Proposal for a Department of Data Science

April 2021

Table of Contents

Summary of Facts and Justification for this Proposal

History

The Data Science Program was founded in 2017 in response to the nation-wide increased demand for data scientists as well as growing campus interest from both faculty and students in an interdisciplinary program that would 1) expose students to complicated problems related to the management, analysis, and dissemination of vast amounts of information, 2) teach students how to define, discuss, and solve those problems, and 3) consider the societal, moral, and ethical ramifications of data-driven decision making. W&M Data Science graduates are trained in cutting-edge methods to manage and analyze data, but also in how to effectively communicate and lead others in making ethical and fair decisions based on sound interpretation of data. Between 2017 and 2020, 47 students graduated with self-designed interdisciplinary majors in Data Science. In 2020, the Data Science Program introduced a new B.S. in Data Science and is in the process of developing a B.A. degree to meet additional student demand.

Facts and Figures

Student Enrollment and Demand

- The program has a total of 176 declared students, including 34 students currently pursuing a B.S. major. The program graduated 47 students pursing a self-designed B.S., with the last cohort of 24 students finishing this now phased-out degree. In addition, 71 students either have or still are pursuing a minor.
- Our Intro to Data Science course enrolls approximately 100 students a semester
- Across our courses, we received 175 override requests for additional seats for Spring 2021, indicating that our current course offerings do not meet student demand
- The inaugural 2020 summer Jump Start Data Science Program received twice as many applicants as available seats (25); and is full for the 2021summer program

Interdisciplinarity

- The Data Science Steering Committee has representatives from eight different programs and departments (American Studies, Applied Science, Biology, Computer Science, Data Science, English, Government, Mathematics, and Psychological Sciences)
- The Data Science teaching faculty have Ph.Ds in multiple disciplines, including Biology, Civil Engineering, Computer Science, East Asian Languages, Geography, Government, Linguistics, and Mathematics

Funding

- The Commonwealth of Virginia has established “Tech Talent” and “High-Demand Degree” initiatives from which W&M has received money to grow explicitly its Computer Science and Data Science programs, with the objective to double the number of our “tech talent” graduates by 2030. Our competition in receiving this money was not internally-based (i.e., other W&M programs), but competition with external tech talent programs in other Commonwealth Universities (e.g., University of Virginia, Virginia Tech, George Mason)
- As a result, the proposed department will be financially independent, does not burden our standard A&S operating budget, and is funded by money specifically ear-marked for Data Science in the context of tech talent.

Why does Data Science need to become a department now?

Data Science has a unique, interdisciplinary culture, expanding upon, but distinct at the same time from that in related fields, such as computer science or mathematics, which is reflected in both our teaching and research. For example, the W&M Data Science Program is firmly committed to the idea that Data Science should be accessible to all students, even those who do not think of themselves as “STEM students.” Thus far we have succeeded: the students who are secondary majors or minors in Data Science are pursuing primary majors in 15 different fields, mostly in the humanities and social sciences. **Data**

Science needs to be a stand-alone unit to allow this culture to flourish and to maintain autonomy over the interdisciplinary curriculum and opportunities offered to students and faculty. However, our current structure as a program, not a department, is inhibiting the growth needed to meet demand and reach our potential. Joint hires commonly and naturally revert to their home department, limiting their ability to fully contribute to the growth of the program in a manner proportional to the investment in it. So, while the present model of a Program was adequate to stand up the BS program, it is inadequate to sustain this effort. **Establishing a Department of Data Science (DDS) is necessary to allow the program to reach its full potential**. Many of our peer institutions have already moved forward with implementing departments of Data Science that share common characteristics: they are high-visibility programs that serve as academic discovery incubators and accelerators and are attracting high-caliber faculty hires who oversee highly successful research labs. **We aspire to the same goals but are hindered by the limitations of our current status.**

As a department, we can:
1. Recruit and retain the highest quality Data Science faculty. Data Science is a competitive field and the best candidates will want to join a department where they can contribute meaningfully to faculty governance in an autonomous unit and where a cohort of their peers will evaluate their records for merit and promotion in a manner consistent with their background. We have been unable to hire as many highly qualified TE and even NTE faculty as we have been approved to hire, due in part to the low visibility of the Data Science Program and the need to find "tenure homes" for TE faculty. We cannot continue developing a Data Science unit of the caliber of W&M by hiring non-Data Science faculty in it.
2. Better serve students from across campus with robust course offerings and research opportunities. Because most core Data Science Program faculty are NTE faculty, there are a limited number of mentored research experiences available for students. Establishing a department will also facilitate our ability to offer a B.A. in Data Science that is explicitly designed to cater to a wider range of students, whose substantive interests are in fields not traditionally associated with STEM.
3. Incentivize and reward the service commitments of affiliated faculty and strengthen faculty governance. As a program, the vast majority of TE faculty have a primary commitment to their home department and faculty with joint appointments devote time and effort to the department overseeing their retention and promotion. Despite the heroic efforts of our NTE faculty in carrying an inequitable service load, we are not able to fulfill the service obligations and governance needed to successfully administer one of the largest academic units in A&S (as measured by number of graduates).

Inclusive Plan for Departmental Development

Spring 2021	Representatives from Data Science meet with more than a dozen A&S programs and departments to solicit input and feedback Cluster hire of 2 TE and 1 NTE faculty
April 2021	FAS Meeting and Vote Board of Visitors Meeting and Vote
May 2021	Steering Committee formalizes timeline for transition in consultation with current teaching faculty and A&S Dean's Office Departmental Transition Committee appointed by Data Science Program Director, in consultation with the Steering Committee, the teaching faculty, and A&S Dean's Office
May 2021 – June 2021	Campus-wide Coordination of Transition: • Inventory of resources and needs to better inform decisions on next hires and further program development • Faculty interested in formally affiliating with the department contact Transition Committee Chair

	• Faculty interested in contributing to the development of a B.A. in Data Science contact the Data Science Program Director
July – December 2021	Departmental Transition Committee leads the effort for the structural, administrative, and governance documents and consensus needed for the new department. New TE and NTE faculty are included in process, as well as faculty who indicate an interest in affiliating with the department. Formal Application to SCHEV
Summer 2021	Development of SCHEV proposal for the Data Science Department
Fall 2021	Submit Department proposal to SCHEV Development of graduate portfolio (incl. certificate, possible M.S. and PhD pathways)
Fall 2021 & Spring 2022	Development of B.A. in Data Science Submit to SCHEV for approval

Full Proposal for a Department of Data Science at William & Mary

Slight Update of an Original Document Prepared in Fall 2020 by (in alphabetical order):

Joshua Burk (Psychology), Gregory Hunt (Mathematics), Matthias Leu (Biology), Michael Lewis (Computer Science), Elizabeth Losh (English and American Studies), Daniel Parker (English and Linguistics), Daniel Runfola (Applied Science), Jaime Settle (Government), Anthony Stefanidis (Computer Science), and Janice Zeman (Data Science Contact Dean)

I. Introduction

Since the inception of the Data Science Program in 2017, 47 self-designed majors have graduated from the program, with student enrollment steadily increasing. This student demand has been accompanied by an influx of base support for Data Science from the Commonwealth of Virginia, in the form of a statewide effort to increase the number of graduates prepared to enter into the "tech workforce" (the Commonwealth's "High-Demand Degrees" and "Tech Talent Pipeline").

With the formal Commonwealth approval of a B.S. in Data Science in the Fall of 2020, William & Mary students – inclusive of the humanities, social and physical sciences – can now choose to gain in-depth knowledge in the analysis of exceptionally large and high velocity information. Based on our experience to date, we anticipate that the interest in Data Science will drive demand primarily in the form of dual majors (i.e., students that seek application in analytical and computational advancement in addition to the foundational major).

The limitations inherent to programs at William & Mary are now restricting our capability to both meet existing student demand (self-designed Major and Minor in Data Science) and to support majors in Data Science. This document presents the challenges currently faced by the program, and our case for how a Department of Data Science would provide benefits not only for students seeking to major in the field, but also for the entire campus. Finally, acknowledging the unique global context this proposal was drafted within (namely the COVID-19 effects on society), we provide a brief discussion on how a Department of Data Science would enhance the fiscal sustainability and student-oriented outcomes at William & Mary.

II. Why William & Mary Needs a Department of Data Science

While mentions to Data Science in the broader world context may be found as early as 30-40 years ago, Data Science as we understand it today has been shaped over the past decade. Broadly, we can view Data Science as the growing scientific discipline that occupies, explores, and refines the space where computing, statistics, and data analysis meet the many disciplines that model and explain the functions and complexities of the world around us.

To better convey the operational scope of the proposed department, Figure 1 shows the five key pillars of Data Science as presented by UC Berkeley's School of Information. The five pillars demonstrate the interdisciplinary approach in Data Science, highlighting the need for a hub to foster and maintain network connections among pillars.

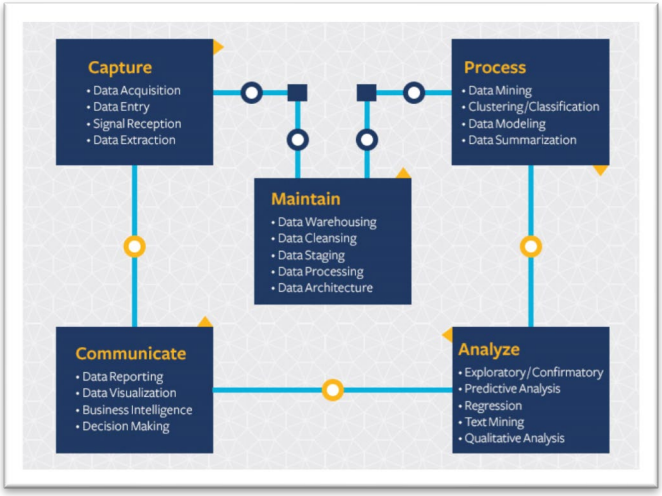

Figure 1: The five pillars of Data Science

Broadly, many of our peer institutions have already moved forward with implementing stand-alone units, departments – or, even schools – of Data Science. These include the Data Science Institute at Columbia University, the School of Data Science at the University of Virginia, the Computing, Data Science, and Society Division at UC Berkeley, and most recently the Academy of Science at Virginia Tech. These are part of a growing number of Data Science undergraduate and graduate level formalized academic divisions that are shaping the relevant academic landscape.

These units share some common characteristics. They are all **high-visibility programs** within their institutions thereby attracting donations to enhance computing and data science programs that ultimately benefit an institution as whole (see UC Berkeley donation https://bit.ly/2WIYdnR). Furthermore, they tend to function as campus-wide **academic discovery incubators and accelerators**, supporting a wide array of activities, from computational social science, to digital humanities, health informatics, and business. They also tend to attract **high-caliber hires** from diverse backgrounds to advance their broad academic mission, and produce **heavily recruited graduates** at both the undergraduate and graduate level.

A Department of Data Science would enable students to pair Data Science majors or minors with degrees from other units within William & Mary – leading to potential improvements in their prospects on the **job market** post-graduation. Market demand for data scientists is huge (over 6500 current job openings), with data scientist listed as the top job in America for four years in a row by Forbes (see https://bit.ly/2vRVbmw), and also currently listed in the third spot by Glassdoor (see https://bit.ly/39gR8O3). William & Mary is just beginning to meet this workforce need through our tailored Data Science minor and major, which provide students with the technical breadth and theoretical depth required to apply computational analyses to exceptionally large data sets. The success of our 2019 and 2020 cohorts of students (47 self-designed majors in total) on the job market provides anecdotal evidence that the combination of a Liberal Arts approach to communication, creativity and exploration provides a potent combination with the Data Science curriculum.

Creating a Department of Data Science would further allow William & Mary to grow into a key player responding to the "*Tech Talent"* and "*High-Demand Degree"* initiatives of the Commonwealth. These initiatives aim to grow the number of skilled graduates in computing and data science. Just this past November, William & Mary received substantial *additional* support from Richmond to support the growth of these fields (starting with $1.5M in 2020-21), including funding for three faculty positions to support the Data Science program (see: https://bit.ly/2UehFY2) with the objective to grow the number of our "tech talent" graduates by nearly two-fold by 2030. Furthermore, in response to a proposal by the University, the Commonwealth is currently planning to add to this budget an additional $700K, for the hiring of faculty who are underrepresented in Data Science. To increase number of graduates, a department would increase the ability for students considering William & Mary to easily locate a programmatic major in Data Science that could be attractive to them. The department provides an easily identifiable home for such students.

III. Limitations of Programs at William & Mary and Implications for Data Science

In 2017, the Data Science Program was formally created by a unanimous vote of the Faculty of Arts and Sciences in favor of the creation of a Minor. This minor has – to date – been housed within the Charles Center, with the faculty (both TE and NTE) sitting either within the Charles Center itself (i.e., Center for Geospatial Analyses and Data Science Program), or a range of affiliated academic departments/programs (i.e., Applied Science, Computer Science, Gender, Sexuality & Women's Studies, Government, Linguistics, Mathematics).

Programs serve a critical role at William & Mary – allowing for the flexible exploration of emerging areas of academic inquiry, establishment of interdisciplinary degrees that do not have a clear departmental home, and discoverability for students. However, programs at William & Mary are critically limited in four ways: (1) they are typically not provided with any annual operational funding; (2) they typically do not have dedicated administrative support; (3) they rely on joint faculty appointments; and (4) they cannot award tenure, or house tenure-eligible positions.

These limitations dramatically inhibit the growth and sustainability of Data Science at William & Mary. To specifically highlight the limitations as they relate to Data Science:

1) **Annual Funding.** Data Science prides itself on being highly diverse – we intentionally aim to provide students who would not traditionally enter into STEM fields with an opportunity to do so. This model has proved highly successful to date. However, to support this model requires modest funding for activities such as: (a) peer learning and support, and (b) computing environments for low income students.

2) **Dedicated Administrative Support**. As the number of students in the Data Science program has grown, so too has the need for dedicated administrative support for a range of faculty travel, procurement, scheduling, recording, reporting, and related needs. Many of these tasks are currently handled by Data Science Program faculty, to the detriment of their ability to advise students, teach, or transition curriculum to online platforms to expand the reach of W&M. As an example, the inaugural summer 2020 Jump Start Data Science program was so popular with our students that we received almost 50 applications for the 25 seats that we could offer. The 2021 Jump Start cohort was filled in January 2021 for the summer 2021 program.

3) **Joint Faculty Appointments**. Joint appointments often create challenges for these faculty as their duties can be unevenly distributed across multiple programs and stretching the capability of faculty to meet their responsibilities in both programs/departments. This can prevent joint faculty members from making a full impact in either their home department or, in this case, the Data Science program.

4) **Tenure**. There are three critical challenges the program is and will be facing if we are unable to hire Data Scientists in tenure track positions.

A) As noted earlier in this document, Data Science is currently in very high demand on the job market. Because academic rates for TE positions are not competitive with the private sector, this significantly inhibits our ability to recruit top talent into the program. Elevating the Data Science Program to a department would increase visibility and thereby help us overcome this substantial challenge of attracting top research faculty for TE faculty positions.

B) Data Science is moving exceptionally quickly as a discipline, necessitating extensive collaboration with like-minded peers to stay at the cutting edge of teaching and research; this is not a unique challenge, and is faced by many departments across campus. However, as a Program without dedicated tenure positions, Data Science faculty rarely have the opportunity to collaborate on a regular or extensive basis due to the demands of home departments and interrelated tenure decisions.

C) Currently, there are certain upper-level subfields we currently cannot cover given a lack of research-active faculty. One affiliated faculty has over 60 research students in his lab. Not only does such a high enrollment reduce the research quality of each student, it also highlights the need for additional faculty to offer research opportunities. Clearly hiring TE faculty would help alleviate the high student demand on a few labs that currently include undergraduate students in research.

D) As there are only two tenured professors formally 'employed' with the program (i.e., with formal responsibilities and funding provided by the program), there is a significant service load to accompany the growth of the program. We desperately need more peers to shoulder the load, and also to better collaborate with the numerous disciplines within William & Mary that will benefit from our data science offerings.

As a result of these core challenges, the Data Science program in its present form lacks the visibility that it can and should have within the William & Mary community and beyond, and is unable to reach its full potential and foster meaningful collaborations at the undergraduate, graduate, and faculty level across disciplines and Universities. Allowing the current program to grow and become the foundation of a Department of Data Science will allow us to overcome these challenges, and will provide our William & Mary community with a new, vibrant academic hub to attract funding not only for Data Science itself, but also for its affiliated academic units.

IV. Blueprint for the Creation of the Department of Data Science

The proposed Department of Data Science is a natural next step for a foundation that is already in place at William & Mary:

- The B.S. in Data Science, approved by SCHEV in March of 2020, serves as the flagship educational offering for this new unit (in addition to supporting the existing Minor in Data Science).

- A number of faculty members with robust expertise and interest in components of the data science academic footprint are already part of our community, forming a core faculty foundation upon which to build.
- Additional funds are already in place from the Commonwealth to support the creation of a department in a way **that will not siphon existing resources**.

As a Department, Data Science would continue to seek both (a) hires specifically within Data Science, and (b) joint hires with other units on campus (i.e., joint appointments). Figure 2 shows current cross-disciplinary data science-driven collaborations within William & Mary (orange), as well as growth opportunities to be pursued by the Department of Data Science (green). As one can see in that Figure, a foundation of collaboration is already in place. For example, Data Science and Biology faculty jointly submitted a proposal to fund collaborative research. Other collaborations are seen as low hanging fruit that have not been explored yet due to the above-mentioned bandwidth limitations of the current set-up as a program.

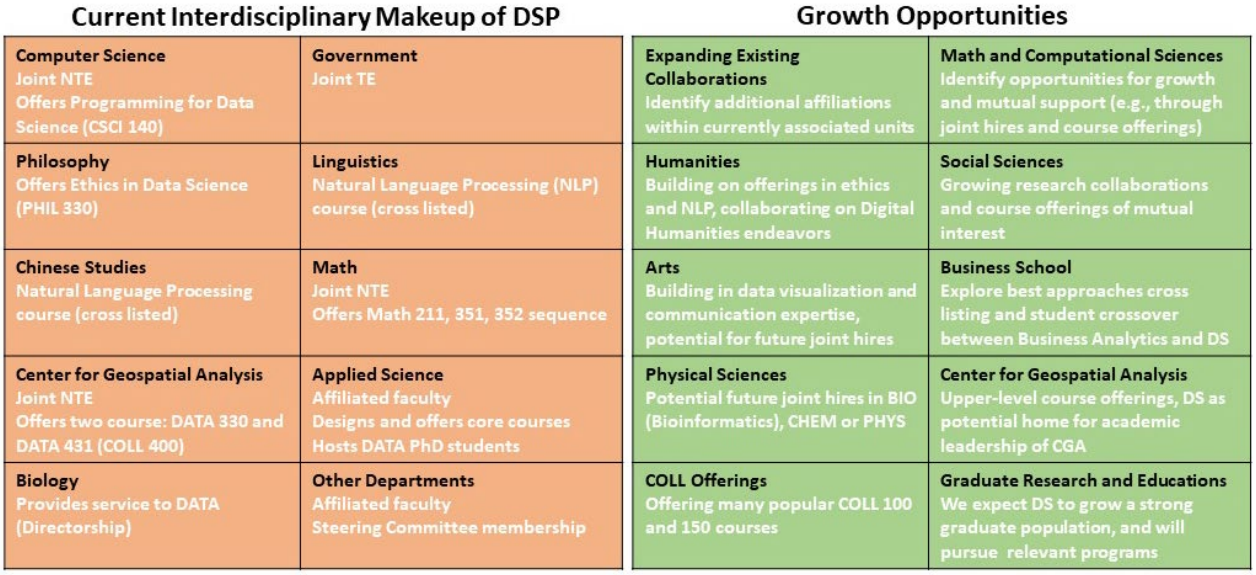

Figure 2: Department of Data Science interdisciplinary collaborations: partnerships already in place (orange; Data Science Program [DSP]), and aspirational goals (green; Department of Data Science [DDS])

The above-presented staffing approach would help promote and foster the interdisciplinary nature of Data Science. Specifically, we envision a moderate group of 'core' faculty seated fully within Data Science collaborating closely with a number of jointly appointed faculty in other disciplines.

Our goal is soon after establishing the Department of Data Science to reach an operational capacity that will approach the blueprint presented in Figure 3, comprising:

1) A core group of approximately 10 Department of Data Science faculty, requiring 2-3 new hires (i.e., 2 TE and 1 NTE, funding for positions already allocated by Commonwealth of

Virginia) in addition to 7 already-present faculty. We would aim to recruit bridges to the mathematical and physical science communities, which are currently absent or limited within the program.
2) A small group of 3-4 faculty in other departments who teach courses for Department of Data Science (pursuant to existing JAMOU arrangements).
3) An administrative assistant.
4) Three undergraduate pillars: (a) the B.S. in Data Science; (b) the B.A. in Data Science (currently in the planning phase), and (c) the Minor in Data Science.
5) An undergraduate research pillar, focused on engaging large numbers of students in both basic and applied research.
6) A graduate research pillar, focused on Ph.D. education in partnership /collaboration with the Department of Applied Science. In a later phase, we would seek to grow this pillar to include a Masters program.

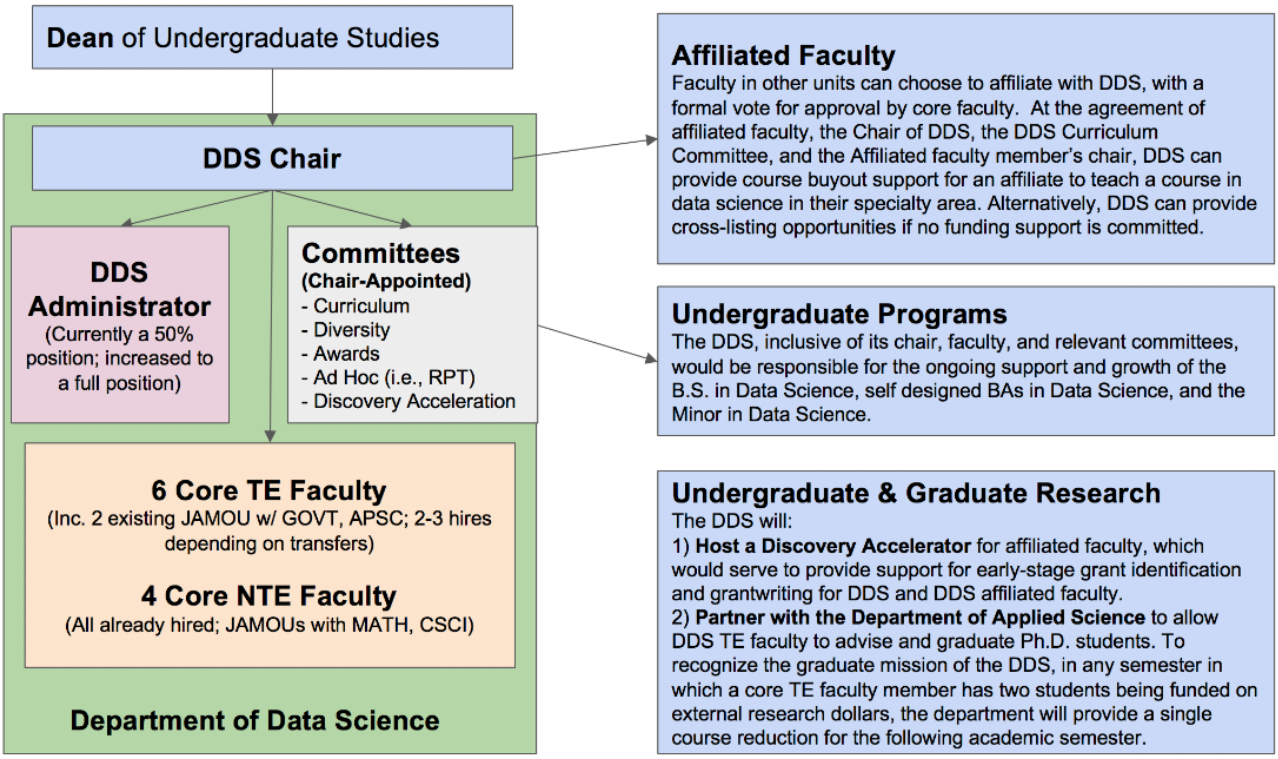

Figure 3: Operational Blueprint for the new Department of Data Science (DDS)