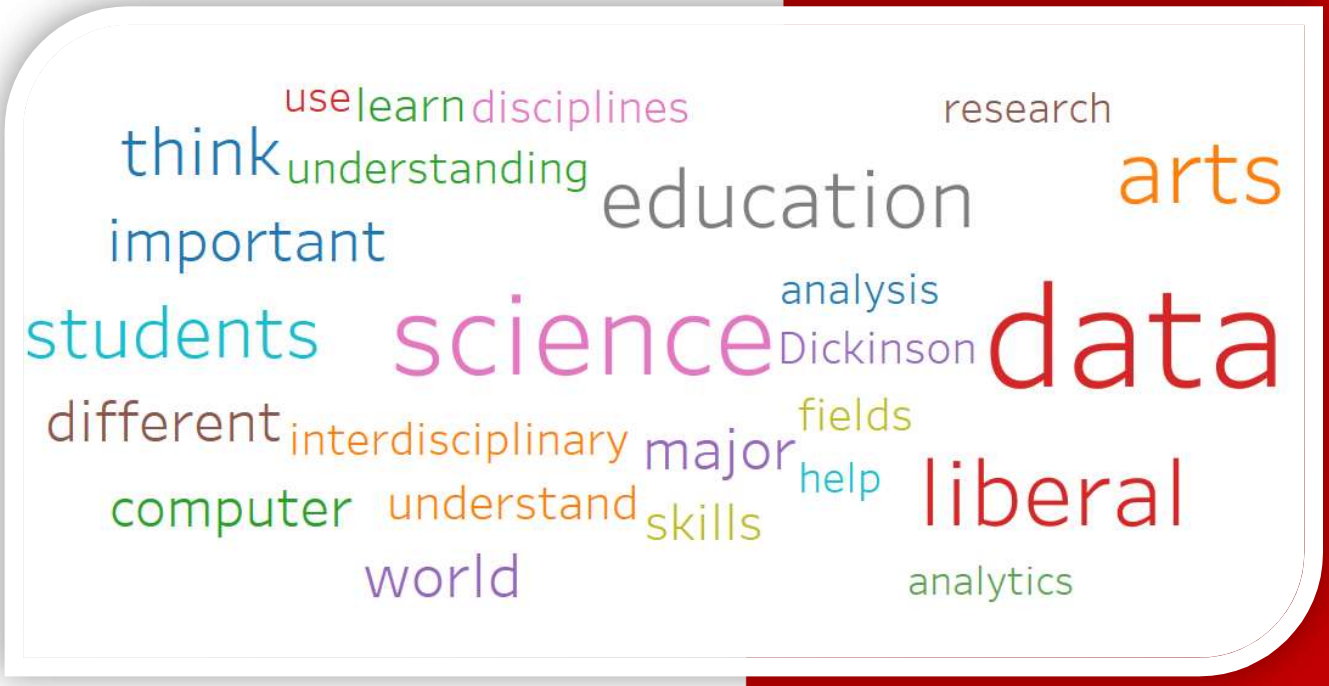
Data Science Initiative



Emily C. Marshall and David Richeson

Revolutionary Challenge Expanded Proposal

6/1/2020

The word cloud displayed on the title page was generated from survey data of Dickinson students who were asked, "What are the distinctive elements and advantages of data science in the liberal arts? Why is data science an important component of a liberal arts education?"

Summary of Proposal

Data—quantitative information, be it financial, temporal, spatial, scientific, medical, computational, operational, combinatorial, or textual—is ubiquitous in the modern world. The prevalence, quantity, and availability of data is constantly increasing in both volume and complexity and is at the center of today's greatest experiences and challenges. Academic disciplines, business models, public health plans, ideas of social justice, community programs, and governmental operations have evolved and transformed to seize the opportunities of data-driven environments.

As an innovative liberal arts institution and a leader in "useful" and "innovative"[1] education, a data science initiative will position our students to be critical consumers, users, explorers, and communicators of data. A data science initiative emphasizing crucial analysis, contextualization, interpretation, and humanization of data is inherently consistent with college's values and leverages its strengths as a liberal arts institution.

[1] Taken from the "Mission of Dickinson College" (https://www.dickinson.edu/homepage/284/about_overview).

Acknowledgements

We would like to thank our President's Panel on Innovation (PPI) liaisons Julie Johnson (Dickinson College class of '82, program director, Division on Research & Learning, National Science Foundation), Jennifer Schaefer (associate professor of mathematics), and Andrei Torriani (Dickinson College class of '89, CEO/advisor and independent, non-executive board member) for help and guidance throughout the Revolutionary Challenge process. Lindsey Davis from Level5 helped us synthesize our ideas and present the proposal to a general audience. In addition, Laura Wills (associate director of alumni relations) and Brian Falck (associate vice president for college advancement) were instrumental in building our connections and relationships with Dickinson College alumni.

Finally, we are sincerely grateful to have worked with an incredible team of enthusiastic, supportive, and innovative Data Science Initiative contributors. Working with this group was truly one of the most rewarding and exciting experiences of our Dickinson College careers.

Table of Contents

I. Project Team

The Data Science Initiative team is a collection of 45 Dickinson College faculty, staff, administrators, alumni, and students. The faculty and students represent 16 different academic departments: biology, chemistry, computer science, earth sciences, East Asian studies, economics, environmental science, environmental studies, international business and management (INBM), mathematics, philosophy, political science, psychology, sociology, Spanish, and the writing program. The members of this team are enthusiastic supporters of the Data Science Initiative and have helped move the proposal forward in various capacities. The website also contains a complete list of Data Science Initiative partners[2] The Data Science Initiative was started by a smaller working group of faculty that participated in an intensive week-long brainstorming exercise in the Summer of 2019.

Another subgroup of this Data Science Initiative team was more deeply involved in developing and writing the expanded proposal. The proposal authors are listed below in Table 1.

Table 1: Revolutionary Challenge expanded proposal authors

Proposal		Supplemental documentation	
Name	**Title**	**Name**	**Title**
Tara Dedrickson	Dickinson College Class of '21	**Jeffrey S. Forrester**	Associate Professor of Mathematics
Maggie Douglas	Assistant Professor of Environmental Science	**Richard Forrester**	Professor of Mathematics
Emily C. Marshall	Associate Professor of Economics Faculty Director QR Center	**Chauncey Maher**	Associate Professor of Philosophy
Lan Pham	Dickinson College Class of '22	**Tracy McKay**	Lecturer in Mathematics
Michael Roberts	Associate Professor of Biology	**Anthony S. Rauhut**	Professor of Psychology
Dana J. Somers	Assistant Professor of Biology	**David Richeson**	Professor of Mathematics
Anthony Underwood	Associate Professor of Economics	**Farhan Siddiqui**	Assistant Professor of Computer Science
		Michael Skalak	Lecturer/Technician in Computer Science
		Eddie Tu	Assistant Professor of Mathematics

These faculty members have extensive research experience using data. Their research expertise include: applied microeconometrics, computer science and operations research, mathematical biology and genetic networks, population and quantitative genetics and evolution, probability, statistics, and stochastic processes (specifically, Markov processes), quantitative ecology and meta-analysis, and time series econometrics. Collectively, their academic records include many publications in these fields.[3]

Jointly, the authors of this proposal have taught classes in:

[2] In addition, we have communicated with over 50 alumni supporters that have expressed interest in the Data Science Initiative.

[3] Curricula vitae (CVs) and short biographies of all proposal authors are attached in the supplementary materials (Section **Error! Reference source not found.**).

Advanced Econometrics	Integrative Environmental Science
Analysis of Algorithms	Linear Algebra
Analysis of Psychological Data	Mathematical Statistics
Artificial Intelligence	Numerical Analysis
Biostatistics	Operating Systems
Computer Science I	Operations Research
Computer Science Theory	Population Genetics
Data Structures	Precision Medicine
Discrete Mathematics	Probability and Statistics I
Econometrics	Probability and Statistics II
Elementary Statistics	Programming Languages
Fermentation Microbiology	Statistical Methods in Economics
Genomics, Proteomics & Bioinformatics	
Two different first-year seminars on using data to make decisions	

One of the proposal authors is the faculty director of the Quantitative Reasoning (QR) Center, which provides peer tutoring and faculty support in courses with quantitative content. Many of the other faculty and student authors above have been involved with the QR Center in a variety of ways. Two of the authors developed the quantitative economics major in the spring of 2018 to meet growing demand for empirical courses in the economics department. The authors have also been involved in initiatives supporting data at the college including the 2017 Big Data Clarke Forum faculty seminar and the Writing with Numbers learning community.

II. Problem/Need/Situation Description

A. Problem

If the technology platform for society can now turn over in five to seven years, but it takes ten to fifteen years to adapt to it: We will all feel out of control, because we can't adapt to the world as fast as it's changing.[4]

In today's world, we are inundated with numerical information. Data touches all aspects of our lives, and it is increasing exponentially in both volume and complexity. Data science is at the heart of many of today's greatest opportunities and most daunting challenges, including climate change, the future of work, globalization, healthcare, artificial intelligence, and critical aspects of privacy and security.[5] Data analytics transcends the traditional disciplinary boundaries of the mathematics, computer science, natural sciences, social sciences, arts, and humanities, and pervades a multitude of the postgraduate degrees and careers. A liberal arts education that prepares students for effective communication, a nuanced global perspective, and rational and ethical decision-making cannot be complete without the ability to meaningfully comprehend, critique, and construct data-driven arguments.

The problem is simple: the data-saturated world of the 21st century demands data-savvy generators,

[4] CEO of Google X, Eric Teller, from Thomas Friedman, *Thank You for Being Late: An Optimist's Guide to Thriving in the Age of Accelerations.*

[5] Throughout this proposal, we have decided on the terminology of proposing a Data Science Initiative that includes a data analytics major, understanding that data science and data analytics have different meanings.

consumers, and interpreters. These skills are essential for engaged citizens and agents of change in today's data-rich environment. The college's mission to "provide a useful, innovative and interdisciplinary education in the liberal arts and sciences to prepare students to lead rich and fulfilling lives of engaged global citizenship, working for the common good," requires that its students have the opportunity to address the rhetorical nature of conclusions based on numbers and analyze and recognize bias in data. The proposed Data Science Initiative is inherently interdisciplinary; it is a true manifestation of a liberal arts education that includes competence in technical skills and content expertise but transcends vocational learning to achieve high levels of critical thinking, analysis, understanding, and communication.

We seek to graduate Dickinson students who are prepared to adapt rapidly to the data-driven global society of today and tomorrow.

B. Need

To continue to be an innovative liberal arts institution and a leader in "useful" and "innovative"[6] education, we need to graduate adaptable students capable of meeting the challenges of an information-accelerated world. To this end, we propose a data science initiative that includes the following:

- Core courses in data analytics to provide foundational competencies
- Opportunities to explore specific topics in depth and apply learned skills
- Foundational domain knowledge
- Experiential learning
- Ethical considerations

Our proposed data analytics major will complement existing Dickinson academic programs and learning goals, while leveraging our strengths as a leading liberal arts institution. It will give students a depth and breadth of knowledge, including the technical expertise to perform sophisticated analyses. The envisioned curriculum will include core courses in data science, advanced data science topics courses, and core courses in the selected discipline where data science skills will be applied to a research question. Moreover, an increasing number of academic disciplines have added data analysis to their existing collection of methodological techniques. Addressing data-driven research questions within a discipline is best accomplished by combining data science tools and skills with discipline-specific grounding.

The impetus for the Data Science Initiative comes, in part, from the preexisting proven demand for data analytics courses on campus. Many professors in numerous disciplines are interested in teaching data-oriented courses, working with students on empirical projects, and broadening pedagogical applications; but, they also cite capacity and resource constraints. Course development and research opportunities are dependent on the contributions of existing faculty and the acquisition of additional areas of expertise through new faculty hires. A coordinated initiative will help us meet this need effectively.

Furthermore, students are interested in working with data. In the fall 2020 semester, a new course is

[6] Taken from the "Mission of Dickinson College" (https://www.dickinson.edu/homepage/284/about_overview).

being offered for the first time in the mathematics department entitled, "Introduction to Data Science." Although the course was not advertised and is not required for any major, more students signed up than there were seats available, indicating significant organic demand for the topic. In the spring of 2020, we conducted a survey of Dickinson students asking for responses on a scale of strongly disagree (1) to strongly agree (5) with the following statements:

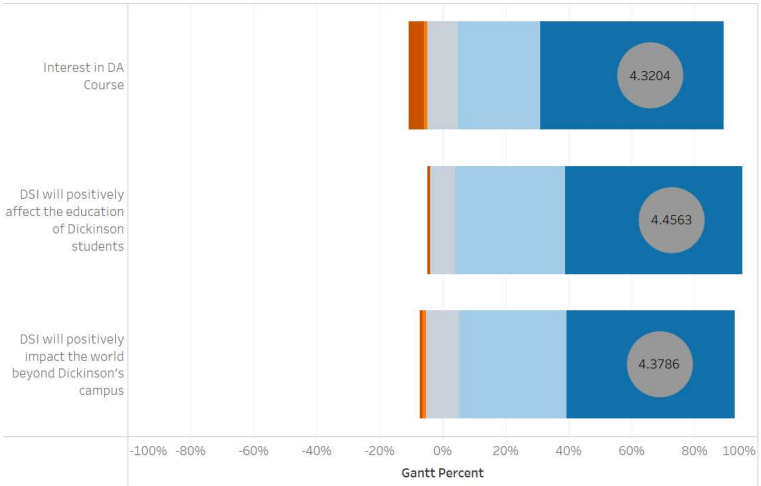
- I would be interested in taking an introductory course in data science with no prerequisites.
- The Data Science Initiative will positively affect the education of Dickinson students.
- The Data Science Initiative will positively impact the world beyond Dickinson's campus.

Figure 1a shows the responses for all students, with the mean answer in the gray circle. It is evident that many students would be interested in taking a data science course. In addition, current students feel strongly that the Data Science Initiative will positively impact the education of Dickinson students and the world beyond Dickinson. Figure 1b shows the survey responses on a scale of strongly disagree (1) to strongly agree (5) for the statement, "I would have pursued a data analytics major if it existed when I arrived at Dickinson." These responses are only reported for current computer science, mathematics, and quantitative economics (Qecn) majors.

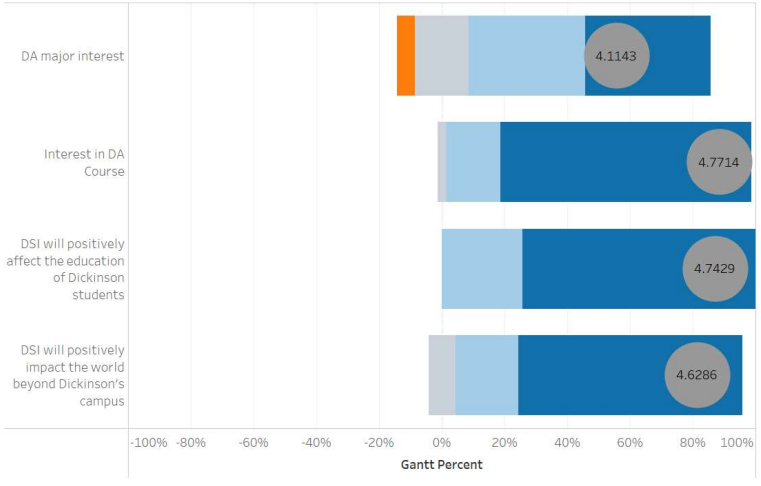
Current computer science, mathematics, and quantitative economics majors report even higher levels of interest in taking a data science course and are even more likely to believe that the Data Science Initiative will positively impact the education of Dickinson students and the world beyond Dickinson. Additionally, these students indicate a high level of interest in the data analytics major, with a mean response of 4.11 (5 indicating strong agreement) for the statement "I would have pursued a data analytics major if it existed when I arrived at Dickinson."

Figure 1: Survey responses

a. All students (n=103)



b. Computer science, math, and Qecn (n=35)

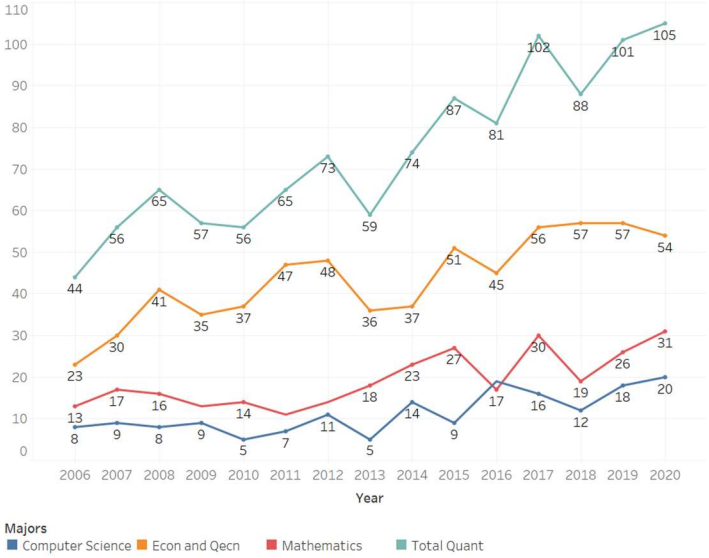


The growing desire for majors in highly quantitative and data-oriented fields is clearly evidenced by the increasing number of students majoring in computer science, mathematics, and economics or quantitative economics. Figure 2a shows the number of majors in each of these disciplines from 2006 to 2020. Over this time frame, the number of computer science, mathematics, and economics or quantitative economics majors has more than doubled (from 44 to 105). Figure 2b shows the number

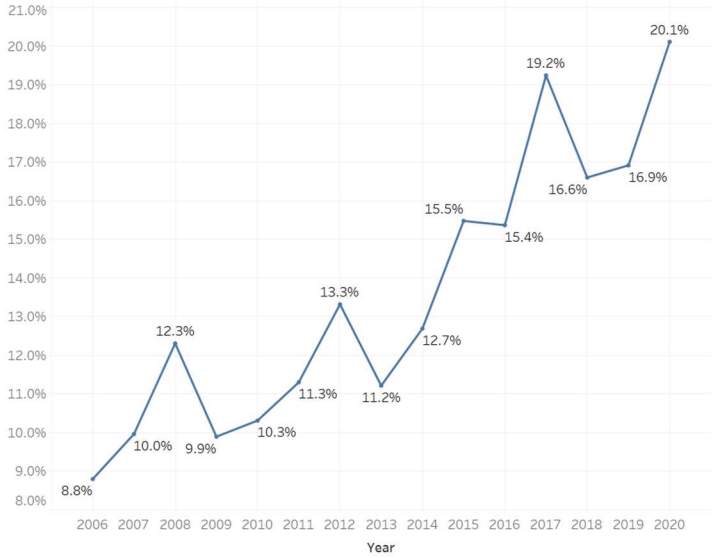
of combined computer science, mathematics, and economics or quantitative economics majors as a percentage of total graduates. In 2020, approximately 20% of the graduating class majored in computer science, mathematics, and economics or quantitative economics.

Figure 2: Computer science, mathematics, and economics or quantitative economics majors

a. Total number of majors



b. Percent of graduates



The Data Science Initiative encompasses a longer-term goal of incorporating "data across the curriculum" that will provide all students with the useful education that Dickinson College guarantees.

C. Situation Description

Current Dickinson faculty members from mathematics, computer science, and other quantitative disciplines already possess the expertise to launch the data science curriculum as it is currently envisioned. Other faculty members from across the curriculum are eager to develop and mentor discipline-specific research projects that apply data analytics skills to important problems in their fields. Because of this existing expertise, we are well-positioned to start this initiative, and it will enable a phased rollout of the program as we determine student demand and prospective student interest.

However, these same faculty members are now delivering some of the most sought-after programs at the college (see Figure 2, for instance). They are already committed to offer courses for the majors in their respective departments that are in high demand (almost always fully/over enrolled); thus, it is impossible to implement a data analytics major without additional faculty. The anticipated shift of students from existing majors into the data analytics major will enable the reassignment of some teaching responsibilities. However, several factors contribute to the eventual need to hire new faculty including the number of data analytics growing and reaching its equilibrium, a new pool of prospective students attending the college, and the need to broaden areas of faculty expertise.

An informal survey of current Dickinson students indicates significant interest in data science courses (see Figure 1). Faculty members in many departments, such as biology, chemistry, economics,

environmental science, international business and management, psychology, and sociology, report significant student interest in taking courses and pursuing independent research projects that involve data. Administrative departments such as Athletics, Admissions, the Center for Sustainability Education, and Institutional Research have also expressed interest in working with students on projects requiring data analytics. Student and faculty interest in data science at the college reflects the growing interest of prospective college students and researchers in this field. Offering a program in data science is expected to draw students and faculty to Dickinson who otherwise might look elsewhere for their education and careers, respectively. The mission of the college to provide "a useful education" is consistent with this initiative, as reports from alumni and other contacts confirm that employers in most sectors are looking for employees with data analytics training. As data science continues to impact all aspects of our lives, preparing students to competently embrace this movement will better prepare them for future careers and personal success.

The opportunities provided by this initiative include increasing the number of students interested in a Dickinson education from a declining demographic pool. Broadening the college's appeal to prospective students, while maintaining our commitment to a liberal arts education and meeting our current students' interests, will help ensure the economic and intellectual health of the institution. The opportunity for our students is clear—a useful education in data analytics will position them for careers of achievement in a variety of fields and prepare them to become engaged, active citizens. The world of data is here to stay and will continue to increase in importance over time, impacting many aspects of our personal and professional lives. Dickinson College can become a leader among undergraduate, liberal arts colleges by developing a robust data science program *now*. In the next decade, liberal arts institutions will begin developing data science programs; instead of adapting to survive, Dickinson should lead the evolution.

To accommodate the expected high demand from students and to administer this complex major, we will have to hire an associate director and two new faculty members. Professional development for affiliated faculty members, new course development, and opportunities for interdisciplinary student-faculty research all require additional resources. Creating the physical and intellectual space for interdisciplinary collaboration is expensive. The necessary talent, commitment, and conviction is here—the financial resources are the key missing component.

III. Work Plan/Specific Activities

The primary target audience is students, both current students who will benefit from the addition of data science courses within the curriculum, as well as future students who may be attracted to Dickinson because of the unique offering of in-demand data-related courses within a liberal arts framework. Additional beneficiaries of the program include faculty across all divisions who will have new opportunities for teaching and scholarship using techniques from data science and administrative offices across campus that can use students with data analysis skills to help their departments and programs make data-driven decisions. Finally, we plan to engage the wider local and Dickinson communities and provide opportunities for collaborations and mutually beneficial partnerships.

The Data Science Initiative team[7] has collaborated for over a year to research, explore, and discuss the Data Science Initiative, including the data analytics major curriculum. The central component of our

[7] See Section **Error! Reference source not found.** for CVs and biographies of the Data Science Initiative proposal authors.

proposal is a new major in data analytics that has been thoughtfully designed by a group of faculty (see Section 0) in coordination with the larger Dickinson community. Our research has consisted of reviewing the curricula of 62 undergraduate data science programs,[8] reading and discussing topical journal articles and curricular recommendations,[9] consulting other liberal arts institutions that have implemented similar programs, and speaking with alumni who work in the field.

A. Output and Initiative Activities

The Data Science Initiative has several key components, outputs, and deliverables. Our proposal includes:

1) a highly interdisciplinary **data analytics major** involving foundational courses in mathematics and computer science; a course in the philosophy and ethics of data; an experiential component such as an internship or an independent research project; and a three-course disciplinary focus in a field that allows students to meaningfully apply their skills in data analysis to a particular subject area;
2) a **data analytics course with no prerequisites** that is accessible to all Dickinson students and is specifically designed for non-majors;
3) an **expanded Quantitative Reasoning Center** that is integrated with the Data Science Initiative; and
4) a **culture of collaboration with data** across campus and the creation of regular opportunities for professional development in data science.

This Data Science Initiative will help prepare our students to become the citizen-leaders and agents of change to advance the lot of humankind, as originally envisioned by Benjamin Rush in a world increasingly populated with data.

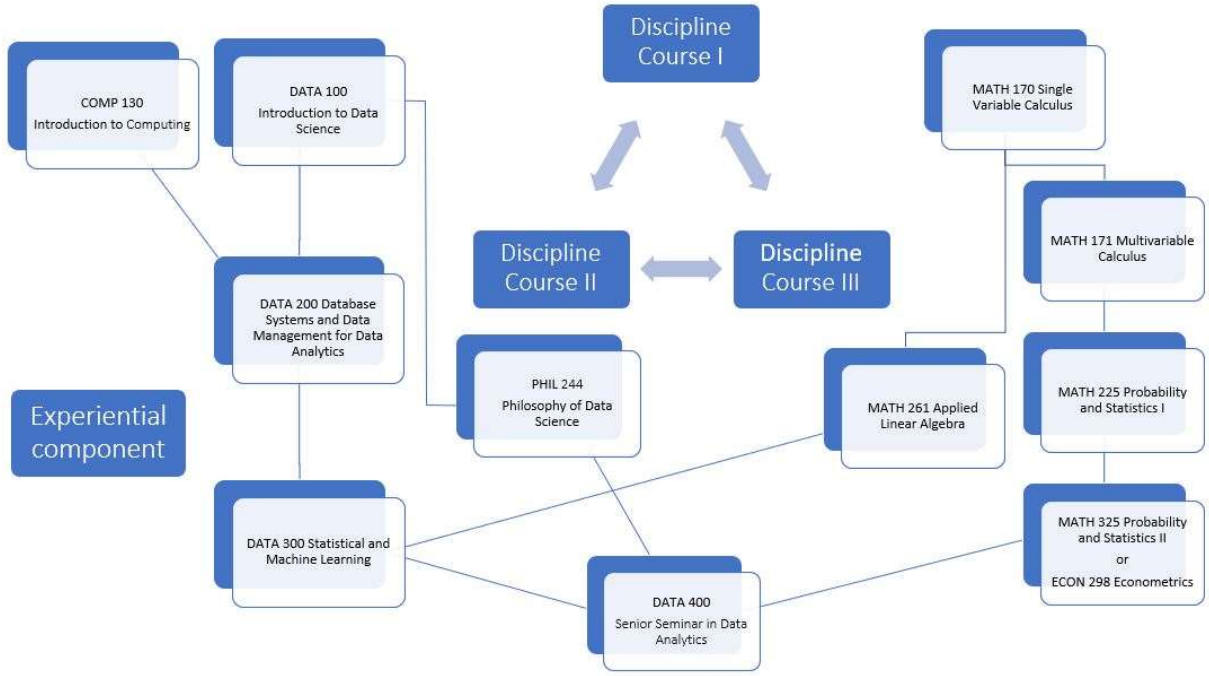
The field of data analytics evolved from existing domains including computer science, economics, and mathematics and has differentiated itself as a distinct field over the last decade. Our curriculum takes advantage of the inherently interdisciplinary nature of the subject by thoughtfully combining courses from the aforementioned disciplines and creating new data analytics courses with additional content expertise to form a coherent, rigorous, and relevant major. Figure 3 presents the courses required for the proposed data analytics major.

The major includes foundational courses in data analytics, mathematics, and computer science to build student knowledge of the tools required to perform data analysis. Several other factors differentiate the data analytics major at Dickinson from those at other higher-education institutions. First, our major requires a course in the philosophy and ethics of data. This course is intended to help students think about issues associated with data use, including privacy, security, reproducibility, limitations, and integrity. These concepts will reappear in the subsequent coursework, in particular, during the senior seminar. Second, an experiential learning opportunity (an internship, independent research, community project, or on-campus partnership) is an important part of the data analytics major. Students will be able to apply their coursework to a hands-on experience. Third, it includes a three-course disciplinary focus in a field that allows students to meaningfully apply their skills in data analysis

[8] www.discoverdatascience.org/programs/bachelors-in-data-science
[9] See, for example, Aoun (2017), Cassel and Topi (2015), De Veaux et al. (2017), Johnson (2018), Miller (2014), and Walker and Kelemen (2010).

to a particular subject area culminating in a senior seminar project. The three-course sequence in another field will provide students with a foundation to know what models are appropriate to use and what are the key assumptions for analysis in a discipline outside of data science. They will learn how to communicate the inputs and outputs of analysis across teams and skill sets and engage with subject-matter experts in problem solving with data. In addition, there will be an emphasis on written communication in the writing-in-the-discipline course and in the senior seminar.

Figure 3: Data analytics major curricular flow



The data analytics course with no prerequisites, DATA 100, will be accessible to all Dickinson students, providing them with the opportunity to explore the complexities of data wrangling, visualization, and analysis. In addition, students will gain an understanding of the limitations associated with data analysis and how to translate data-based conclusions into language meaningful to a general audience.

The QR Center provides peer tutoring and faculty support for courses with quantitative content. QR tutors assist students with quantitative reasoning across the curriculum—focusing on introductory courses that familiarize students with a field of study and a potential major. They are available to assist students in general quantitative areas and with discipline-specific gateway courses based on the tutor's area of expertise. An expanded QR Center would be fully integrated with the Data Science Initiative and support all aspects of data analytics on campus including resources for students, professional development for faculty, and project-based guidance for problem-solving with data in other departments (non-academic) at the college.

The activities described above would create a culture of collaboration with data across campus, facilitate conversations about using data, and develop a community of educators and students interested in asking and answering questions with data.