

# A Proposal for the Future of Data Science at SMU

A report prepared for the Office of the Provost

July 20, 2017

## Contents

	Page
Introduction .....	1
What is Data Science? .....	1
Why Data Science? .....	2
Data Science in Higher Education .....	3
Data Science at SMU, a SWOT Analysis .....	4
Recommendations .....	6
Appendix 1: Data Science Programs at Aspirational Peer Institutions .....	8
Appendix 2: Data Science Programs at Universities Visited by the Task Force .....	9
Appendix 3: SMU Degree Programs Involving Data Science .....	12
Appendix 4: Known Employers of SMU Graduates in Data Science and Analytics .....	13
Appendix 5: Centers and Institutes Representing Potential Resources for a Data Science Initiative .....	14
Appendix 6: Centers and Institutes Representing Potential Customers for a Data Science Initiative .....	15
Appendix 7: Data Science of Human Rights and Social Justice .....	18
Appendix 8: Tentative Outline for a New PhD in Data Science .....	20
Appendix 9: Contributions to Goals of the 2016-2025 Strategic Plan .....	21

# A Proposal for the Future of Data Science at SMU

## Introduction

In the fall semester, 2016, Provost Currall assembled a Task Force<sup>1</sup> 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 presents the findings of the Task Force and concludes with recommendations to ensure that SMU is well positioned and competitive in the increasingly important field of Data Science.

## 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 Task Force's definition of Data Science.

*Data Science is an interdisciplinary field consisting of methods and systems to extract knowledge and insights from data. It encompasses statistics, machine learning, visualization, business analytics, data analytics, and scientific computing.*

---

<sup>1</sup> Members: Alejandro Aceves, Marci Armstrong, Tom Fomby, Suku Nair, Jim Quick, Santanu Roy, Lynne Stokes, and Halit Uster.

## 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.<sup>2</sup> Current projections estimate that the rate of data production will reach 1.7 megabytes of new information every second for every human being on the planet by the year 2020.<sup>3</sup> Expressed in more tangible terms, the information generated in one second in 2020 would be the equivalent of more than 8 billion copies of the novel, Tom Sawyer, which if stacked would reach 2/3 of the distance to the moon. In today's parlance, "Big Data," has come to encapsulate the concept of this rapidly expanding informational resource, its potential, and the challenges it poses.

### Scale and Impact of Big Data

- By the year 2020, about 1.7 megabytes of new information will be created every second for every human being on the planet.
- For a typical Fortune 1000 company, just a 10% increase in data accessibility will result in more than \$65 million additional net income.
- Retailers who leverage the full power of big data could increase their operating margins by as much as 60%.
- 73% of organizations have already invested or plan to invest in big data by 2016

(<https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#4674f14017b1>)

The impact and potential of big data is now widely recognized. Almost every modern organization, be it a company, non-profit organization, health care organization, or government agency is inundated with data derived by the digital collection of information (credit card swipes, satellite photos, remote sensors, brain and CAT scans, security cameras, on-line transactions, etc.). As noted by Shmueli, et al., "If you consider the traditional statistical study (15 variables, 5000 records) to be the size of a period at the end of a sentence, then the Walmart database is the size of a football field."<sup>4</sup> The analysis of such huge datasets for insights require integration of the special tools that make up the field of Data Science.

Data Science is a new field offering important and long-lasting benefits to society. According to the McKinsey report Big Data: The Next Frontier for Innovation, Competition, and Productivity by James Manyika, et. al.<sup>5</sup>, "Data has swept into every industry and business function and now are an important factor of production. Big Data will become a key basis of competition and growth for individual firms and likewise will underpin new waves of productivity growth and consumer surplus." While storage

---

<sup>2</sup> <https://www.sciencedaily.com/releases/2013/05/130522085217.htm>.

<sup>3</sup> <https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#4674f14017b1>

<sup>4</sup> Shmueli, G., P.C. Bruce, N.R. Patel, Data Mining for Business Analytics, 3<sup>rd</sup> ed. (Wiley, 2016), p. 7.

<sup>5</sup> Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. H. Byers, Big data: The Next Frontier for Innovation, Competition, and Productivity (McKinsey Global Institute Report, May, 2011), p. vii and pp. 1 - 13.

capabilities and computing power have increased in tandem, there is a scarcity of ability to effectively distill and evaluate these data. Again, quoting the McKinsey report, “The United States alone faces a shortage of 140,000 to 190,000 people with analytical expertise and 1.5 million managers and analysts with the skills to understand and make decisions based on the analysis of Big Data.” The Business Higher Education Forum reports that there were 78,959 job postings asking for data science and analytics skills in the Dallas-Fort Worth-Arlington Metroplex in 2015!<sup>6</sup> Thus, the promise of Big Data challenges institutions of higher education to train an interdisciplinary workforce capable of managing and distilling insights from Big Data.

## Data Science in Higher Education

An increasing number of universities are rushing to respond to the “Big Data challenge.” Eleven of SMU’s twelve aspirational peer institutions have degree or certificate programs in **Data Science** and/or Data Analytics (Appendix 1), and a growing number of leading universities, including UC Berkeley, University of Michigan, NYU, MIT, University of Rochester, Columbia, and Harvard to name a few, have launched aggressive Data Science initiatives with significant seed funding to create Data Science institutes or centers.

Following a review of university websites, Rice University, Columbia University, the University of Rochester and New York University were approached for information given that they are respected private institutions, which have recently directed significant financial resources toward Data Science initiatives. Of the four, Rice University’s program is the least evolved, with no plans to create a Data Science institute or a center, no degree program existing or planned, and a relationship undefined. Columbia University, the University of Rochester and New York University have degree programs and have established a center or institute with a clear charge and governance. Benefits cited as a result of these initiatives are: increased income through gifts, increased tuition revenue from degree programs, increased research grants and corporate partnerships, improved faculty recruiting, and improved student quality.

Although the initiatives of these universities differ in detail (Appendix 2), the following practices were identified by the Task Force as important elements of success:

- **Fostering an interdisciplinary community** within the university organized around Data Science.
- **Governance and financial solutions that enable faculty participation and encourage school and department support.**
- **Staff support** consistent with the scale and mission of the initiative.
- **Building on existing institutional strengths and external relationships.**
- Focus on a **limited number of important themes.**

---

<sup>6</sup> Business Higher Education Forum, 2017, Investing in America’s data science and analytics talent, the case for action. (<https://www.pwc.com/us/en/publications/data-science-and-analytics-skills.html>)

- **Emphasis on degree programs** as foundational elements in the program.
- **Creation of external partnerships** with corporations and the local community.
- **Seed funding** to launch the initiative.
- **Support for essential tools, such as high-performance computing and visualization.**

## Data Science at SMU, a SWOT<sup>7</sup> Analysis

As a growing number of universities scramble to jump on the Big Data bandwagon, it is unlikely that all new programs will succeed, and it is in SMU's interest to assess its potential for success. The following SWOT analysis identifies a number of SMU's strengths, weaknesses, opportunities and threats that should be considered in moving forward with a Data Science initiative.

### **Strengths:**

- Considered collectively, **SMU degree programs** involving or contributing to Data Science (Appendix 3) constitute a major strength compared to many universities. Cox, Dedman and Lyle collectively offer 7 such masters programs, and the Office of the Provost oversees an interdisciplinary MS in Data Science. In addition, Cox offers an MBA with a concentration in business analytics and a Graduate Business Analytics Certificate, Dedman offers a PhD in Biostatistics, and Lyle offers a PhD in Operations Research. Lyle also offers an MS in Datacenter Systems Engineering, which may be viewed as a program supporting Data Science.
- **The Cox School of Business brand** increases the visibility of Data Science at SMU. The Cox MS in Business Analytics is currently ranked 25<sup>th</sup> in the nation, and graduates of Cox programs in analytics are highly sought by many corporations (Appendix 4).
- **Strong programs in Applied Mathematics, Statistical Science and Computer Science** provide the expertise required to develop new methods to manage and evaluate data, utilize machine learning and artificial intelligence, and insure cyber security.
- **Extensive Research** in fields increasingly influenced by Big Data is conducted by faculty widely across SMU in Cox, Lyle, Dedman, Simmons, and Meadows. The Task Force identified 6 SMU centers and institutes with expertise that could contribute directly to the success of a Data Science Initiative (Appendix 5), and 17 SMU centers and institutes representing potential customers for a Data Science Initiative because of their focus on issues in which big data plays an increasing role (Appendix 6). These centers will contribute research directions and benefit directly from an SMU Data Science initiative.
- **Collaboration with UT Southwestern** represents both a strength and an opportunity. The SMU-UTSW PhD in Biostatistics creates a foundation on which expanded collaboration in the fields of population health, computational biology, individualized medicine, human genome research, to mention a few possible collaborative research areas.
- **The Dallas Metroplex** is home to an increasing number of corporations for which Big Data plays an increasing role in their strategic decisions and everyday business operations. These

---

<sup>7</sup> SWOT = Strengths, weaknesses, opportunities, and threats.

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 is noteworthy that initiatives at Rice, Columbia, Rochester and NYU were each launched with financial commitments on the order of tens of millions of dollars (Appendix 2).
- **Disciplinary silos** have historically minimized interdisciplinary collaboration at SMU. Although this is changing and there are a growing number of interdisciplinary programs on campus every year, a successful Data Science initiative will require unprecedented collaboration by the majority of the University's schools.
- **Lack of coordination of Data Science resources** obfuscates the scale of the University's considerable footprint in Data Science.

#### ***Opportunities:***

- **Creation of a high-profile, nationally ranked top 10 program** is within SMU's reach given its current assets in Data Science and location in Dallas provided adequate investment is assembled by the University.
- **Increased collaboration with a medical schools and research centers** such as UTSW, or the UNT Medical School in the field of population health. Population health has emerged as a Data Science focus at a number of universities. SMU intellectual and computational resources partnered with medical expertise and data would constitute a nationally competitive program in population health.
- **Federal funding for Data Science** has increased in recent years. For example, NSF's Transdisciplinary Research in Principles of Data Science (TRIPODS) supports development of interdisciplinary research programs in Data Science. With proper coordination, SMU's strengths in Data Science and computing resources can position the University to be internationally competitive.
- **Increased community engagement.** The City of Dallas and numerous foundations in the Dallas area have expressed interest in obtaining assistance from SMU in dealing with Big Data, opening the door to development of community-focused Data Science projects.
- **The North Texas Economic Development District** was recently recognized by the Department of Commerce, opening the possibility of Federal funding for Data Science projects focused on community economic development.

- **Increased corporate engagement** would be stimulated by a University Data Science initiative. Enamored by the promise of Big Data, corporations are seeking immediate solutions in the short term and desire a workforce trained in Data Science both immediately and in the long term.
- **The Ford Research Building** with its planned visualization center and high-tech features is an optimal space to house an interdisciplinary Data Science initiative. The successes of Data Science at Rochester and NYU are due in no small part to university commitments to dedicate similarly equipped space to their Data Science initiatives.
- **Human Rights and Social Justice** are topics with many opportunities for Data Science, but there appear to be no universities to date with centers that focus on these topics (Appendix 7).
- **No Texas institution of higher education currently offers an undergraduate degree in Data Science.** An SMU undergraduate degree program in Data Science would distinguish the University from other Texas institutions.
- **Creation of a highly ranked interdisciplinary PhD program in Data Science** represents a real opportunity for SMU at this time. With its multiple degree programs in Data Science, SMU already has the courses in place to support an interdisciplinary Data Science PhD structured to provide students with a strong analytical foundation coupled with an opportunity to specialize in a number of fields (Appendix 8). It will be easier to create a highly ranked PhD program now, when there are few competing programs, than in a few years, when there are likely to be many.
- **Interdisciplinary collaboration** will be a significant outcome of a successful Data Science initiative. Big Data has emerged as an issue that has generated degree programs and impacts research across the disciplines. A Data Science initiative is a mechanism to bring together the Data Science expertise distributed across multiple schools and departments at SMU and create an interdisciplinary community of scholars that will generate exciting synergies, in turn, increasing the research productivity of the University.

**Threats:**

- **Inaction** poses the greatest threat to SMU. SMU's portfolio of Data Science educational programs and research skills coupled with its location in a major metropolitan area positions the University to succeed provided it does not delay. However, absent a Data Science program, SMU will become less competitive in this rapidly emerging field as more and more universities stand up serious initiative. At the same time, other University educational and research programs will also fall behind as their fields become more involved with Big Data. This evolution will adversely affect faculty and student recruitment and diminish the stature of the University.

## 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	<ul style="list-style-type: none"> <li>• Concentration in Business Analytics</li> </ul>
Brandeis University	<ul style="list-style-type: none"> <li>• New MS in Strategic Analytics offered by the Division of Graduate Professional Studies and focused on business analytics.</li> </ul>
Carnegie Mellon University	<ul style="list-style-type: none"> <li>• Numerous degree programs at the masters level grouped and introduced in single website</li> </ul>
Emory University	<ul style="list-style-type: none"> <li>• Nell Hodgson Woodruff School of Nursing's Center for Data Science focuses on utilizing data to improve health outcomes.</li> <li>• Emory continuing education offers 10-day sessions on Big Data and Data Analytics.</li> <li>• Computer Science offers an MS in Computer Science with a Data Science concentration.</li> <li>• Goizueta Business School offers MS in Business Analytics.</li> </ul>
Lehigh University	<ul style="list-style-type: none"> <li>• Data Science is one component of the Lehigh's Data "X" Initiative which aims to prepare students to use the latest technologies and methodologies.</li> <li>• College of Business and Economics offers a Business analytics Certificate.</li> </ul>
Tufts university	<ul style="list-style-type: none"> <li>• Still exploring options.</li> </ul>
Tulane University	<ul style="list-style-type: none"> <li>• School of Public Health and Tropical Medicine has a Department of Global Biostatistics and Data Science offering MS and PhD in Biostatistics.</li> </ul>
University of Notre Dame	<ul style="list-style-type: none"> <li>• Online interdisciplinary MS in Data Science in collaboration with AT&amp;T.</li> <li>• Mendoza College of Business offers an MS in Data Analytics.</li> </ul>
University of Rochester	<ul style="list-style-type: none"> <li>• Goergen Institute for Data Science</li> </ul>
University of Southern California	<ul style="list-style-type: none"> <li>• Department of Data Science and Operation in the Marshall School of Business (30 faculty).</li> <li>• MS in Computer Science with specialization in Data Science.</li> </ul>
Vanderbilt University	<ul style="list-style-type: none"> <li>• School of Medicine Big Biomedical Data Science (BIDS) Program offers PhDs.</li> </ul>
Wake Forest University	<ul style="list-style-type: none"> <li>• MS in Business Analytics</li> </ul>

## **APPENDIX 2**

### **Data Science Programs at Universities Visited by the Task Force**

#### **Rice University**

Rice University has undertaken a university-wide initiative to increase both the quantity and quality of data-driven knowledge-discovery at Rice. Their Data Science Initiative is funded by roughly \$45 million of the university's endowment as part of the university's recently announced \$150 million investment in research excellence. The Data initiative is intended to (1) bring new data scientists to Rice as tenure-track faculty members, (2) support collaboration-building activities for Rice's existing data-science community, and (3) develop educational offerings in Data Science at both the undergraduate and graduate levels.

Rice University has chosen to focus its Data Science initiative on urban analytics, medical analytics, and development of methods. The university is well placed to capitalize on the data-rich resources of the United States' fourth largest city, and is located across the street from the world's largest medical center. Rice has long ties to Houston's major industries, its school districts, and its civic organizations. Rice has a data-sharing arrangement with the City of Houston and works with city, county, and regional organizations on a variety of problems including urban flooding, air quality and health, education policy, and many more.

Rice University currently offers no degrees in data or business analytics, but is currently considering the creation of an undergraduate program with one or more minors in analytics, visualization, communication that would be open to majors in any field. The university has no plans to create an institute or center.

As part of the initiative, Rice expects to hire between six and ten new faculty members over the next three years with expertise at the interface of disciplines. Applications are being accepted at all ranks for employment as early as July 1, 2017, and are vetted by a university faculty committee formed in 2016. Participating departments must commit to support the Data Science curriculum for interdisciplinary minors.

#### **Columbia University**

Columbia University's initiative is based on a partnership with the City of New York. A new Institute for Data Science and Engineering was created in 2012 in response to the need for the acquisition and analysis of "Big Data" and with the overarching goal of job creation and economic impact. New York City obligated \$15 million in support of the initiative. Columbia commitments include \$80 million in private investments, 44,000 square feet of new space on the Columbia campus, and hiring 75 new faculty over the next 15 years. The goal of this partnership is job creation and economic impact.

The institute is administered by Sharon Sputz, Director of Operations and Academic Programs, who is supported by an Assistant Director of Operations, an Assistant Director of Admissions and Academic Affairs, and a Program Coordinator. The director currently reports to the Dean of Engineering but will likely be transferred under the Provost to emphasize and support the university-wide mission of the institute.

The Institute coordinates 7 Research Centers and 3 working groups, provides guidance for interdisciplinary research and degree programs, and has an extensive Industry Affiliates program. The institute has a curriculum committee, and currently offers an MS and a certificate in Data Science and is working to establish a PhD program.

Responsibilities of the Institute include negotiation of complex interdisciplinary grant proposals and creation of corporate partnerships through an industry affiliate program. To encourage collaboration, indirect cost recovery is shared with schools and agreements are negotiated to support faculty course buy outs.

## **University of Rochester**

The University of Rochester Data Science initiative was created as a redirection of an earlier initiative in systems science following the 2012 release of the McKinsey Global Institute Report, Big data: The Next Frontier for Innovation, Competition, and Productivity. In 2012, the University of Rochester committed \$100 million to construction of a new Data Science building, purchase of new high-performance computing equipment, and 20 new faculty positions dedicated to Data Science. In 2014, the Wegman foundation donated \$10 million in support of Data Science, and in 2015 Robert and Pamela Goergen donated an additional \$10 million to name the Goergen Institute for Data Science.

The Institute reports to the school deans. It seeks to capitalize on existing university strengths in health care, neuroscience, cognition and linguistics, and is focused on Health Science and information, artificial intelligence, and machine learning, with an emphasis on both computational and developmental research. Institute responsibilities include organizing workshops and seminars to stimulate faculty involvement and ideas, providing seed funding to facilitate external applications, and recruiting industry partners, which is an important goal of the university, and plans exist to create an executive-level position within the Institute to manage corporate relations.

Degree programs are considered foundational components of the initiative. The College of Arts and Sciences currently offers a BS/BA in Data Science and a 1-year, concentrated MS with an option for a second year that includes an internship. The Business School MS in business analytics was launched with 4 students in 2014 and expects to enroll 60 students this year. The university is actively planning to add a PhD program in Data Science.

## **New York University (NYU)**

NYU launched its Initiative in Data Science and Statistics with the creation of its Center for Data Science in 2013. Within the same year, the Center for Data Science was awarded grants by the Moore Foundation and Sloan Foundations totaling \$12.6 million as part of a collaboration with the University of Washington and the University of California, Berkeley. The university has provided space for the Center, but it is unclear if NYU it has contributed significant funding.

The Director of the Center for Data Science reports to the Provost, an administrative arrangement strongly recommended by the Acting Director based on the experience of the university in operationalizing the center because it reduces the competition for resources associated with coordinating interdisciplinary programs across school boundaries. A memo of understanding between

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 Data Science is charged under the university's Data Science initiative with collaborating with the Center for Urban Science and Progress (CUSP), and the Center for Promotion of Research Involving Innovative Statistical Methodology (PRIISM). A total of 17 faculty are in the Center for Data Science; some have joint appointments and some are seconded to the center. All center faculty receive 1 month summer salary.

The Center for Data Science has produced significant benefits for NYU. An MS in Data Science offered by the center has become the flagship master's program of NYU with a national ranking of #2, and the "cash cow" of the center. The PhD program recently launched by the center is expected to earn a similarly prestigious ranking. Grant activity is increasing as a result of center activities, and the center has become a faculty "recruiting magnet."

### APPENDIX 3 SMU Degree Programs Involving Data Science

School	Department	Plan	Enrolled AY 2017	Graduated	
				AY 2016	AY 2017*
Research & Grad	Research & Graduate Studies	Data Science - MSDS	385	13	48
Cox	Info Tech and Operations Management	Business Analytics-MBA	93	36	56
Cox	Info Tech and Operations Management	Business Analytics-MN	24	3	3
Cox	Info Tech and Operations Management	Business Analytics-MSBA	65	63	63
Cox		Graduate Business Analytics Certificate		63	57
Dedman	Economics	App Eco & Pred Analytics-MSAEPA	17	10	10
Dedman	Statistical Sciences	Applied Stats & Data Analytics-MS	45	24	24
Dedman	Statistical Sciences	Biostatistics-Ph.D.	9	0	0
Lyle	Computer Science & Engineering	Computer Engineering - MSCPE	8	9	6
Lyle	Computer Science & Engineering	Computer Science - MS	75	25	35
Lyle	Engineering Management, Information, & Systems	Operations Research - MS	27	18	7
Lyle	Engineering Management, Information, & Systems	Operations Research - PhD	14	1	0
Lyle	Lyle School of Engineering	Datacenter Systems Engineering -- MS	21	2	8

## 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; CheapCarribbean.com; Chickasaw Nation Industries; 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 Resources; TriscendNP; TXU; UNT; VRtices Corp; Walmart; WellCare; Toyota; Ericsson; inCode Consulting; American Airlines; PwC Consulting; Sabre; Amazon; AT&T; Essilor; Intel (McAfee); Alcon, a Novartis company; Axtria; Blue Shield of California; Deloitte Consulting; Enforce Consulting; Equifax; FedEx; General Mills; Kimberly Clark; KPMG; McMaster-Carr; Methanex; RBC; Rent A Center; Staples; Texas Rangers; Whirlpool; Riverbend Growth Partners.
Dedman College	Economics	Federal Reserve Bank Dallas; Prudential, Citibank; King Saud University; Comerica Bank; Blue Cross Blue Shield of Texas; Mary Kay, Inc.; State Farm Insurance; Capital One; Citibank; SABRE; KPMG; FTI Consulting; Moody's; Accenture; Fannie Mae; Bureau of Labor Statistics; Cottonwood Financial; Americredit; Compass Lexecon; Analysis Group; Disney; Equifax; FEDEX; Annadale Capital; Google; TXU.
Dedman College	Statistical Sciences	Abbott Laboratories; Cottonwood Financial in Dallas; Assurex Health; Baylor Scott & White Health; Fannie Mae; Dell; John Deere; Experis; Proctor and Gamble; Littler Mendelson; General Motors Finance; AgilityDocs; Baylor Scott & White; Blue Cross Blue Shield of Texas; Abbott Laboratories; Cottonwood Financial in Dallas; Assurex Health; Mary Kay; Bank of America; Capital One; GM.
Lyle School of Engineering	Engineering Management and Information Science	Accenture, Adani Ports Special Economic Zone (APSEZ), Advice Interactive Group, Alcatel, American Airlines, AT&T, Bain & Company, Bank of America, Centendera (sp), Chevron, Cindero Consulting, Cisco, Citi Technology, Citigroup, Defense Information Systems Agency, Deloitte, Disney, Epsilon, Ernst & Young, Ethos Group, First Southwest Company, Frito Lay, GM, Hewlett Packard, Hitachi Consulting, INDESA, IBM, ING Capital, Kone, L3 Communications, Lockheed Martin, Merrill Lynch, PricewaterhouseCoopers, Quorum, Raytheon, Royal Bank of Canada, Sabre, Southwest Airline, Texas Instruments, Towers Perrin, Toyota, US Army, Verizon.
Lyle School of Engineering	Computer Science and Engineering	7-Eleven, Accenture, Alcatel, Amazon, American Airlines, Apple, AT&T, Bank of America, Cisco, Citigroup, Deloitte, Ernst & Young, Ericsson, Frito Lay, Fujitsu, GM, Google, Hewlett Packard, Hitachi Consulting, Huawei Technologies, IBM, L3 Communications, Lockheed Martin, Merrill Lynch, Microsoft, National Security Agency (NSA), PricewaterhouseCoopers, Quorum, Raytheon, Sabre, Southwest Airline, SPAWAR, State Farm, Texas Instruments, Toyota, US Airforce, US Army, Verizon.

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

<b>Center/Institute Name</b>	<b>Center/Institute Administration</b>	<b>Center/Institute Description</b>	<b>Potential Relationship to an SMU DSI</b>
Center for Scientific Computation	Central	An interdisciplinary research center devoted to the application of computational techniques to problems in mathematics, engineering, and applied sciences.	Direct Support for DSI and Research Collaboration
The AT&T Center for Virtualization	Central	This recently created center addresses challenges of security, performance, quality of service, interoperability and scalability as the number of devices taking advantage of the Cloud increases.	Research Collaboration
Center for Creative Computation	Meadows School of the Arts	An interdisciplinary research and teaching center exploring computation as a universal generative medium—integrating creative development, quantitative analysis and interdisciplinary synthesis. The Center offers both a major and minor in Creative Computing, combining core coursework from the Meadows School of the Arts and the Lyle School of Engineering.	Direct Support for DSI and Research Collaboration
Darwin Deason Institute for Cyber Security	Lyle School of Engineering	The mission of the Darwin Deason Institute for Cyber Security is to advance the science, policy, application and education of cyber security through basic and problem-driven, interdisciplinary research.	Research Collaboration
NSA Center of Academic Excellence	Lyle School of Engineering	Cybersecurity research and education has long been a priority at SMU's Lyle School of Engineering. Researchers in the High Assurance Computing and Networking (HACNet) Lab work on a broad range of cybersecurity topics.	Research Collaboration
Dedman College Interdisciplinary Institute	Dedman College	The Dedman College Interdisciplinary Institute (DCII) encompasses all the divisions within Dedman College: humanities, social sciences and sciences, which includes mathematics and statistics. It also reaches out through its programs to the rest of the University and to the local community, creating a culture of interdisciplinary inquiry and bridging the physical sciences and the humanities, the liberal arts and the professional schools, and academia and the broader community.	Research Collaboration

**APPENDIX 6**  
**Centers and Institutes Representing Potential Customers**  
**for a Data Science Initiative**

<b>Center/Institute Name</b>	<b>Center/Institute Administration</b>	<b>Center/Institute Description</b>	<b>Potential Relationship to an SMU DSI</b>
Center for Global Health Impact	Central	The center aims to save lives in developing countries and in poor settings through innovation approaches that bring effective and affordable health solutions to those who need them and in ways they will use them.	Research Collaboration
The Guildhall at SMU	Central	The premier graduate video game education program in the nation. Many of the school's founders are industry icons, and classes are run by industry veterans. Alumni have worked at more than 80 video game studios in North America. The program offers a Master's degree in Interactive Technology and a graduate Professional Certificate.	Research Collaboration
National Center for Arts Research	Meadows School of the Arts	The center conducts data-based research on the economic health of the arts and culture industry. The NCAR Fundraising Report is the only report to examine fundraising across genre, organization size and geographic location, based on data provided by over 2,700 organizations across 11 arts disciplines, showing trends over the four-year period 2011-2014.	Research Collaboration
Temerlin Advertising Institute	Arts	The institute strives to advance the state of advertising communication through partnerships with both industry and government and through programs to blend the research interests of the academy and the profession. While enhancing the effectiveness of advertising domestically and abroad, Institute professors with high levels of advertising expertise and experience provide undergraduate instruction to a select group of students, within an environment that fosters creativity, energy and ethical conduct.	Research Collaboration Education
Center for Research in Real Estate and Land Use Economics	Cox School of Business	The Center provides opportunities for students to deepen their exposure to critical issues in the current real estate industry and acts as a conduit for research which it conducts through the Costa Institute of Real	Research Collaboration Education

		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  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  Education
Kitt Investing and Trading Center	Cox School of Business	The Kitt Investing and Trading Center is a state-of-the-art instructional and research facility designed to integrate finance curriculum, enhance innovative faculty research, and teach students practical finance and investment applications.	Research Collaboration  Education
JCPenney Center for Retail Excellence	Cox School of Business	The Center is a leading source of academic expertise on consumer shopping behavior and the effects of retailer activities on shopping behavior.	Research Collaboration
Maguire Energy Institute	Cox School of Business	The Institute promotes the study of policy, marketing, and management issues related to oil, natural gas and electricity.	Research Collaboration
Brierly Institute for Customer Engagement	Cox School of Business	The institute trains students and conducts research to promote business practices that build customer relationships, leverage digital and traditional media, measure financial impact, and create customer experiences which engage customers and create loyalty and value.	Research Collaboration  Education
Hunter and Stephanie Hunt Institute for Engineering and Humanity	Lyle School of Engineering	The Hunter and Stephanie Hunt Institute for Engineering and Humanity is dedicated to using the power of engineering, collaboration and the free market to develop and implement solutions to the problems of the poor, both here and abroad.	Research Collaboration
NSA Center of Academic Excellence	Lyle School of Engineering	Cybersecurity research and education has long been a priority at SMU's Lyle School of Engineering. Researchers in the High Assurance Computing and Networking (HACNet) Lab work on a broad range of cybersecurity topics.	Research Collaboration

Tsai Center for Law, Science and Innovation	Dedman School of Law	The Tsai Center for Law, Science and Innovation is a research-focused academic center exploring how law and policy affect scientific research and discovery as well as the development and commercialization of new technologies.	Research Collaboration
The Center for Drug Discovery, Design, and Delivery (CD4)	Dedman College	The Center for Drug Discovery, Design and Delivery (CD4) is a novel multi-disciplinary focus for scientific research targeting medically important problems in human health. Using innovative approaches, CD4's mission is to potentiate the development of new therapeutics, their delivery methods as well as the translation of these new therapeutics to clinical studies.	Research Collaboration
Institute for Evidence-Based Education	Simmons School of Education and Human Development	The Institute promotes academic excellence in K-12 education by conducting and disseminating cutting edge research on effective instruction, teachers, and schools.	Research Collaboration
Center on Research and Evaluation (CORE)	Simmons School of Education and Human Development	CORE conducts research and evaluations on programs, practices, and policies designed to strengthen families and communities and enable evidence-based decisions.	Research Collaboration

## **APPENDIX 7**

### **Data Science of Human Rights and Social Justice**

One potential focus for a Data Science initiative at SMU is human rights and related social justice issues. This is an area that has many opportunities for the application of the tools of Data Science. Two non-profits that are active in this area are the Human Rights Data Analysis Group ([hrdag.org](http://hrdag.org)) and DataKind ([datakind.org](http://datakind.org)). To date, there appear to be no universities who have centers that focus on human rights issues using analytics. However, there are currently three U.S. university Data Science centers that have public policy and justice issues as a component of their work. We provide examples of the types of questions and research that the nonprofit and university centers produce, as well as the role these types of projects play in the institutions where they occur.

Both HRDAG and DataKind were founded specifically to address human rights problems using data. DataKind is headquartered in New York and has a staff of about 20, about a third of whom have the title of data scientist and are trained in computer science and statistics. Several others have social science backgrounds. HRDAG is much smaller, with a staff of five, 3 of whom are trained in statistics/computer science. Both groups have advisory and technical boards with university, corporate, government and/or foundation representation.

A visible contemporary example of the kind of questions these organizations address is recent work by HRDAG to produce estimates of the number of deaths in Syria due to the current conflict there. The data scientists use lists of the dead produced by NGO's and other organizations. These lists must be first be de-duplicated, using various text comparison algorithms. Then models estimating the probabilities of inclusion on each list are used for predicting the number of deaths missing from all lists. An example of a recent DataKind project was carried out for Microcred, an organization that makes small loans to entrepreneurs in developing countries. Their loan application and status data was used to build predictive models to help select promising applicants to receive their loans.

The three university Data Science centers who explicitly state their intention to address questions of social justice and related public policy are at University of Chicago, University of Washington, and Georgia Tech. All three have (or have had) summer internship programs for students designed to train students in Data Science methods as they work on problems related to public policy. All three internship programs are affiliated with a Data Science center or institute somewhere in the university. The University of Chicago program was established first and the other two claim that their programs were modeled on it. The projects tackled by these centers have some overlap with those of the nonprofits; for example, HRDAG and University of Chicago's center have carried out studies about policing and its impact on minority communities.

The University of Chicago Data Science center, called the Center for Data Science and Public Policy, is affiliated with its public policy school (Harris School). Its stated goal is to use Data Science to improve public policy. Their internship program, begun in 2013, is called the Eric and Wendy Schmidt Data Science for Social Good Summer Fellowship program (DSSG). It brings students to campus to train in Data Science as they attack problems "that really matter... in education, health, energy, public safety, transportation, economic development, [and] international development," according to their website. The Fellowship program also has a staff housed in the Center, and directed by Rayid Ghanim who was the former Chief Data Scientist for the 2012 Obama Presidential campaign.

The University of Washington program is affiliated with its eScience Institute, which is a Microsoft-funded Data Science center that is a collaborative effort between University of British Columbia and

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/](http://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.)

SMU seems especially well positioned to have both human rights and social justice/public policy issues as a focus of its Data Science initiative. A high percentage of our undergraduates major in a social science and SMU is one of only a handful of universities that has a human rights degree program. Both the Embrey Human Rights Center and the Hunt Institute for Engineering and Humanity have these areas central to their mission, and have affiliated faculty who are knowledgeable about the areas. We are situated in the middle of Dallas, which has many problems related to poverty, human trafficking, and other similar issues, all of which can be examined using Data Science methods.

## 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 (ARC-GIS)
- 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**  
**Contributions to Goals of the 2016-2025 Strategic Plan**

<b>Task Force Recommendations</b>	<b>Strategic Plan Goals</b>	<b>Contribution</b>
<p><b>1.</b> Create a Data Science Institute (DSI).</p>	<p><b>Goal One:</b> Enhance the Academic Quality and Stature of the university.</p>	<p><b>Objectives One and Two:</b> Based on observed impact at other universities, an SMU DSI will enhance the University’s ability to recruit outstanding faculty and students.</p> <p><b>Objective Five:</b> An SMU DSI will support interdisciplinary research and education utilizing high-performance computing to address issues confronting cities of the future and businesses dealing with big data.</p>
	<p><b>Goal Three:</b> To Strengthen Scholarly Research, Creative Achievement, and Opportunities for Innovation.</p>	<p><b>Objective One:</b> The mission of an SMU DSI will be to engage faculty across the campus in interdisciplinary research (see Appendices 5 &amp;6).</p> <p><b>Objective Two:</b> The DSI should be charged with identifying funding sources and potential external partnerships and with assembling faculty and student teams to engage these opportunities.</p> <p><b>Objective Three:</b> An SMU DSI will facilitate wider use of SMU’s powerful ManeFrame II by the University community.</p>
<p><b>2.</b> Create an interdisciplinary undergraduate program in Data Science.</p>	<p><b>Goal One:</b> Enhance the Academic Quality and Stature of the university.</p>	<p><b>Objective Two:</b> Results at other universities demonstrate that Data Science degree programs are in high demand and attract high quality students. At this time, no university in Texas offers an undergraduate major in Data Science, and this innovation at SMU will distinguish it from other Texas universities and help recruit outstanding students.</p>

	<p><b>Goal Two:</b> To Improve Teaching and Learning.</p>	<p><b>Objectives One:</b> As the first in Texas, an undergraduate degree program in Data Science would be an innovation that equips SMU students with methodologies that enhance creative thinking and innovative problem solving informed by research and high-performance computing.</p> <p><b>Objective Two:</b> The undergraduate program would serve as a leading example of an interdisciplinary program that draws on the expertise of multiple disciplines and high-performance computing to equip students with tools to address the needs of the modern city, state, nation and world.</p>
<p>3. Stand up an interdisciplinary PhD in Data Science.</p>	<p><b>Goal One:</b> Enhance the Academic Quality and Stature of the university.</p>	<p><b>Objectives One and Two:</b> Based on observation of other universities, an SMU interdisciplinary PhD in Data Science will be a “top-10 program,” enhancing the University’s ability to recruit outstanding faculty and students.</p> <p><b>Objective Five:</b> An SMU interdisciplinary PhD in Data Science will engage graduate students in interdisciplinary research utilizing high-performance computing to address issues including but not limited to those confronting cities of the future and businesses dealing with big data.</p>
<p>4. Hire a minimum of 10 tenured/tenure track faculty.</p>	<p><b>Goal One:</b> Enhance the Academic Quality and Stature of the university.</p>	<p><b>Objective Two:</b> A strong, interdisciplinary faculty will support nationally recognized degree programs that will be recruiting tools to attract and graduate academically gifted undergraduate and graduate students from diverse backgrounds.</p> <p><b>Objective Five:</b> Development of an interdisciplinary Data Science faculty will be a strategic investment in instruction and research that utilizes high-performance computing and addresses issues including but not limited to those confronting cities of the future and businesses dealing with big data.</p>

	<b>Goal Two:</b> To Improve Teaching and Learning.	<b>Objectives One and Two:</b> Development of an interdisciplinary Data Science faculty will support and expand student access to an innovative curriculum that will prepare them to address the needs of the modern city, state, nation, and world.
5. Endowment of the DSI and the PhD program in Data Science.	<b>Goal One:</b> Enhance the Academic Quality and Stature of the university.	<b>Objective Five:</b> Endowment of an SMU DSI will constitute a strategic investment in instruction and research in but not limited to specifically targeted areas of interdisciplinary research, high performance computing, cities of the future, and entrepreneurship.
6. Add staff to facilitate access to ManeFrame II.	<b>Goal Three:</b> To Strengthen Scholarly Research, Creative Achievement, and Opportunities for Innovation.	<b>Objective Three:</b> Additional staff will enable greater access to ManeFrame II by the University community, ensuring return on the equipment investment and enhancing faculty recruitment.
7. Advertise SMU's current programs and strengths in Data Science	<b>Goal One:</b> Enhance the Academic Quality and Stature of the university.	<b>Objectives One and Two:</b> Communicating SMU strengths in Data Science will enhance faculty and student recruitment in the growing number of fields concerned with big data.
	<b>Goal Six:</b> To Increase Revenue Generation and Promote Responsible Stewardship of Resources.	<b>Objective Four:</b> Communicating SMU's current and developing strengths in Data Science will elevate the University's national profile.