WILLIAM & MARY
CHARTERED 1693

DATA SCIENCE AT WILLIAM AND MARY: Rationale and Motivation

What is Data Science?

Data Science is a blend of many fields, including many subdomains of mathematics, computer science, computational science, statistics, and information science. In contrast to "pure" mathematicians, statisticians, or computer and information scientists, a data scientist has a breadth of experience across all of these fields, but may not have as much knowledge as a specialist in any particular field. Critically, a data scientist has a depth of substantive knowledge and applied experience within at least one knowledge domain – for example, linguistics, history, or political science, and is trained to understand the moral implications of new algorithms and data analysis methods. This combination will allow William and Mary Data Science students to (a) efficiently conduct computational analyses within his/her own knowledge domain, (b) manage teams of more specialized individuals to answer far-ranging questions, and (c) communicate technical findings to new types of audiences. Individuals with this set of knowledge are revolutionizing a wide set of domains, and are in very high demand not just by faculty researchers at William and Mary, but also by the public and private sector. While there is a projected shortage of nearly 200,000 individuals with Data Science skills in 2018, that number jumps to a shortage of 1.5 million when "managers and analysts with the know-how to use analysis of big data to make effective decisions" are considered.[1]

Why Data Science at William and Mary?

The College of William and Mary is uniquely positioned to build on existing strengths, not only in our world-class High Performance Computing (HPC) infrastructure, computer science and mathematics departments, but also the unparalleled liberal arts education offered in departments such as government, anthropology, and the interdisciplinary experiences of the Applied Science program and Charles Center interdisciplinary majors. Further, the high quality of scholarship offered by William and Mary has continued to attract the best scholars across all disciplines, resulting in a critical mass of faculty members pursuing computational and statistical analysis in fields that might not traditionally be associated with such work.

Data Science provides a strong venue to improve the postgraduate experience of multiple classes of students across campus - be it students interested in pursuing further academic knowledge in graduate school settings or those seeking to enter into the workforce. In both cases, the coursework provided to students will make them more competitive in STEM roles, and further increase the breadth of scholarship

[1] Considering only the United States; McKinsey Global Institute, "Big data, The Next Frontier for Innovation

offered as a part of the liberal arts education at William and Mary for students that don't participate in the full minor. Furthermore, a burgeoning population of cutting-edge faculty members at William and Mary strongly seek to engage students in research, but are frequently unable to do so due to the lack of appropriate coursework and coordination that would enable these faculty-student research collaborations. This lack of preparation has implications not only for existing initiatives on campus — including the fulfillment of COLL 400 — but also broader ramifications if William and Mary seeks to continue to attract the best scholars and students in face of a rapidly modernizing world. Acknowledging similar challenges, universities in William and Mary's peer group have begun to take similar steps, following in the footsteps of many public and private Ivies:

- "New quantitative science major unites big data with the liberal arts" - Emory University (http://news.emory.edu/stories/2014/11/er_quantitative_sciences_major/campus.html)
- Certificate in Data Science - Georgetown http://scs.georgetown.edu/programs/375/data-science/
- Department of Biomedical Data Science - Dartmouth https://bmds.dartmouth.edu/
- Data Analytics Group - Clemson http://citi.clemson.edu/ciprac/da/
- Certificate in Data Science - UCI http://unex.uci.edu/areas/it/data_science/
- MS in Data Analytics - Rutgers http://mbs.rutgers.edu/programs/analytics-data-sciences

Minor Program Proposal: Catalog Listing

Required Credit Hours: 18

Students trained in Data Science will study a blend of topics from many subdomains of communications, philosophy, mathematics, computer science, and information science. A data scientist has a breadth of experience across all of these fields, but may not have as much knowledge as a specialist in any particular field. Furthermore, a data scientist trained at William & Mary is equipped to consider the philosophical and moral implications of algorithm development and data collection, and the societal ramifications that new approaches to data manipulation could have. This combination allows William and Mary Data Science students to (a) efficiently conduct computational analyses within their own knowledge domain, (b) manage teams of more specialized individuals to answer far-ranging questions, and (c) communicate technical findings to new types of audiences. Individuals with this set of knowledge are revolutionizing a wide set of domains, and are in very high demand not just by faculty researchers at William and Mary, but also by the public and private sector.

The College of William and Mary offers a minor in Data Science, which draws on faculty expertise from many departments. There are four key pedagogic pillars students will be expected to engage with during their time in the program: Computation, Application, Communication, and Deliberation.

- Computation - the computer science and mathematics required to responsibly use large datasets to create new knowledge. This is a focus of the introductory coursework, as well as the elective data specialties.

- Application - the skills and creative thinking required to identify novel ways to apply computation to new problems. Application is present in all core courses within the Data Science program, including DSP 140, DSP 146, and APSC 200.
- Communication - the confidence to present in front of a crowd, and creativity to communicate or visualize technical concepts for new, likely non-technical audiences. Courses such as public speaking and art help promote such confidence and creativity; while many students will naturally receive some communications training as a part of their time at William and Mary, the Data Science program promotes an additional depth of skill due to the challenges in communicating large sets of data. Further, communication is a strong theme within all DSP core courses.
- Deliberation - the ability to consider the societal, moral, and ethical implications of Data Science. Data Science minors are required to take either a COLL 150 examining these topics or broader philosophy coursework.

The Data Science minor is designed to be paired with a wide variety of majors across William & Mary, so there are no restrictions on the primary major pursued in conjunction with the Data Science minor. Under most circumstances the Data Science minor should be declared no later than the second semester of the Junior year to ensure the minor can be completed. Two courses may be counted toward both your primary major and minor; some courses may be substituted with permission from the director. In addition to the required work, various other courses as well as non-classroom training (such as internships, research projects with faculty, participation in study abroad programs, or off-campus study) are strongly recommended.

Minoring in Data Science

How to Declare a Data Science Minor

1. Choosing an advisor

Below is a list of faculty who advise students in the Data Science program. It is highly recommended that you contact an advisor with expertise that closely matches your primary major and/or long term career goals before matriculation into the program. If you are unsure who to talk with or have general questions, feel free to reach out to the program director.

- {Link to Advisor Page}[2]

2. Choose your Coursework

Consult the information below detailing courses required for a minor in Data Science, and choose the course of study that best matches your interests. In many cases, substitutions are available, so you can choose to discuss options with your Data Science Program advisor about possible course substitutions before submitting your selections.

[2] Interested departments can volunteer an advisor. Initial volunteer advisors would include Jaime Settle (Government), Ariel BenYishay (Economics), Dan Runfola (Applied Science), Michael Lewis (Computer Science), Dan Parker (Linguistics), Carrie Dolan (Kinesiology), Rob Rose (Center for Geospatial Analysis), Gregory Hancock (Geology).

Core Data Science Coursework
(3) DSP 140 - Introduction to Data Science (currently CS 140)
(3) DSP 146 - Reasoning Under Uncertainty (currently CS 146)
(3) DSP 201 - Data Driven Decisionmaking (currently APSC 490)

Choose one from the Below:
(3) PHIL 215 - Right and Wrong in the Contemporary World
(3) DSP 151 - Human Development and Data Science (Currently INTR 150)
(3) PHIL 303 - Ethics

Applications (Minor: Choose 2 in consultation with your advisor)
Courses chosen should engage with two or more of the pedagogic pillars of the Data Science Program (computation, application, communication and deliberation), and have a focus on the analysis of temporal, spatial, or numerical data. It is suggested that students take this requirement within their home major department when feasible.

3. Completing declaration forms
To declare your Data Science minor, you will need to complete the Declaration of Academic Minor form. On this form you will be able to indicate and plan the exact courses you will use to fulfill the requirements for a Data Science minor degree. Once the form is ready, you must have the program Director co-sign the declaration before it is turned into the Registrar.