core consideration to combat this trend. Accomplishing this requires the establishment of thought-leaders/researchers in cyber and innovative education.

All major research universities have developed core Data Science initiatives.
University of Michigan to spend $100 million over the next 5 years in a new Data Science Initiative (DSI), including 35 new faculty.

In September, 2017, Vanderbilt received $71 million in initial NIH funding to explore data science and big data over five years.

In 2017 Rice announced a $45 million in initial funding (out of a $150 million commitment) to support Data Science.

Columbia University Data Science Institute is expected to have ‘$3.9 billion in overall activity over the next 3 decades’. From a 2016 press release: ‘Columbia will begin the development of the first of two phases of Institute immediately, first creating 44,000 square feet of new … facilities in existing buildings by August 2016 for the Institute for Data Sciences and Engineering. … Columbia will hire 30 new faculty members as a part of the first phase, and ultimately expects to expand the Institute faculty to 75 by 2030.’
Federal Spending

2017 FY Federal Grants Expenditures or Budgets for Data Science related research (in 1000 dollars)

- **NSF** 38.6% 1,254,043
- **NIH: BD2K, CIT, NCATS,** 47.7% 1,550,000
- **CDC, DOD** 13.5% 438,000
- **DOE** 0.2%

*NSF & DOE include only funded project from 01/01/2017 - present. DoD indicates procurement budget for Big Data research. NIH numbers are FY 2017 budget allotments.*
DiSCuS Drivers

- **Scalable computation is now a required component of life science, physical science, and business analytics.** This includes infrastructure and software expertise. An institute will meet these requirements for a variety of academic units, preventing the *ad hoc* duplication of computational efforts within the university, and increasing our critical mass in computation while reducing overall costs.

- **Finite intramural and extramural resources require a focused growth model.** Organizing computation and training within a singular resource will allow intentional growth of academic units towards emerging areas of computationally based research by leveraging successful, and ongoing Data Science research currently being conducted at Baylor University. For example, *Joint Baylor-U.S. Army Research and Education Programs* produce vast amounts of data that require **secure data storage and analytics** and currently must satisfy these requirements outside of Baylor.
DiSCuS Drivers

- **It will provide a focal point for external collaborators to commit expertise that immediately impact translational problems.** External partner researchers joining DiSCuS as adjunct Baylor faculty can shape future success through their interactions with graduate students, institute researchers, and other collaborating institutes.

- **Computational fluency in managing and analyzing large datasets is now a requisite for success in STEM, business, and life science domains.** The proposed institute will enable hands-on training in these areas through workshops, short courses, formal classwork, and the development of joint research projects. Students with computational backgrounds will require experience with applied computation in external research areas.

- **Baylor University has the opportunity to shape a new generation of Data Science professionals.** Through the initiation of an executive Master’s of Science degree program in Data Science (MSDS), Baylor can serve to train current information scientists to meet the emerging market needs for well-trained Data Science professionals.
DiSCuS Drivers

- **The trend of analyzing high volume consumer data is driving an increase in the demand for cyber security workers with hybrid skill sets.** Baylor’s security programs are uniquely placed to provide students with the business skills, regulatory knowledge, and data analytical and technical competencies needed to support industrial demand.

- **An increasing number of federal funding agencies require demonstrable computation resources to be competitive.** This includes a critical mass of experts in applied computation and an institutional infrastructure in place to support computation-based research.

- **Baylor University will need to have a proven resource to attract and maintain high quality research faculty.** As graduate students across all areas become increasingly computationally savvy, access to high quality computation resources and infrastructure that supports transdisciplinary research will be critical for success. Baylor has the opportunity to position itself to attract the highest quality faculty and graduate students.
Benefits of an ‘Institute’ Organization

- Cross-school interdisciplinary relationships.
- Flexible allotment of resources, particularly in the creation of Centers and research focus, aimed at rapid adaptation.
  - Institutes, at their best, can rapidly adapt to accommodate emerging trends or shift priorities from stale initiatives.
- Ability to retain non-Tenure Track Research Faculty (a new paradigm to the Baylor campus).
- Institutes are more easily dissolved if the research environment changes.
Institute Administration

- **Director, Associate Director**
  - Curriculum Coordinator
  - Administrative Staff
  - Research Staff
  - Research Faculty

- **External Advisory Board**

- **Baylor Governing Board**
  - Report to EVP
  - Dean’s meet yearly to provide feedback.
  - Five Year Extension Model
Industrial Contracts
Corporate sponsors regularly fund laboratory space and R&D.

Diversified Revenue Streams
Funding for DiSCuS will rely heavily on external funding to grow and meet the expectations of a rapidly evolving digital landscape. To ensure sustainability, we will focus on a combination of professional education, short courses, granting, and industrial cooperatives.

Federal and State Granting
NIH BD2K and NSF Big Data and CyberSecurity Initiatives account for >$3 Billion per year in funding.

Professional Education
There is an immediate need of a professional degree in Data Science to retrain existing IT & CS specialists in this area.

Short Courses & Certificate Training
Cybersecurity lends itself to certificate training. Given Waco’s geographic location we can establish ourselves as a leader.
Baylor University Institute for Data Science and Cyber Security

DiSCuS

- Professional Master’s of Data Science (MDS)
- DiSCuS Education and Research Coordinator
- Short Course and Certificate Training
- Baylor University Associated Faculty: Data Science, Bioinformatics, Business, Mathematics, Statistics, Biology, Geology, Engineering, Environmental Science, Information Technology
- External Collaborators: Waco Family Health, Scott & White, Baylor College of Medicine
- Scalable Resource Sharing

- Adjunct Research Faculty
- Dedicated Research Faculty and Graduate Students
- External Advisory Panel
- Internal Baylor Oversight (Deans, EVP)
- DiSCuS Director, Assoc. Dir, Staff

- Center for Business Analytics
- Center for Competitive Learning
- Center for Medical Informatics

Future Expansion of Domain Based Centers
Infrastructure Requirements

- Office space to house up to 10 new faculty members, including administrative space.
- Graduate office research space for 30 - 50 graduate students.
- Data science lab and cooperative space: collaborative research space, conference rooms, facilities for small - medium lectures.
- Computing infrastructure. Most of the infrastructure can extend existing Baylor compute and storage servers, but some cloud based internal systems will be required.
Graduate Students

- Graduate lines funded by DiSCuS will receive degree from the home departments.
- The Institute will provide funding fellowships, laboratory space, and transdisciplinary mentoring.
- Only faculty members associated with DiSCuS will be able to apply for graduate student fellowship hours.
- And it is expected the concentration of degrees shared among these graduate students will shift to meet departmental needs.
7.5 million dollars over 10 years. This is the initial amount requested as part of the pro forma budget. Five (5) and ten (10 ) year outcomes will be measured against this initial support.

15 million dollars over 10 years. Five and ten year outcomes will be measured against this initial support. This approach will keep 5 million dollars as an endowment in addition to the 10 million dollar operating budget.

30 million dollar initial investment. Five and ten year outcomes will be measured against this initial support. This approach will keep 15 million dollars as an endowment in addition to the 15 million dollar operating budget.
# Doctoral Student Production

<table>
<thead>
<tr>
<th></th>
<th>Fiscal Year Funded</th>
<th>Fiscal Year Funded + $5 M endowment</th>
<th>$15 M Initial Funding + $15 M endowment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total # of Ph.D. student funding years (over 10 years)</td>
<td>30</td>
<td>42</td>
<td>55</td>
</tr>
<tr>
<td>Expected avg # of Ph.D. graduates per year after year 5</td>
<td>3</td>
<td>4.5</td>
<td>6</td>
</tr>
<tr>
<td>Initial number of Ph.D. students</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Total number of Ph.D. graduates at year 10</td>
<td>15</td>
<td>19</td>
<td>25</td>
</tr>
</tbody>
</table>
Scholarly Output

Expected Institute Publications per Year

- .75 M/yr
- 1.5 M/yr
- 3.0 M/yr

Year 1, Year 2, Year 3, Year 4, Year 5, Year 6, Year 7, Year 8, Year 9, Year 10
### Additional Opportunities

<table>
<thead>
<tr>
<th>Research Faculty</th>
<th>Fiscal Year Funded</th>
<th>Fiscal Year Funded + $5 M endowment</th>
<th>$15 M Initial Funding + $15 M endowment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>4 (+1 Research Staff)</td>
<td>6 (+ 2 Research Staff)</td>
</tr>
<tr>
<td>Post-Doc Researchers (grant funded)</td>
<td>0-3</td>
<td>2-4</td>
<td>4-6</td>
</tr>
<tr>
<td>Professional Education</td>
<td>Yes, with no Substantial Variation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undergraduate Education</td>
<td>No explicit goals</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Institute Revenue

Yearly Expenditures (in thousands of dollars)
Institute Revenue

Cumulative Revenue Surplus (in thousands of dollars)
Authors

Erich J Baker, Ph.D.
Professor, Department of Computer Science

Michael J. Donahoo, Ph.D.
Professor, Department of Computer Science

Greg Hamerly, Ph.D.
Associate Professor, Department of Computer Science

Chris Pullig, Ph.D.
Professor and Chair, Marketing Department

James Stamey, Ph.D.
Professor, Department of Statistics

Randal Vaughn, Ph.D.
Professor, Department of Information Systems