Report by the Taskforce to Envision Data Science at Boston University

Draft of April 1, 2019

Taskforce Committee Members

- **Azer Bestavros (co-chair)**, Director, Hariri Institute, Professor, Computer Science
- **Daniel Kleinman (co-chair)**, Associate Provost for Graduate Affairs, Professor, Sociology
- **Margrit Betke**, Professor, Computer Science
- **Josée Dupuis**, Professor and Chair, Biostatistics
- **Jacob Groshek**, Associate Professor, Emerging Media Studies
- **Evan Johnson**, Associate Professor, Medicine and Biostatistics
- **Tasso Kaper**, Professor and Chair, Mathematics and Statistics
- **Clem Karl**, Professor and Chair, Electrical and Computer Engineering
- **Stan Sclaroff**, Dean ad interim, College of Arts and Sciences, Professor, Computer Science
- **Marshall Van Alstyne**, Professor and Chair, Information Systems
- **Tanya Zlateva**, Dean, Metropolitan College, Professor of the Practice, Computer Science

Staff to the Committee

- **Courtney Mansfield**, Administrative Coordinator, Undergraduate & Graduate Affairs

Executive Summary

Data science is widely seen as a promising field, one which is transforming business, government, and many areas of social life and is also having, and likely to continue to have, a substantial impact on the way in which scholars from fields as wide-ranging as English and astronomy think about research problems and seek answers. Given the growing importance of data science, President Robert Brown assembled a taskforce of faculty to envision the future of data science at Boston University (BU). Imagining a set of proposals that would promote collaboration and innovation across departments, programs, schools, colleges and campuses, and invite engagement by faculty, students, and staff, the taskforce met over two semesters, sought input widely from across BU and did research on initiatives proposed or underway across US institutions of higher education. As a result of its work, the committee developed a set of recommendations:

- **Vision & Mission:** BU should establish a mission statement that highlights the need for the institutional home for data science at BU to be deeply inclusive, promote collaboration and cooperation across the University and have porous boundaries, allowing easy connection.
- **Data Science Unit:** To facilitate integration of data science research and education, to serve as a hub for data science, and to coordinate data science efforts at BU, the committee recommends the creation of an autonomous but collaborative Data Science Unit. The report outlines specific organizational and governance structures for the Unit. Importantly, the elements of its governance should involve voices from across BU's schools and colleges.
- **Education:** While the new Data Science Unit can be a partner and collaborator across related educational programs, the committee concluded that much data science education can and should be left to established departments, programs, schools and colleges at BU. At the same time, BU will benefit if certain data science educational initiatives are led by the Unit. In particular, to be a leader in data science education, the committee believes the University needs to be able to offer undergraduate and graduate programs that focus on data science core methodological competencies. These programs should be overseen by an arm of the Data Science Unit that the committee calls the Division of Data Science.
- **Faculty:** As the academic hub for data science at BU, the Division of Data Science will need to have Faculty Members who also hold budgeted appointments, and will need to put in place a set of search and hiring practices for the development of diverse and inclusive set of data science faculty. The committee expects most of these will be joint appointments, involving schools and colleges across BU, but some faculty will likely be 100% appointed in the Division. To encourage broader participation, the Division should offer affiliated appointments to faculty who are engaged in research or education consistent with its mission.

- **Research:** To leverage the symbiotic relationship between research and education programs at BU, the Data Science Unit should also support university-wide interdisciplinary research centers and labs that pursue research or provide research support relevant to data science. These centers and labs, which involve affiliated faculty and students from across the University would operate under the Data Science Unit as a federation that share common resources and services, each subject to a specific mission.
- **Incentives:** To be successful, the new Data Science Unit will need to have its own well-provisioned budget. However, to ensure that data science pervades BU and is undertaken in a spirit of collaboration and interconnection, the committee believes that the University will need a carefully crafted system of incentives.
- **Outreach:** To enhance data science research and education at BU, the new Data Science Unit should make building collaborations with companies, government agencies, and non-profit organizations a high priority.

Figure 1 depicts the extent to which the organization of data science at BU under a new Data Science Unit would catalyze interconnectedness between the proposed Division of Data Science, cognate academic units, interdisciplinary research centers, industrial collaboratories, and civic society partnerships.

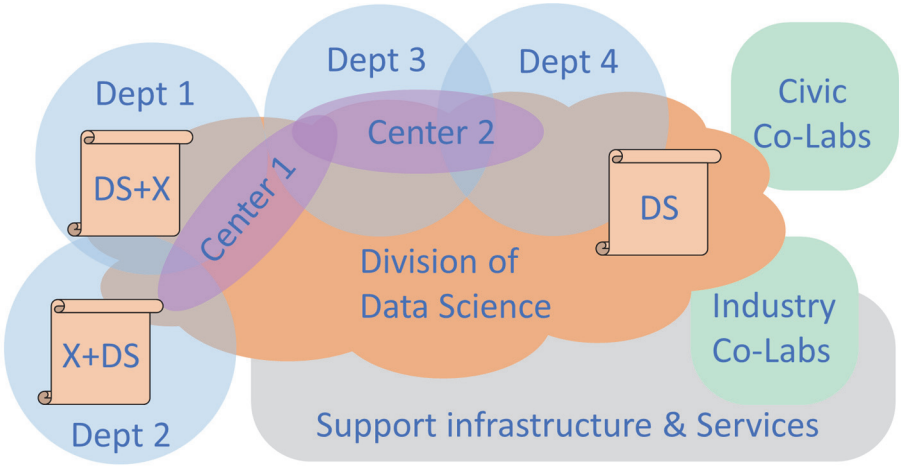

Figure 1: Interconnectedness as an organizing principle for a new Data Science Unit in support of interdisciplinary research and education at Boston University.

The committee believes that its recommendations constitute the blueprint for organizing data science at BU in a way that will enable the University to remain at the cutting edge of research, to establish itself at the forefront of educational offerings, to collaborate with area corporations, government agencies, and civic organizations, and to be highly visible, regionally, nationally and internationally in this fast-evolving field.

1. Introduction and Background

The digital age has brought us previously unimagined quantities and varieties of data in structured and unstructured forms. In this context, a new field is developing: ***data science***.

Operating at the nexus of mathematics, statistics, and computer science and engineering, data science offers significant promise as a means to better understand the social and natural world, to inform decisions in areas from business to government, and as a way to prompt collaboration across disparate academic fields. As a result, the tools and platforms of computing in general, and of data science in particular, have emerged as the *lingua franca* not only for research across the landscape of disciplines in academe, but also for bridging systems and solutions that span multiple sectors of the economy.

In the world of business, companies are using machine learning to guide strategic decisions and massive online experimental frameworks for product development.[1] In the US, some predict that data driven technologies in healthcare alone will result in as much as $300 billion in added value per year,[2] and we may need as many as 1.5 million data-science skilled managers to oversee the growing place of data analytics in corporate management and planning.[3]

In academia, data science can provide tools for scholarly research in areas as different as ecology and literary studies, and the field itself is evolving, as researchers consider new approaches to collect, ingest, manage, mine, and analyze data, as well as synthesize models that uncover the underlying processes and behaviors exhibited through such data. Our new data age is also raising important social research questions about privacy, security, and the possible impacts that the pervasive application of data science, and the emergence of artificial intelligence could have on work and social inequality.[4]

Tools from data science allow communication scholars to extract and analyze information from the Twittersphere, and empower literary scholars to glean patterns in large arrays of novels. Modern telescopes produce terabytes of data per observation, which astronomers analyze to derive complex models that shed light on their hypotheses and findings. Biologists mine microbial genomes to predict gene clusters that might produce useful compounds, and sociologists are using data science methods to reveal racial disparities or inequitable patterns in incarceration.

---

1 https://hbr.org/2018/08/what-data-scientists-really-do-according-to-35-data-scientists

2 https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

3 https://www.nsf.gov/cise/ac-data-science-report/CISEACDataScienceReport1.19.17.pdf

4 https://www.mercurynews.com/2019/03/15/stanford-unveils-new-ai-institute-built-to-create-a-better-future-for-all-humanity/

Across both the Charles River Campus and the Medical Campus hundreds of faculty members consider data science to be an element of their scholarly identities and many more are likely to do so in the future. Students from every major are flocking to take courses that prepare them for a competitive workforce, which is increasingly putting a premium on fluency with the concepts, frameworks, and tools in the broad areas of data science.

As exciting as it is to our faculty and students, much of BU's educational and research initiatives in data science are not well coordinated. This state of affairs is not unique to BU, and in fact has been evident at many peer institutions. This challenge is due, in part, to the pervasive nature of data science and the large number of scholarly fields embracing its transformative potential.

Yet, all around us leading universities have established or are planning integrated data science initiatives that go beyond cosmetic branding and rebranding of existing units and the frenzied introduction of ad-hoc professional degree programs to meet skyrocketing job-market demand. Among other noteworthy efforts nationwide, in May of 2017, the University of California—Berkeley announced the establishment of a Division of Data Science, and in October of 2018, MIT revealed plans to create a college of computing, described as *"an interdisciplinary hub for work in computer science, AI, data science, and related fields."*

In the context of this rapidly changing ecology, President Robert Brown assembled a faculty committee to serve as a taskforce to envision the future of data science at Boston University. To succeed in the world of academic data science, President Brown's charge to the committee underscored the need to breach disciplinary and organizational silos and integrate research and education, while promoting appropriate differentiation, but avoiding needless duplication and competition across departments, schools, and colleges.

With reference to his challenge to the university community[5] in the Spring of 2018 to make BU *"the most integrated university in the United States,"* a university that "seamlessly connects programs and schools," the committee envisioned an integrated, but porous initiative, one that draws in faculty, staff, and students from across the whole of BU and invites new forms of collaboration and engagement—an initiative that embodies President Brown's depiction (reproduced in Figure 2) of the interconnectedness of academic units and interdisciplinary programs and centers in a truly integrated university.

To understand what kinds of organizational structures would allow BU to lead in the world of data science and the kinds of educational programs it should host, the committee—a diverse group of faculty and academic leaders from across BU—met regularly for nearly six months.

[5] https://www.bu.edu/president/envisioning-our-future-boston-university-in-2030/

Throughout its deliberations, the committee's guiding principle has been to ensure that the integrative nature of data science is deeply infused into BU's forward looking 21st century culture—a culture that contributes to weakening existing disciplinary silos and promoting collaboration between units, faculty, and students.

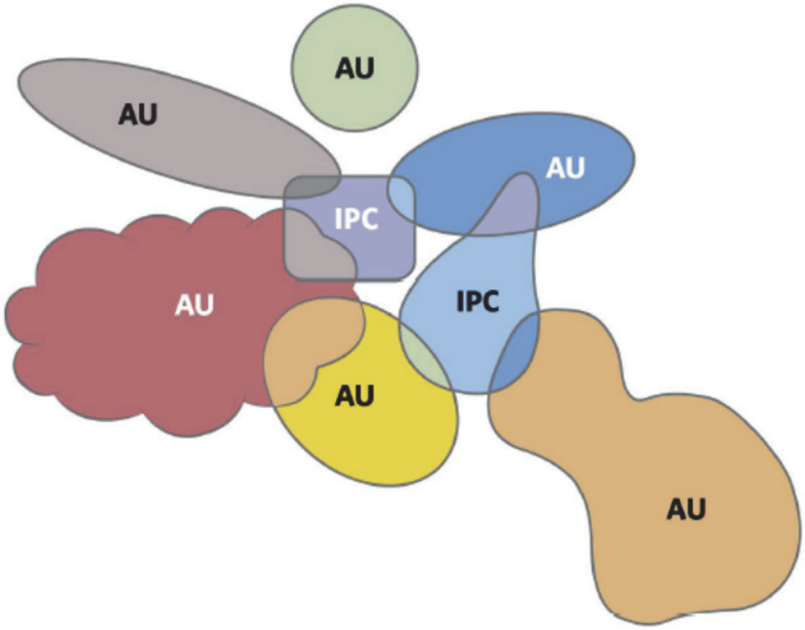

Figure 2: Overlapping academic units (AUs) and interdisciplinary programs and centers (IPCs) are effective catalysts that promote integrative research and education [Reproduced from President Brown's letter to the BU community, Spring, 2018].

2. The Data Science Landscape

Well before "data science" became part of the vernacular, the importance of computational and mathematical approaches to the extraction of knowledge from data (whether large or small) prompted the academic community to integrate these approaches into a growing number of disciplines through the development of new degree programs, research centers, initiatives, and departments in the United States. As depicted in Figure 3, by the early 1990s, "Data Mining" and "Computational Statistics" were well-established subfields of computer science and statistics.

Origin and Evolution of Data Science

In the late 1980s, computational statistics emerged as a distinct subfield of statistics, focusing on the use of computer-intensive methods to deal with very large sample sizes and heterogeneous

data sets.[6] Contemporarily, data mining emerged as a distinct subfield of the broader field of "Knowledge Discovery in Databases" (KDD),[7] which in the late 1980s emerged out of the then-decades-old "Data Bases Management Systems" (DBMS) field of Computer Science.

The introduction of scalable cloud computational platforms in the early 2000s, coupled with the data deluge that resulted from the digital transformation of our economy and society, provided the perfect setup for rapid advances in the use of machine learning in "big data" settings, which in turn fueled significant interest in data science as a profession.[8]

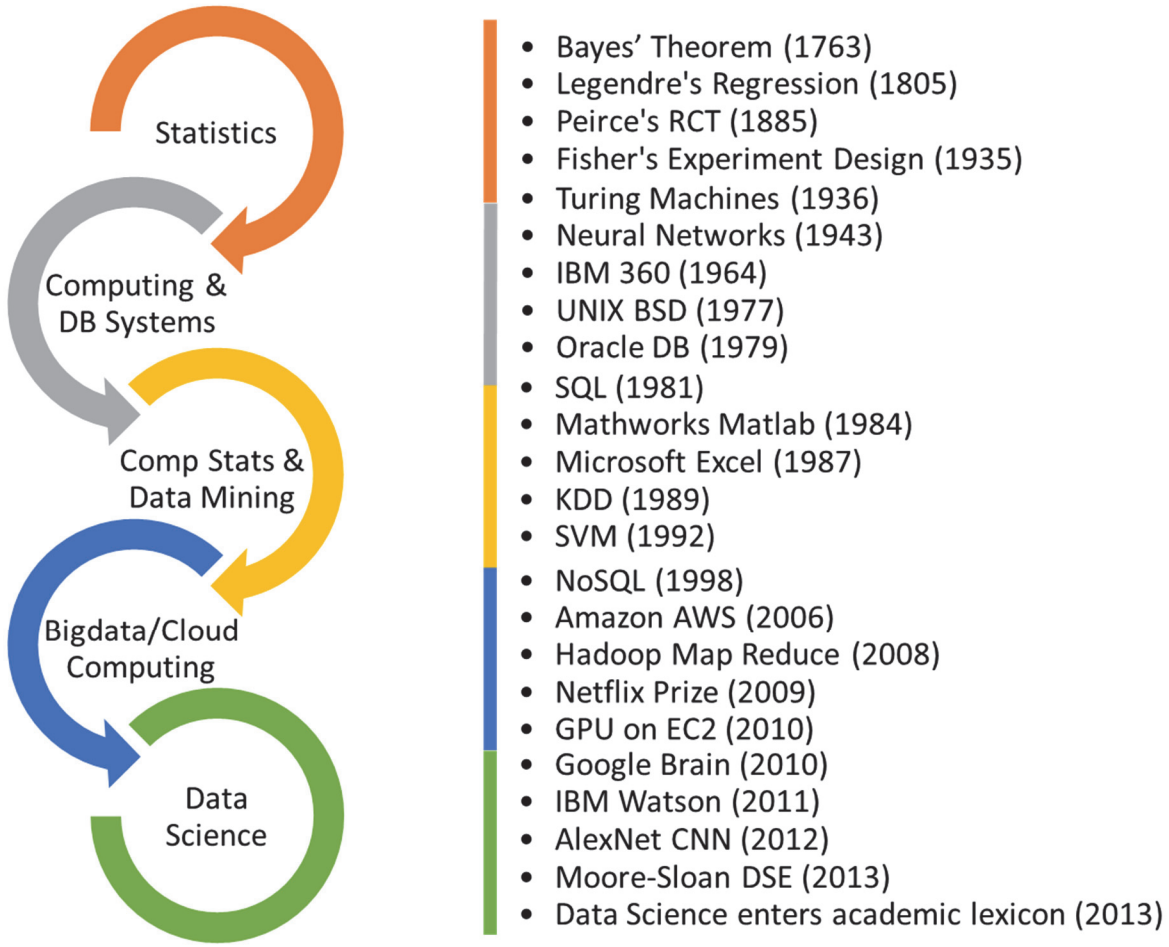

Figure 3: Origin and Evolution of Data Science prior to its embrace by academia in 2013.[9]

[6] Wegman, Edward J. "Computational Statistics: A New Agenda for Statistical Theory and Practice." Journal of the Washington Academy of Sciences, vol. 78, no. 4, 1988, pp. 310–322. JSTOR

[7] https://www.kdd.org/

[8] https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

[9] For a data-mining-centric timeline, see https://dataconomy.com/2016/06/history-data-mining/.

Data Science in Academia

The term data science[10] entered the mainstream academic lexicon around 2013 with the launch of the Moore-Sloan Data Science Environments (MSDSE) initiative.[11] By 2017, a study by the Moore Foundation revealed that more than 80% of 116 research universities considered in their report included at least one data science offering.[12]

In the Fall of 2018, a report about the top twenty data science initiatives in the US was released by ABT Associates to the institutions that were selected for the survey, including BU. The report was commissioned by the Moore and Sloan Foundations to assess the impact of their 2013 MSDSE investments (primarily in data science initiatives at the U of Washington, UC Berkeley, and NYU). It revealed some common characteristics of these leading programs. Most notably, all but one (at the University of Massachusetts, Amherst) were set up as independent or intercollegiate units (centers or institutes) that are not part of any existing college; fourteen of these units had dedicated faculty lines; and eleven offered degree programs.

Since the issuance of the report from ABT Associates, the introduction of data-science degree programs at leading institutions has accelerated to a frantic pace, with the announcements of the Division of Data Science at UC Berkeley,[13] the School for Data Science at the University of Virginia,[14] and the College of Computing at MIT.[15]

A common attribute of these new programs and initiatives is the cross-cutting nature of units in which these programs are or will be housed. At UC Berkeley, the Division of Data Science is seen as a catalyst that *"will stress interdisciplinary collaboration among existing departments and will work to integrate data and information science into a suite of existing academic and scientific fields."* At the University of Virginia, the School for Data Science *"will set up satellites in UVA's other schools to foster interdisciplinary collaboration, as well as standalone centers focused on topics such as machine learning, analytics and visualization."* At MIT, the College of Computing *"will be an inter-disciplinary hub for work in computer science, AI, data science, and related fields […] allowing the future of computing and AI to be shaped by insights from all other disciplines*

[10] The term "Data Science" is attributed to William S. Cleveland who introduced it as part of his advocacy of "an action plan" for the statistics community to embrace computing by "expanding the technical areas of the field of statistics" – see http://brenocon.com/Cleveland1991DataScience.pdf

[11] https://news.cs.washington.edu/2013/11/12/uw-berkeley-nyu-collaborate-on-37-8m-data-science-initiative/

[12] http://msdse.org/

[13] https://news.berkeley.edu/2018/11/01/berkeley-inaugurates-division-of-data-science-and-information-connecting-teaching-and-research-from-all-corners-of-campus/

[14] https://techcrunch.com/2019/01/18/uva-data-science-school/

[15] http://news.mit.edu/2018/mit-reshapes-itself-stephen-schwarzman-college-of-computing-1015

[and giving] MIT's five schools a shared structure for collaborative education, research, and innovation in computing and AI."

Another noteworthy attribute of these initiatives (as well as many others) is the extent to which societal, ethical, and public policy dimensions of computing in general, and of data science and AI in particular, are front and center. At MIT, the new College of Computing *"will educate students in every discipline to responsibly use and develop AI and computing technologies to help make a better world, [and] transform education and research in public policy and ethical considerations relevant to computing and AI."* At UC Berkeley, the new Division of DS programs aims to ensure that *"the next generation of tech comes from people who really understand social science, and the next generation of humanities thinkers really understand what the tech people are doing."* At Stanford,[16] the new Institute for Human-Centered Artificial Intelligence offers courses on *Politics of Algorithms*, on *Ethics, Public Policy, and Technological Change*, and on *Regulating AI*.

Finally, it is important to note that as with many terms in the public lexicon, the meaning of "data science" is increasingly elusive, encompassing much of what has been traditionally associated with computing. This includes such diverse fields as artificial intelligence (AI), natural language processing (NLP), network science, cloud computing, cybersecurity, data and information privacy, graphics and visualization, among others. At Columbia, the Data Science Institute includes research centers and master's degree specializations in theoretical foundations, computing systems, smart cities, cybersecurity, business/health analytics, and social media.[17]

Looking ahead, it is hard to predict whether "data science" will hold on to the mantle of interdisciplinary *lingua franca* or whether it will be eclipsed by some other disruptive information technology that captures the public imagination. The recently announced Presidential "Executive Order on Maintaining American Leadership in Artificial Intelligence," which could be read as a manifesto for investments in data science and indeed in all of computer science, highlights the fluidity of these terms.[18]

Data Science at Boston University

Today, as with the national landscape, a large number of academic departments and programs as well as research centers and initiatives at Boston University are engaged in data science and related activities.

[16] https://hai.stanford.edu/research/human-impact
[17] https://datascience.columbia.edu/data-science-centers
[18] https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/

Academic Units and Programs

BU's presence in data science is defined by undergraduate and graduate programs in a number of cognate disciplines. Academic departments and associated faculty primarily engaged in data science include Computer Science (College of Arts and Science [CAS]), Mathematics & Statistics (CAS), Electrical & Computer Engineering (College of Engineering [ENG]), Computational Biomedicine (School of Medicine [MED]), Information Systems (Questrom School of Business [QST]), Biostatistics (School of Public Health [SPH]), Actuarial Science (Metropolitan College [MET]), and Computer Science (MET). In addition, a number of academic divisions and special programs contribute to the BU data science landscape, for example the University-wide Bioinformatics Program, the Emerging Media Division in the College of Communications (COM), the Systems Engineering Division in ENG, and the Tech & Cyber Law Clinic in the School of Law (LAW).

These units offer a broad range of degree programs, many of which have emphases or specializations that are directly related to data science, such as the MS in Computer Science with specialization in Data-Intensive Computing, the MS in any Engineering discipline with specialization in Data Analytics, the MS in Statistical Practice, the MS in Business Analytics, and the core data science undergraduate curriculum in engineering. New data science-related programs are also under consideration, some of which are slated for 2019, including the MS in Artificial Intelligence (CAS) and the MS in Applied Data Analytics (MET). In addition to on-campus degree programs, some of these units provide options for on-line degrees and certificates, such as the Digital Business Certificate from Questrom Digital and the Computer Information Systems with a concentration in Data Analytics MS and Graduate Certificates in Health Informatics as well as Data Analytics from MET.

While varied, these programs can be grouped into two major categories: those focusing on the underpinnings of data science (core) and those focusing on the applications of data science (applied). Within each of these categories, further distinctions may be made based on the target student population, *e.g.,* residential versus online, full-time versus part-time, undergraduate versus professional.

Enrollment Trends

Undergraduate enrollments in departments that offer degrees related to the core disciplines of data science (Computer Science, Electrical and Computer Engineering, and Mathematics and Statistics) have accelerated significantly in recent years, mirroring national trends (Figure 4), and

these enrollments show no signs of slowing down.[19] Interest in these programs is also evident in the number of undergraduate students who choose them as a minor. In the Fall of 2018, a total of 185 students declared computer science as their minor, 164 students declared mathematics and statistics as their minor, and 32 students declared electrical engineering or computer engineering as their minor. The graduate enrollments in master's programs in these disciplines have also accelerated significantly (Figure 5). Similar trends are also evident in the growth of graduate programs in a number of applied data science fields, such as Mathematical Finance in Questrom (Figure 6). This trend is shared across various student populations, especially for master's degree programs for full-time students in core and applied areas of data science as well as for part-time, professional education students (Figure 7).

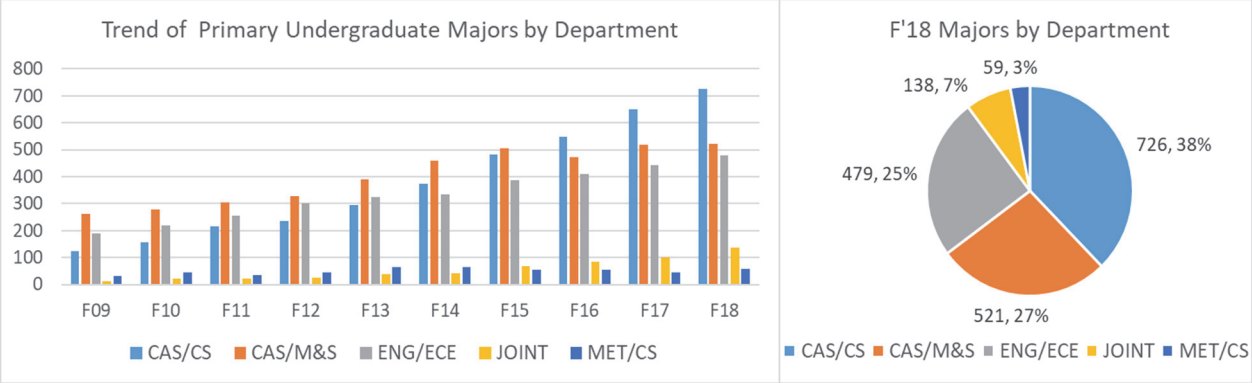

Figure 4: Undergraduate enrollments in majors related to core data-science disciplines at BU.

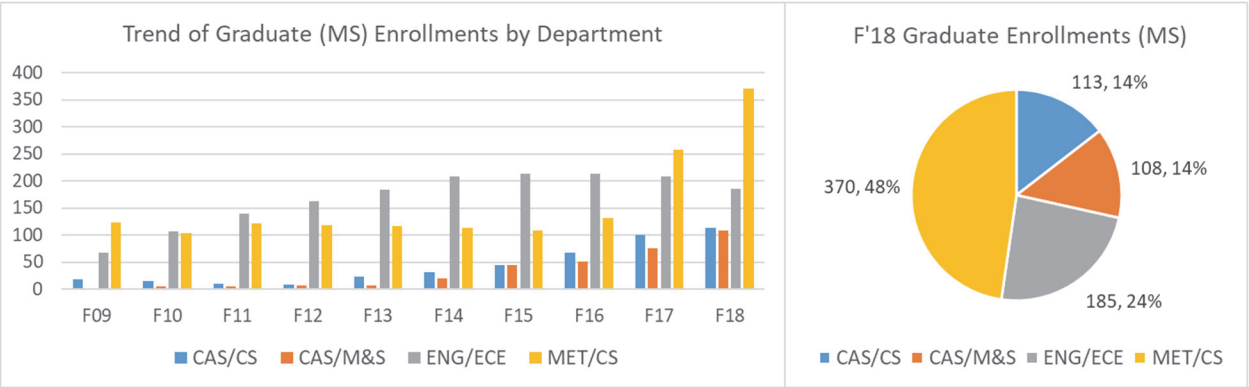

Figure 5: Graduate (MS) enrollments in programs related to core data-science disciplines at BU.

[19] Data in Figure 4 and Figure 5 represent the total number of students enrolled in degree programs and joint degree programs offered by the Computer Science Department in CAS (CAS/CS), the Mathematics and Statistics Department in CAS (CAS/M&S), the Electrical and Computer Engineering Department in ENG (ENG/ECE), and the Computer Science Department in MET (MET/CS).

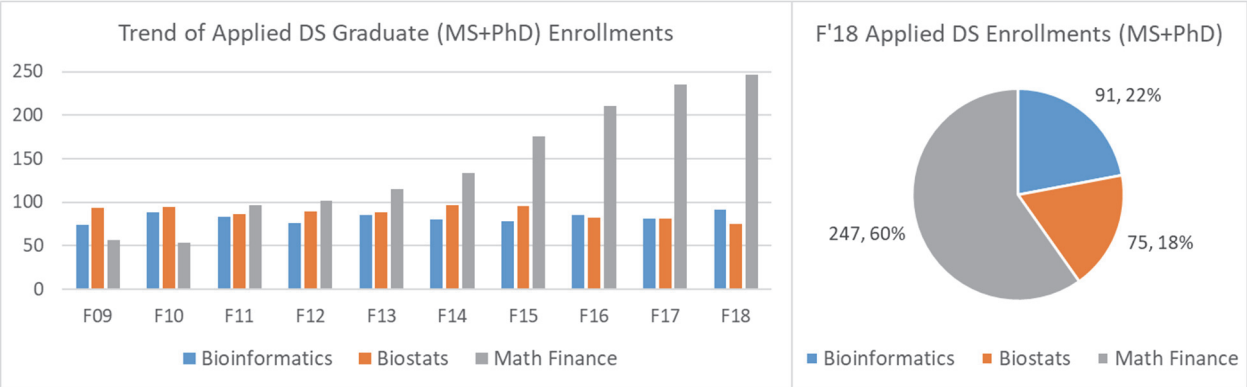

Figure 6: Graduate enrollments in representative applied data-science programs at BU.

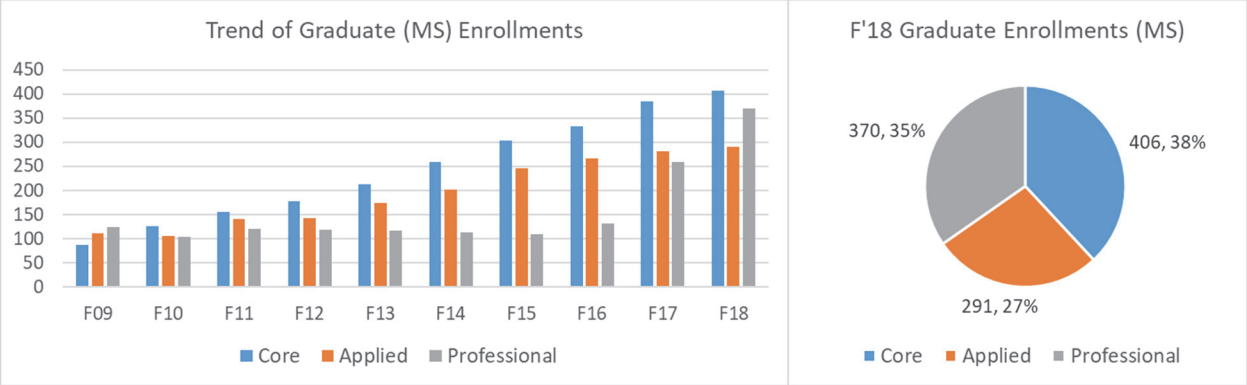

Figure 7: Graduate (MS) enrollments in data science for different student populations at BU.

Research Centers and Institutes

In addition to academic units and programs, a number of research centers have been integral to the ethos of data science at BU. Dating back to the early 1980s, BU has a proud history of investments in research areas that contributed to the emergence of data science, most notably in cognitive and neural systems[20] (the precursor to deep neural networks) and in scientific high-performance computing[21] (the precursor to cloud computing). This includes the Center for Computational Science (CCS), which was set up in 1990, and the Center for Information and Systems Engineering (CISE) and the Center for Reliable Information Systems and Cyber Security (RISCS), which were established in the early 2000s.

[20] http://cns-web.bu.edu/about/about.html
[21] http://www.bu.edu/today/2006/researchers-get-to-know-new-supercomputer/

The Hariri Institute for Computing

Recognizing the potential for data-driven inquiry to transform entire disciplines, in 2011 BU made a conscious decision to invest in creating the Hariri Institute for Computing – a hub that connects researchers who advance the science and practice of computing with researchers who pursue computational and data-driven research in various disciplines. Inaugurated in 2012, the Hariri Institute for Computing[22] was organized as a federation of units, including the existing University-wide Center for Computational Sciences (CCS) and Center for Reliable Information Systems and Cyber Security (RISCS). Since its inception, the Institute has served as an incubator for a number of highly successful labs and initiatives, including the Cloud Computing Initiative, the Cyber Security, Law, and Society Alliance, the AI Research Initiative, the Digital Health Initiative, the Red Hat Collaboratory, the Software & Application Innovation Lab, and BU Spark.

The Data Science Initiative

To *"leverage BU's existing strengths and to further expand its capacity to compete and lead in the big-data revolution,"* in the summer of 2014, the Provost launched the Data Science Initiative (DSI).[23] Based at the Hariri Institute for Computing, DSI had two complementary goals to (a) energize, strengthen, and connect BU faculty working on advancing data science methodologies with those looking to apply these methods in their own disciplines, and (b) recruit interdisciplinary faculty with proven records of accomplishment in data science and strong potential for long-term impact at BU and beyond. In 2017, the Provost expanded DSI with the announcement of the Data Science Faculty Fellows program.

Experiential Learning Programs and Practica

A number of programs at BU engage students in experiential learning opportunities that expose students to real-world data science applications, which are sourced from BU projects and labs as well as from external partnerships. At the undergraduate level, a particularly successful program has been the BU Spark! X-Lab program at the Hariri Institute,[24] which matches curated projects from internal (BU) as well as external (industry, local government, and non-profit) organizations with students in a set of BU courses that have project requirements. MetroBridge is a similar program, focusing on applications from partnering local governments, that has also been recently launched through the BU Initiative on Cities.[25] The Engineering Senior Design Project[26] offers yet

[22] http://www.bu.edu/computing
[23] http://www.bu.edu/datascience
[24] http://www.bu.edu/spark/students/spark-programs/#x-lab
[25] https://www.bu.edu/ioc/metrobridge/
[26] https://www.bu.edu/eng/prospective-undergraduate/design-project/

another model for experiential learning through undergraduate engagement with external partners in projects, many of which have data science components. At the graduate and professional levels, a number of programs exist, including CAS MSSP Consulting,[27] and Law School Technology Clinic.[28] Direct practical training with real-world problems is at the heart of MSSP's consulting model, in which graduate students work continuously through all phases of their study directly with clients, who are researchers in a number of BU schools/ college as well as in organizations in the greater Boston Area.

Research Support and Services

A number of units within BU support research and education in core and applied areas of data science. This includes the BU Information Systems & Technology Research Computing Group, the BU Library Office of Digital Initiatives & Open Access, the BU Software & Application Innovation Lab at the Hariri Institute, the BU Bioinformatics Hub in MED, the Statistical Consulting arm of the Master's of Science in Statistical Practice program in CAS, the Technology & Cyber Clinic in LAW, and the BU Office of Digital Learning & Innovation.

3. Interdisciplinary Models at Boston University

There are many interdisciplinary models across BU on both campuses, including the Systems Engineering and the Materials Science & Engineering Divisions in the College of Engineering. However, few are completely independent of existing schools and colleges. Two, which the committee investigated, are the program in Bioinformatics and the Graduate Program in Neuroscience. Each of these programs offers valuable lessons, models and warnings.

Bioinformatics

In 1998, BU established one of the first comprehensive graduate training programs in Bioinformatics. The program was designed in anticipation of the need for close integration of computation and wet lab biology. Trainees are strongly encouraged to have two mentors, one computational and one experimental, and to spend time performing experiments to develop a better feeling for the origin of the data to which bioinformatics methods would be applied. The Bioinformatics Program effectively brings together faculty members from the Medical Campus with faculty from Engineering and Arts and Science on the Charles River Campus, to participate together in research projects and graduate training in Computational biology. Approximately 60

[27] http://sites.bu.edu/mssp-consulting/
[28] https://sites.bu.edu/techlaw/practice-areas/

Ph.D. students and 25 Master's students are enrolled at any given time. PhD students are equally divided between the two campuses. Since its establishment, the Bioinformatics Program has been funded by federal training grants. However, the program has found it difficult to kept pace with other programs around the country (and world). This is largely because the Program only has three direct faculty lines, and lacks resources for joint hiring. This makes the program dependent on hiring and participation of faculty housed in other departments and prevents the program from defining its own independent educational or research agendas.

Neuroscience

The broad range of neuroscience research at BU is coordinated through a unified community of investigators from multiple research groups of our Charles River and MED campuses. Centers bridge the efforts of individual laboratories and their leaders via scientific retreats, innovative workshops, and collaborative research grants that stimulate new scientific partnerships and a vibrant community of scholars with university wide affiliations. In addition to their training efforts for department based programs that have a rich history at BU (Anatomy & Neurobiology and Pharmacology & Experimental Therapeutics), neuroscience faculty also support two major independent, University-wide training programs: Undergraduate Program in Neuroscience and Graduate Program for Neuroscience (GPN). The goal of these programs is to allow students to move freely between departments, schools, and colleges. The Neuroscience community is enriched by students coming from a variety of undergraduate majors including biology, chemistry, computer science, engineering, mathematics, physics, and psychology. However, the Neuroscience community does not have its own faculy lines or independent resources and this constrains efforts to offer targets graduate courses, build programs, and chart and independent path.

4. Committee Work and Process

The committee met regularly during the fall and spring semesters, including holding a half-day retreat midway through its work. Members read widely about data science, undertook research on extant data science research, faculty, and education programs at Boston University, and completed an investigation about data science initiatives at other major US universities, considering their genesis, structure, and public profile. The Committee also sought written input from interested members of the BU community, held two town hall meetings for interested members of the community, one on the Charles River Campus and the other on the Medical Campus, and held a briefing during the 2019 BU Data Science Day. These were opportunities for faculty, staff and students to express their vision of data science at Boston University. In addition,

the Committee met with President Brown to discuss his thoughts about data science and his impressions of efforts currently underway across the United States.

During initial meetings, discussion was wide-ranging. The co-chairs presented information on data science at BU and beyond, offered alternative organizational models for data science at BU, and facilitated discussion. In preparation for the retreat, groups of committee members were assigned to focus on specific areas where the committee hoped to offer recommendations about data science at BU. Each group made a presentation at the retreat, and the committee came to broad agreement on several areas. In terms of recommendations, the co-chairs encouraged the committee to arrive at consensus, but short of consensus called for super-majorities for any recommendation to be included in our report and offered members who had concerns about any recommendation to provide dissenting statements.

To facilitate sharing of documents, reports, and minutes, and also to enable committee members to comment on or exchange notes, a BU password-protected wiki was set up as a repository for the taskforce. Courtney Mansfield kept complete and careful committee minutes and managed all committee logistics.

5. Committee Recommendations

Recommendation #1: Vision and Mission Statement

To energize all those at BU whose research and educational efforts touch on data science, the committee believes that the mission of any new high-profile institutional home for data science at BU should be deeply inclusive and that its boundaries should be porous. While the committee believes a new data science unit creates an opportunity for new forms of collaboration across schools, colleges, and disciplines, there is a risk that the creation of any such entity at BU could leave some feeling that established research and education programs that intersect with data science could be weakened by the establishment of such a home. To avoid undermining the potential for collaboration and weakening the spirit of innovation across BU, the committee believes that the mission statement of any such entity should stress that the establishment of a new home for data science at BU is intended to be porous and inviting and should not control or weaken existing efforts across the University. The new unit should help coordinate and integrate, allowing the emergence of a synergistic enterprise that is larger than the sum of its components. In addition, given the historic underrepresentation of women and people of color in many of the fields associated with data science, the new data science effort should articulate a commitment to having a diverse faculty and student body.

Consistent with what the committee sees as the promise of data science at BU, it proposes that the mission statement for any new Data Science Unit (DSU) that serves as the institutional home for data science at BU incorporate language exemplified by the following text:

> *Data Science is everywhere and promises to revolutionize areas from education to commerce. In keeping with Boston University's history of innovative interdisciplinary research and education and its focus on excellence across a remarkable range of undergraduate, graduate, and professional programs, with a strong foundation in the liberal arts and sciences, DSU serves as a hub of data science related activities across the Charles River and Medical Campuses. The DSU aims to build a diverse faculty and student body and to develop a vibrant and inclusive environment. The DSU aims to facilitate the creation, implementation, and coordination of data science-related programs and to provide expertise, resources, advice, and support to faculty, students, and staff in units across the University. The DSU seeks to promote programmatic integration and synergy of data science at Boston University, especially facilitating collaboration and the integration of research and education. The DSU endeavours to project a cohesive view of data science at Boston University that avoids needless duplication and competition.*

Recommendation #2: Organization and Governance

Given the current and projected pervasiveness of data science at BU, the committee recommends that a new unit—henceforth referred to as the **Data Science Unit** (DSU)—be set up as the institutional home for data science at BU.

As depicted in Figure 8, while the DSU should be established as a separate entity that is independent of existing schools and colleges, it must also be permeable—allowing people and ideas to move in and out of the unit with ease and permitting easy and effective collaboration between the DSU and established schools, college, departments, and other entities at BU.

The mandate of the DSU should be to support and advance both data science education *and* research at BU. Towards this end, the DSU should develop its own strategy and pursue its own objectives, but it should do so with a collaborative spirit, seeking input from across BU. It should also strive to support and enhance existing and future educational and research programs in other units. The DSU should be led by a **Head** who is appointed by and reports to the Provost.

The internal organization of the DSU should reflect the symbiotic relationship between education and research and should recognize the diversity of research areas that are germane to data science, even if such areas are not exclusively contained within data science. As such, the DSU should have a number of subunits that operate in tandem. The DSU should also include subunits

that oversee and support the administrative and technical infrastructures and services needed for both education and research.

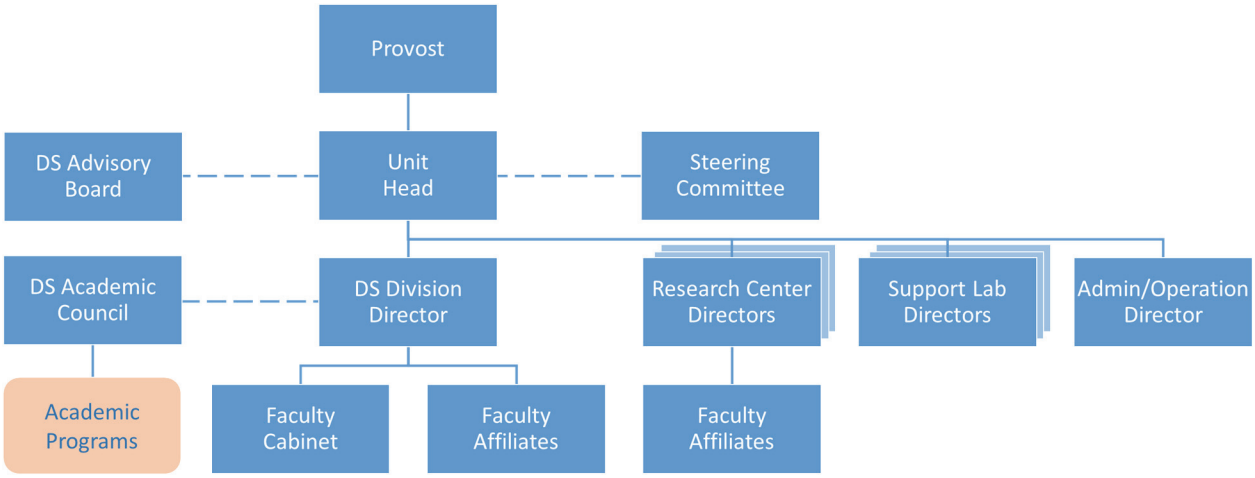

Figure 8: Data Science Unit Organizational Structure (for a functional view, refer to Figure 9).

The DSU's academic programs should be under the purview of a single subunit to be constituted as an **Intercollegiate Division of Data Science**. The Division of Data Science should be headed by a **DS Division Director** who reports to the Head of the DSU.

The DSU's research programs should be under the purview of a number of university-wide **Thematic Research Centers and Labs** focusing on interdisciplinary areas of DS-related research. The DSU Research Centers and Labs complement centers and labs, which are set up under other schools and colleges to be better aligned with the application of data science to specific disciplines in these schools and colleges. Each DSU research center or lab should be headed by a **Center/Lab Director** who reports to the Unit Director and should be open to wide participation from interested faculty, staff, and students from across BU. As university-wide centers, the operation of each DSU research center should be governed by a charter approved by the Provost and should be subject to renewal as a result of periodic sunset reviews conducted by the Vice President and Associate Provost for Research.

To avoid duplication and to leverage the efficiencies gained from pooling of resources, technical support, consultancy services, and administrative capacities that are commonly needed for data science research and education should be organized as DSU subunits. Each such DSU subunit should be headed by an **Executive Director** who reports to the Unit Head.

Both the DSU Head and the DS Division Director should be appointed by the Provost based on a deliberative process similar to that used for identifying deans or major university officers.

The DSU's governance structures should ensure that constituents act collaboratively as a federation of subunits that leverage each other and benefit from a common infrastructure of shared resources and services. As such, the DSU's governance should be carried out through four bodies, which are described next.

The **DS Faculty Cabinet** oversees the faculty search, hiring, tenure, and promotion processes of the DS Division. The DS Faculty Cabinet comprises faculty members with 50% or greater appointments in the DS Division. As such, the DS Faculty Cabinet is meant to be representative of all the disciplines invested in the division, especially through joint faculty appointments (see recommendation #4). This should ensure engagement with, and representativeness of, a broad set of disciplines, while allowing the DS Division to pursue its mission relatively autonomous of cognate units.

The **DS Academic Council** oversees "DS-branded" academic programs and curricula. This includes review and approval of new DS Division programs and curricula, periodic evaluation of learning outcomes for existing DS Division programs and curricula, and providing assistance with, and feedback on, DS-related programs and curricula offered through other units. The DS Academic Council is comprised of a certain number of members appointed by the Faculty Cabinet, an equal number appointed by the Provost and the Council of Deans, and an equal number of "at large" members elected for fixed terms by all faculty affiliated with the DS Division. As such, the DS Academic Council is meant to be representative of expertise that spans a broad set of disciplines, including those concerned with the human, societal, and ethical dimensions of data science and its applications, and to ensure broad input from across BU.

The **Research Steering Committee** oversees the allocation of resources to various DSU research activities, including research development, incubated initiatives, and research support. The Research Steering Committee comprises all DSU Center and Lab Directors and an equal number of "at large" faculty appointed for a fixed term by the Vice President and Associate Provost for Research upon the recommendation of the Unit Director. As such, the Research Steering Committee reflects a federated approach to governance of the research arm of the DSU.

The external **Advisory Board** provides feedback on overall direction and strategy for both education and research. The external Advisory Board is assembled by the Head of the DSU in consultation with the DS Division Director and with the Directors of the DSU Centers and Labs. Ideally, this body should include a diverse group of representatives from business, government, and civil society as well as academics who work directly or indirectly in the data science world.

Recommendation #3: Data Science Education

Data Science Degree Programs

The committee reached broad agreement on various aspects related to the DSU's educational mission under the purview of the DS Division. In its deliberations, the committee focused on three distinct categories of data science programs and curricula: (1) Unqualified Data Science Programs; (2) Data Science + X Programs; and (3) X + Data Science Programs.

Unqualified **Data Science programs** (DS programs for short) refer to programs and curricula that are focused exclusively on *core data-science competencies*. Core data-science competencies refer to theoretical and methodological dimensions that underlie the design, analysis, and implementation of the tools used by data scientists in the field. Advances along these dimensions are the primary drivers for innovation in a broad range of applications that leverage these tools. DS programs are starting to develop at a number of leading institutions, including undergraduate programs at UC Berkeley, US San Diego, UC Irvine, and the U of Michigan; master's programs at a host of schools such as Columbia, NYU, Harvard, Brown, Duke, and U of Virginia; and PhD programs at a few institutions, such as NYU and Tufts. The committee recommends that the DS Division be charged with the development and offering of standalone unqualified programs in Data Science at all educational levels.

The **Data Science + X programs** (DS+X for short) refer to programs and curricula that approach data science from the perspective of one of its methodological source disciplines. Examples of DS+X programs include "Computational DS" and "Systems Data Science" at CMU, "Statistical DS" at UC Davis, "Human-Centered DS" at CMU and at U of Toronto. The committee recommends that the DS Division be charged with the development and offering of DS+X programs at all levels, in close collaboration (and possibly jointly) with the cognate (X) fields providing the specific perspective for degree program.

The **X + Data Science programs** (X+DS for short) refer to programs and curricula that are primarily focused on the application of DS to a particular substantive domain (X) or on a DS specialization within a particular field (X). Examples of X+DS programs include "Biostatistics," "Bioinformatics," "Medical Informatics," "Health Data Science," "Media Analytics," "Business Analytics", as well as specializations of disciplinary degrees such as "CS with specialization on Data-Centric Computing" and "ECE with specialization in Data Science." The committee recommends that X+DS program be primarily managed through cognate (X) units, with support from the DS Division as needed. We expect that some of these X+DS programs will be offered online in order to enhance access

for a broad student population, including working adults. While these programs will likely be developed outside of the DSU, the DSU should facilitate and support such efforts.

Supporting a Common Core Curriculum for Data Science

In its deliberations, the committee recognized the importance of developing a set of courses that could be used in various combinations as a Common Core for data science education and training, especially for X+DS degrees to be developed in fields that do not have the capacity to teach basic DS courses. Such a common core would establish a standard for what constitutes minimal preparation for a specialization (or even a minor) in data science. In addition, the availability of a common core provided through the DS Division would remove a significant hurdle for a number of departments to consider the introduction of DS specialization or tracks. To that end, the committee recommends that a common DS core curriculum be developed by, and managed through, the DS Division. Such a curriculum may include courses developed in collaboration with and offered by other academic units, and should be viewed as a resource for existing and future X+DS programs and is not intended to create a barrier to having these programs offer courses/curricular paths that are appropriate for their students.

In reviewing the national landscape of educational programs, two attributes were evident, especially in degree programs at leading institutions, namely the extent to which experiential learning, practica, and internships were integral to these degree programs, and the coverage of aspects related to social sciences in general, and ethical and public policy considerations in particular. The committee recommends that BU consider these attributes to be essential ingredients of any DS-branded degree programs and that the DSU develop the necessary capacities to help integrate these attributes into existing as well as new programs, whether offered through or jointly with the Division of Data Science (DS and DS+X programs) or through other units at BU (X+DS programs).

Accommodating Diverse Constituents

In its deliberations, the committee recognized the diversity of constituents targeted by DS degree programs and the importance of developing programs that are appropriately tuned to the preparation of these constituents. Committee members raised concerns about the mathematical and/or computational maturity of students in disciplines that do not require much background in mathematics or computer science, as well as the preparation of students joining professional certificate or degree programs for career advancement. The committee believes that its recommendation to put DS and DS+X programs under the purview of the DS Division, while leaving X+DS programs under the purview of various units across the BU landscape strikes a good

balance, ensuring uniformity of "DS-branded" programs, while providing flexibility for DS specialization for existing programs and constituencies.

The committee does not foresee the need to introduce new mechanisms for cognate approvals of new programs developed by the Division of Data Science or by other units at Boston University. Existing processes at the University level for cognate approval of new programs or changes to existing programs should be followed, thus ensuring that the DS Division is afforded appropriate opportunity for oversight over X+DS or special programs, and that other units are afforded appropriate opportunity for input and oversight over DS and DS+X programs.

Recommendation #4: Faculty Appointments

Joint and Affiliated Appointments

In order to succeed, the DSU must facilitate faculty research collaboration, enable collaborative teaching across departments, schools, and colleges, make possible teaching on one campus by faculty appointed on the other campus, and systematically lower barriers to integrative data science efforts. Toward this end, as the hub for data science at BU, the DSU Division of Data Science will need to have faculty members who also hold budgeted joint appointments in schools and colleges across Boston University.[29]

Additionally, to encourage broader faculty participation, the DSU should offer affiliated (secondary) appointments to faculty members who are engaged in the academic research and education mission of the DSU, whether through its Division of Data Science or through its research centers and labs. These appointments should be subject to some baseline standards (*e.g.,* co-advising DSU students or teaching courses offered by the Division) to ensure that affiliated faculty are vested in the unit.

Tenured Appointments

In a limited number of instances, the Division of Data Science will need to have faculty for whom the division is their tenure home and from which they receive 100% of their salary. These faculty members, including the Division Director, would serve as the personnel anchor for the Division. Of course, BU will need to be judicious in making 100% data science appointments, since doing so has the potential to result in tensions over resource distribution—perhaps even creating a

[29] The successes and challenges of two of our existing autonomous interdisciplinary graduate programs—Bioinformatics and the Graduate Program in Neuroscience—testify to the need for the DSU to have faculty with budgeted appointments (joint and 100% in the unit). The former has budgeted faculty appointments and consequently finds meeting teaching obligations easier than the latter, which does not have budgeted faculty appointments.

new academic silo. At the same time, it is possible to imagine instances in which the value offered by truly interdisciplinary innovative scholars and teachers will not be fully appreciated by departments committed to disciplinary boundaries. In these instances, 100% appointment in the data science entity may be called for.

An important constraint on collaboration across the Charles River and Medical Campuses is their divergent models for faculty appointment and salary provision. In order to facilitate collaboration across the campuses, a certain number of faculty members will need to be jointly appointed on the Charles River *and* Medical campuses. The committee recommends that BU consider having these faculty members be eligible for tenure on the Charles River Campus. In all likelihood, such faculty members should have most, if not all, of their salaries provided for, independent of successful grant acquisition. In addition to such jointly appointed faculty members, the Division of Data Science should have funds available to support teaching by Medical Campus faculty in programs under the auspices of the Division.

Bootstrapping and Growth of Faculty Body

When the new Division of Data Science is initiated a process will need to be developed and budgeting made available to enable a limited number of existing Boston University faculty to move some portion of their appointments/ faculty lines to the Division. These will be the first jointly appointed data science faculty. Over time, the committee expects that the faculty in the Division of Data Science will need to grow to a size that is commensurate with the scale of the Division's programs through joint recruitment and appointment of new faculty. While the committee does not recommend a target size of the division's faculty body here, the committee noted that recent announcements by peer and peer+ institutions of faculty growth provide a good indicator of the scale needed for the division (anywhere from 20 faculty at Northwestern to 50 faculty at MIT).

Searches for new jointly appointed data science faculty should be carried out in a collaborative manner between a traditional department and the Division. The two units should work together to craft a position description and to develop a search process. They should be clear at the outset how the new faculty member's teaching and service responsibilities will be divided between the department and the division.

Given the historic underrepresentation of women and people of color in the disciplines at the current heart of data science, the new unit should be especially attentive to building a diverse and inclusive faculty. Getting to this end point will require developing a specific set of search and hiring practices.

Recommendation #5: Resources and Incentives

The committee was in strong agreement that DSU will need to have its own well-provisioned budget. The majority of the committee thought it best not make recommendations in matters of budget size and specific areas of funding. The committee does, however, believe that providing the unit with a tuition share return for master's and graduate certificate programs under its auspices (similar to that received by existing schools and colleges) makes sense. This could allow the unit to support first-year PhD students in its programs and provide it with flexible funds to use for faculty start-up packages or innovative initiatives. This would also allow the unit to develop and sustain resources and capacities that could be valuable for other units, including infrastructural and consultancy support for teaching and research and to act as a clearinghouse for information about data science expertise and resources at BU.

To facilitate movement of ideas, faculty, research, and educational programs across Boston University and to build bridges between established departments, schools, and colleges with well-defined territories will require a carefully crafted system of incentives. These incentives will need to take into consideration the distinctive budget models of the Charles River and Medical Campuses and the established commitments of existing schools, colleges and departments.

To facilitate data science educational programming across BU, faculty will need to have easy access to classroom space and, as appropriate, teaching fellows or staff assistants, as well as infrastructural and technical support for courses that require hands-on computational work. Participation in teaching activities by faculty not appointed in the Division will require an array of incentives, including those related to salary, credit for teaching outside of home departments, access to data science PhD students, and mechanisms to facilitate research collaboration (including discretionary funds, grant application support, and pooled research resources).

In its deliberations, the committee noted the importance of giving the leadership of the DSU enough flexibility and agility to use centrally-provided resources to incentivize faculty to engage with it. Some members of the committee also noted that as a system of incentives is put into place, those involved in establishing the incentive model should be keenly aware of the possibilities that certain incentives could reinforce territorial boundaries and undermine synergy.

The availability of adequate physical space is an essential catalyst for research collaboration among faculty and students who hail from different programs. Given the pervasiveness of data science, it is inconceivable to imagine confining such collaborations to a single physical location on campus. In particular, while the planned new Data Sciences Center on the Charles River Campus will serve as an anchor for data science at BU, it is unlikely that it will come close to

providing adequate collaborative space for all potential data science collaborations at BU. As such, the committee believes that adequate, satellite space for groups of collaborators will need to be provisioned across the BU Charles River campus and Medical campus. Thus, while the new building will serve as the hub for data science, BU also needs to have spokes.[30] In a way, this is the nature of data science, and this is how BU can articulate its cross-cutting, integrative nature, physically speaking.[31] Having spokes for data science across BU will also create incentives for cognate units and schools to connect and be invested, and perhaps more importantly, mitigate the risks of DSU being perceived as—or, indeed, becoming—a new silo.

Recommendation #6: Engaging Industry, Government, and Civil Society

Unlike traditional science and engineering disciplines in which academic research tends to be too basic and/or abstract to be of value to business, government, or the non-profit sector, and in which the challenges entities from these sectors face tend to be too translational and/or applied to be of scholarly value, data science offers a point where scholarship meets real-world impact. Indeed, the most relevant and impactful scholarly work in data science tends to be connected with real-world applications, typically through access to unique data sets and practitioners.

In the Moore-Sloan survey of top data science initiatives in the US, eight of the twenty initiatives they considered (at Columbia, MIT, NYU, Stanford, U of Chicago, U of Michigan, U of Rochester, and U of Massachusetts) had thriving industrial partnership programs, with most of the rest being funded by industry on an ad hoc basis. In that survey, industrial funding was seen as particularly attractive to faculty because *"it offers flexibility and larger budgets compared to the government funders, and because the practical nature of the problems of interest to industry resonates with faculty."* Development of a strong partnership program with companies, government, and non-profit organizations around data science should be a priority for BU.

In its review of the national landscape of data-science-branded academic programs, it was evident to the committee that internships and practica were not only essential but also a differentiating component of leading programs. Indeed, of the twelve MS in DS programs reviewed, only one did not have a practicum requirement.

As summarized earlier, BU has many thriving programs that support experiential learning, with several having a practicum component that is fully-integrated into all phases of the MS program

[30] NSF is using the "hub and spoke" model for its designated Data Science research centers across the US, with Big Data Hubs serving as anchors for regional data science research and Big Data Spokes serving as focal areas for collaborations on specific data science research themes. See https://www.nsf.gov/cise/bdspokes/index.jsp.

[31] Incidentally, UVA's plans for its new School for Data Science calls for the school to command physical spaces to act as *"satellites in UVA's other schools to foster interdisciplinary collaboration."*

so that students directly and continuously gain real-world experience from just after matriculation all the way through graduation and internships. These programs present significant assets that should be further developed and leveraged not only in support of DS-related programs, but also as vehicles for engaging industry, government, and civil society.

Given the rich (and data rich) ecosystem of industrial research labs, businesses, non-profit organizations, and governments in the Greater Boston Area, BU is uniquely positioned to build partnerships for collaborative research, contract research, educational programs, and student practica. Such partnerships will be mutually beneficial. In addition to reputational and financial benefits to be accrued from such partnerships, the incentive for BU should be on the capacity of possible partners to provide experiental learning opportunities for students through internships and practica, as well as applied research opportunities for faculty through access to precious data sets. The incentive for industry, government, and non-profit organizations centers on access to students for recruitment purposes and to faculty for research and development purposes. We could also envision direct integration of industry within the classroom through joint symposia, courses, and DS+X or X+DS programs in conjunction with industry sponsors. Establishing innovative ways to integrate industry involvement in our DS curricula could provide students with perspectives they might otherwise not receive and access to industrial connections and experience.

To build such partnerships, the DSU should be staffed appropriately and should consider the extent to which the advisory board described in recommendation #2 can serve as a mechanism for developing strategy in this area and building actual partnerships.

6. Concluding Remarks

Responsive to its charge, in this report, the committee presents a set of recommendations that constitute the blueprint for organizing data science at Boston University in a way that will enable the University to remain at the cutting edge of research, to establish itself at the forefront of educational offerings, and to be highly visible nationally and internationally in this fast-evolving field.

Figure 9 shows the extent to which the proposed organization of data science at BU under a new Data Science Unit (DSU) will catalyze interconnectedness, allowing for the emergence of a synergistic enterprise that is larger than the sum of its components and ensuring that the integrative nature of data science is deeply infused into BU's forward looking 21st century culture.

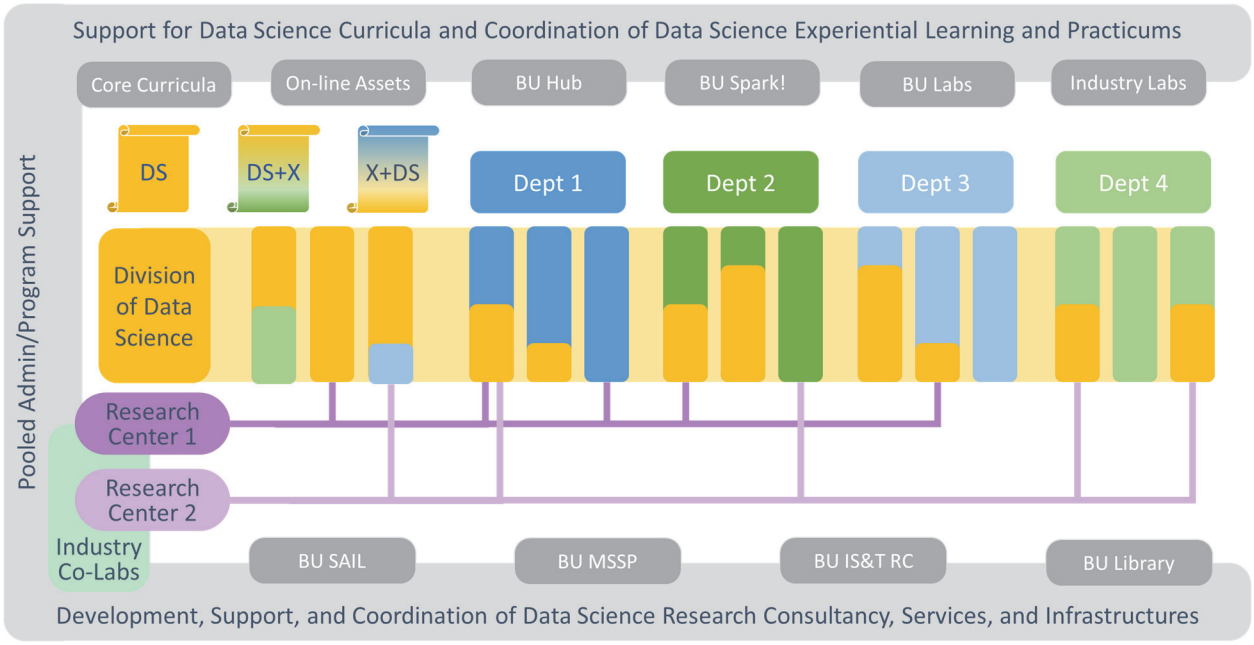

Figure 9: DSU as a catalyst for integration of research and education at BU.

Largely drawn from and jointly appointed at various levels in cognate departments from across BU, faculty in the envisioned DSU Division of Data Science will act as overseers of new educational (DS and DS+X) programs to be offered through the Division, as well as enablers of current and future educational (X+DS) programs offered through cognate units.

Similarly, faculty drawn from across the university will connect with one another through research centers that are thematically focused on broad interdisciplinary computational and data-driven research areas. These centers, along with industry-sponsored co-laboratories, will operate as a federation that leverages shared administrative and programmatic support capacities.

This ecosystem will be made complete through the integration and coordination of functions offered by units (both within and outside the DSU) that support various facets of data science research and education, as well as outreach and engagement with industry and society at large.