# 7 K H  3 D V W  3 U H V H Q W  D Q G  ) X W X U H of Data S
SMU

2 F W R E H 20 20,

Contents

# A Proposal for the Future of Data Science at SMU

## Introduction

In the fall semester, 2016, Provost Currall assembled a Task Force[1] to assess the potential role of Data Science at the University. The Task Force met on November 18, 2016, and established the following objectives:

1. Establish a working definition for 'Data Science'.
2. Inventory relevant SMU degree programs.
3. Inventory SMU centers/institutes involved in Big Data, Data Science, and data analytics.
4. Collect information on successful programs at other universities.
5. Advance recommendations for the future of Data Science at SMU.

This report GHVFULEHV the findings of the Task Force LWV UHFRPPHQGDWLRQV DQG WKH ZLWK LPSOHPHQWDWLRQ RI WKH WKRVH UHFRPPHQGDWLRQV

## What is Data Science?

The analysis of huge datasets to generate insights and support decision making requires special tools that make up the emerging field of Data Science, including mathematics and statistics, computer programming, data management, machine learning, artificial intelligence, geo-spatial analysis, and visualization to extract and communicate meaningful information from enormous and complex data sets with an emphasis on predicting and optimizing outcomes. In this report, education and research described as business analytics or data analytics are included within the

## Why Data Science ?

The rate at which information is being produced is expanding exponentially.  In 2013, Science Daily reported that 90% of all the data in the world had been generated in the preceding two years.[2] Current project

capabilities and computing power

corporations are potential future employers for SMU graduates in Data Science as well as sources of funding for education and research in Data Science and internships.

- The Federal Reserve Bank of Dallas is partnering with SMU and other north Texas universities to establish a Federal Statistical Research Data Center , which will provide access to otherwise inaccessible Federal data.
- ManeFrame II is now operational and provides computation power to SMU researchers second only within the state to The Advanced Computing Center (TACC) at the University of Texas.

Weaknesses:

- Lack of seed funding is a major impediment to advancing a Data Science initiative at SMU. It

- Increased corporate engagement

## Recommendations

The Task Force offers the following recommendations based on its assessment of Data Science initiatives at other universities and its SWOT analysis of SMU's strengths, weaknesses, opportunities and threats. Appendix 9 summarizes the considerable contributions that these recommendations will make, if implemented, to the goals and objectives of SMU's 2016-2025 Strategic Plan.

1. Create a Data Science Institute (DSI) charged with (i) coordinating and facilitating interdisciplinary programs in Data Science, (ii) collaborating with other institutes and centers on campus, (iii) expanding research funding, and (iv) developing research and educational linkages, including strong internship and placement programs, with Dallas corporations, the Federal Reserve Bank of Dallas, the University of Texas Southwestern, and the City of Dallas and surrounding local governments. The Task Force strongly recommends that the DSI be adequately staffed and housed in the Ford Building, and that the DSI Director report to the Provost. A governance structure, seed financial resources, and a business model should be formally established before the DSI is launched with consideration given to:
    - Creation and management of joint faculty appointments with departments.
    - Management of indirect cost recovery.
    - Moving the on-line MS in Data Science to the DSI to immediately provide a supporting income stream.
2. Create an interdisciplinary undergraduate program in Data Science. A bachelor's program would be the first in Texas and serve the exploding demand for expertise in the DFW Metroplex and beyond. A minor could serve a larger number of students and could be step stone to the creation of a bachelor's degree.
3. Stand up an interdisciplinary PhD in Data Science to be operated by the DSI (Appendix 7) with the objective of becoming a top-10 national program.
4. Hire a minimum of 10 tenured/tenure track faculty who will complement the Data Science research of the schools of the University and who will contribute to the teaching courses and directing dissertations of the PhD in Data Science program operated by the DSI.
5. Endowment of the DSI and the PhD program in Data Science should be a major SMU Development priority.
6. Add staff to facilitate access to ManeFrame II . SMU currently supports one FTE to assist faculty and students in parallel computing. A total of at least two FTE should be devoted to this task to enable expanded access to ManeFrame II by faculty and students.
7. Advertise SMU's current programs and strengths in Data Science . As soon as possible, create and maintain a Data Science website that summarizes and points to Data Science degree programs, research, and investments across the University. When appropriate, issue press releases on investments such as ManeFrame II, and initiatives in Data Science.

# APPENDIX 1
## Data Science Programs at Aspirational Peer Institutions

| | |
|---|---|
| Boston College | xConcentration in Business Analytics |
| Brandeis University | xNew MS in Strategic Analytics offered by the Division of Graduate Professional Studies and focused on business analytics. |
| Carnegie Mellon University | xNumerous degree programs at the masters level grouped and introduced in single website |
| Emory University | xNell Hodgson Woodruff School of Nursing's Center for Data Science focuses on utilizing data to improve health outcomes.<br>xEmory continuing education offers 10 day sessions on Big Data and Data Analytics.<br>xComputer Science offers an MS in Computer Science with a Data Science concentration.<br>xGoizueta Business School offers MS in Business Analytics. |
| Lehigh University | |

the Office of the Provost, the School Deans, and the Center for Data Science codifies important relationships such as faculty compensation and course buy-outs, the Center's role in faculty tenure and promotion decisions, and the distribution of indirect cost recovery.

The Center for

# APPENDIX 3
# SMU Degree Programs Involving Data Science

| School | Department | Plan | Graduated Enrolled AY 2017 |
|---|---|---|---|

# APPENDIX 4
## Known Employers of SMU Graduates in Data Science or Analytics

| School | Department | Known Employers |
|---|---|---|
| Cox School of Business | Information Technology and Information Management | Accenture; PwC; American Airlines; Deloitte; Full Contact; Hitachi; Innovate Auto Finance; Money Gram; Think Finance; Worldlink; Albertson's; American 1st Finance; Artis Consulting; AXA Advisors; Baidu Waimai; Bain; Bluehost; Bonjour Mineral Group; BrainScale Inc.; Brierley Partners; Capital One; CheapCarribian.com; Chickasaw Nation Industies; Draper Labs; DTC; Enforce; Equity Metrix; EY; Goldman Sachs; Grand Peaks Real Estate; Greyhound lines; Hotels.com; Infosys; Intuit; ISNetworld; Kohl's; Light Beam Health Solutions; LoopBack Analytics; MedSynergies; Mercer; Nautilus Hyosung America; Pacific Union; PCIC (Primary Care Innovation Center); PeopleAdmin; Red; ROKT; Santander; SAS Institute; SparkHound; Texas Health |

# APPENDIX 5
## Centers and Institutes Representing Potential Resources for a Data Science Initiative

| Center/Institute Name | Center/Institute |
|---|---|
| | Adma72.1 35.5555 5 /I 668.76 <</r08 43aT /Pt3 (I 6o)10 n</MCID 5 >>BDC 6656 |

| | | | |
|---|---|---|---|
| | | Estate Finance and the Folsom Institute for Development and Land Use Policy. | |
| Don Jackson Center for Financial Studies | Cox School of Business | The Don Jackson Center for Financial Studies provides enrichment programs and research opportunities for students and faculty in the Cox School's Finance Department. | Research Collaboration<br><br>Education |
| EnCap Investments & LCM Group Alternative Asset Management Center | Cox School of Business | The Alternative Asset Management Center provides students with hands-on training in the increasingly important financial markets of non-traditional asset classes, including hedge funds and private equity funds. | Research Collaboration<br><br>Education |
| Kitt Investing and Trading Center | Cox School of Business | The Kitt Investing and Trading Center is a state-of-the-art Cox | |

Tsai Center for

UW. The Institute does not focus entirely on public policy problems; in fact it appears to address problems across a broad range of disciplines (escience.washington.edu/). (In fact their byline is “Advancing data-intensive discovery in all fields.”) Their summer internship program, Data Science for Social Good, began  in 2015.

The Georgia Tech program is affiliated with their Institute for Data Engineering and Science,. This center and appears to be less interdisciplinary than the others. It is housed in the engineering school and the projects described concern such engineering issues as transportation and sustainability in the urban environment. The Data Science focus of the summer programs appear to be more about data management than analysis and prediction. Their summer program was begun in 2014 and was partially funded by Oracle. (It is not clear from their website that the summer program is still in operation.)

# APPENDIX 8
# Tentative Outline for a New SMU PhD in Data Science

First Year, Fall Semester
- Introduction to Probability Theory and Inference
- Introduction to Data Collection and Management (SQL, Python, Hadoop)
- Elective (in chosen field of specialization)

First Year, Spring Semester
- Introduction to Data Mining (using R and Python)
- Likelihood and Generalized Linear Models (Using R)
- Elective

First Year Qualifying Exam

Second Year, Fall Semester
- Machine Learning (using R and Python)
- Geo-Spatial Analysis (ARCGIS)
- Elective

Second Year, Spring Semester
- Experimental and Causal Analysis
- Text Analytics, Sentiment and Network Analysis
- Elective

Second Year Qualifying Exam

Third Year, Fall Semester
- Visual Analytics and Virtualization Research
- 6-hours practicum/internship

Third Year, Spring Semester
- 9-hours Thesis Research and presentation of Third-Year Paper (PhD Prospectus)

Fourth Year, Fall and Spring Semesters
- 12-hours Dissertation Research (Defend Dissertation)

Potential Fields of Specialization

| | |
|---|---|
| Applied Economics | Cyber Security |
| Applied Mathematics | Financial and Business Analytics |
| Biostatistics/Bioinformatics | Marketing (Brierley Institute) |
| Computational Biology | Operations Research |
| Computational Chemistry | Smart Cities and Regional Economic Development |
| Creative Computation / Visualization | Human Rights |

Summer months to be utilized for internships, or research/learning activities applicable to the student's education.

# APPENDIX 9

Goal Two: