An Innovative Interdisciplinary Undergraduate Data Science Program: Pathways and Experience

Bojan Cukic
College of Computing and Informatics
University of North Carolina at Charlotte
Charlotte, NC USA
bcukic@uncc.edu

Douglas Hague
School of Data Science
University of North Carolina at Charlotte
Charlotte, NC USA
dhague@uncc.edu

Mary Lou Maher
Software and Information Systems
University of North Carolina at Charlotte
Charlotte, NC USA
m.maher@uncc.edu

***Abstract*— The paper is part of the Difference Makers Track of FIE 2020.**

Many institutions of higher learning are in the process of defining and implementing Data Science programs. The emerging field is revolutionizing scientific discovery and many industries. Broad availability of data sets and emerging analytical techniques for their processing are changing our economies and societies. While there is a broad agreement about underlying principles in this new discipline, defining data science as a pedagogically independent discipline has proven to be a challenge.

In this paper, we present the university-wide effort that led to the creation of a new undergraduate interdisciplinary Data Science program. In particular, we point to the critical role of inclusive pedagogical design and broadening participation criteria in program development. Current and projected job market data clearly indicate that data science programs need to expand traditional STEM workforce and attract students with varying backgrounds and degrees or preparation. Such a focus has led to a broad agreement across our campus regarding the new program and its curriculum.

Keywords— Curriculum innovation, data science, undergraduate programs

I. INTRODUCTION

Data Science has emerged as one of the fastest growing areas of study on college campuses [1]. It is dramatically changing business processes in many traditional domains, creating job opportunities for graduates. As our lives become increasingly dependent on data, university faculty face the challenges of creating undergraduate programs that include data science knowledge and skillsets. While all undergraduates benefit from fundamental awareness and competencies in data science, the challenge is to develop knowledge and skills needed for data collection, storage, integration, statistical inference, communication, and ethics into multiple existing domains of science [2]. In order to create a curriculum that spans across traditional department and college boundaries, we set up a plan to develop a sustainable multidisciplinary faculty structure for curriculum development and pedagogical innovation that can govern the educational program, focus on student success, and facilitate enrollment growth.

This is not an easy task. The current data science workforce is composed of individuals trained in different fields who learned data management and analytics techniques "along the way". At the graduate level, many data science programs emerged within traditional disciplines, including statistics, business, industrial engineering, computer science, health sciences, public policy, sociology, geography, etc. But the applications motivating the development of an undergraduate data science degree should also include those in humanities and social sciences, arts and architecture, even music. In addition, cross-disciplinary aspects of data science education should encompass communication skills and ethical concepts including privacy, justice, fairness, and reproducibility.

From the workforce perspective, the rapidly growing need for data science professionals has been apparent for more than a decade. According to the Burning Glass report [3], in 2015, there were more than 2.3 million data science job listings in the United States. With an annual growth of 15%, data science careers are the fastest growing occupation. High growth rates create difficulties in hiring due to chronic shortages of skilled workforce. Data science jobs remain open about 10% to 20% longer than the market average [4]. There is clearly an imbalance between talent supply and demand in the job market. The described mix of skills and needs justify the emergence of two- and four-year undergraduate data science programs, as well as traditional graduate degree programs and a range of shorter up-skilling and re-skilling educational opportunities.

This work in progress describes the environment in which the undergraduate degree program in data science was developed at our institution. We will outline the key decisions made in order to keep the process productive and, ultimately, successful.

II. The University Environment

The University of North Carolina at Charlotte is a public institution with over 24,000 undergraduate students. Data Science had been recognized as one of the key areas for university growth almost a decade ago. In 2010, close collaborations with the stakeholders in the urban community we serve resulted in the creation of the Data Science Advisory Board. The board assisted the university in the early days of identifying the most promising points of engagement with the local community and the surrounding region. An early outcome of the board's discussions included the recommendation to create professional Masters programs in the disciplines that exhibit the fastest penetration and benefits of data science and analytics. Professional Science Masters (PSM) programs are typically designed in close collaboration with future employers. They also emphasize the principle of learning from practice, which often results in a curricular requirement for mandatory internships or similar capstone experiences [5].

Given the visibility of this initiative, we needed to act fast to address the board's recommendations. Within a couple of years, the University approved two new PSM programs, one in Health Informatics (HI) (2012) and the second in Data Science and Business Analytics (DSBA) (2013). The University also announced data science as a top strategic initiative. The discussion about the governance of the new PSM programs was resolved through the creation of the Data Science Initiative (DSI) under the auspices of the Graduate School. Having the new independent entity, situated outside any of the existing colleges, in charge of managing new data science programs had interesting consequences. As intended, DSI maintained the university focus on data science research and education. Unique industry partnerships and opportunities for research collaborations have been created and benefited faculty and departments across the campus. The resulting growth of interest and opportunities led to increased recruitment of faculty with data science related expertise in many departments. However, the new degree programs evolved very differently. The HI program evolved very narrowly and traded off the generalist data science education in favor of highly specialized education within the public health domain. It grew quickly to ~75 students, but was unable to maintain stable enrollment and became somewhat smaller over time (~55 students). The DSBA program continues to grow and now has over 200 students enrolled. While these are not unexpected outcomes, they point to the difficulty of maintaining a broad set of interrelated, interdisciplinary academic programs.

By the time the University made the decision to create a new undergraduate degree program in data science in 2018, related programs were emerging on campus. For example, the business school created new undergraduate major in Business Analytics in 2017; The Department of Bioinformatics and Genomics had well-established undergraduate minor and graduate programs; The Department of Computer Science introduced a minor in Data Science in 2016, while the Department of Mathematics and Statistics planned their own Data Analytics minor. Such an emerging landscape of data science education is not unusual. These and other examples have been interpreted as indicators that various disciplines are adopting modern scientific practice and are incorporating the data science knowledge and skills in their own academic offerings.

Universities across the world are ramping up academic offerings in Data Science. There are no prescribed roadmaps. Program leaders, experts and champions reside in different parts of campus. The emerging organizational solutions vary significantly. Nevertheless, explaining the variety of Data Science programs and options to incoming students is difficult. The lack of a dedicated data science major on campus may send a message that data science is not present in undergraduate education on campus at all. Furthermore, when data science programs are administered through well-established departments, the preconceived notions of who should (or should not) enroll in their majors are likely to propagate into the future. The National Academies study [1] clearly indicates that data science programs need to expand the traditional Science, Technology, Engineering and Mathematics (STEM) workforce and attract students with varying backgrounds and degrees or preparation. We felt that this goal would be much easier to attain through the creation of the new Bachelor of Science degree in Data Science (BS DS) with careful design of its admission policies and the curriculum requirements.

III. The Process and the Champions

The process of creating the BS DS degree started in late 2018. The University Provost called the meeting with a large number of faculty and administrators involved with data science research and existing related programs on campus. The meeting established that various existing (and future) specialized programs have important role within the University. Nevertheless, the direction agreed upon in the meeting was clear: the University will create BS DS degree program. All colleges interested in having the voice in the design of the curriculum were invited to participate in the ad-hoc committee lead by the Data Science Initiative. Due to various reasons, only the faculty representatives from the College of Liberal Arts and Sciences and the College of Computing and Informatics responded with keen interest in the new undergraduate degree program. This two-college structure enabled the committee members (faculty and administrators) to bring a variety of innovative ideas and proposals about program content and pedagogical approaches specifically tailored for the data science major. In contrast to similar efforts that created specializations within existing majors or disciplines, our aim was to develop an independent new major that incorporates many possible data analytics application domain(s).

Several important aspects of committee work were critical for the success: collegiality and mutual respect of all members, understanding of the common goal, prioritizing the interests of the students over any other group, and general understanding of constraints to resources faced by most public institutions of higher education. This was the place to define the future program's curriculum, admission requirements, progression requirements and all related academic issues. This was not the place to resolve possible implications of curricular decisions for the distribution of resources within the University.

However, frequent communication with College Deans assured that the proposal being drafted was not out of the realm of feasible solutions.

As indicated earlier, one of the most important early decisions guiding program development was to attract students with varying backgrounds and degrees or preparation and in such a way to bring in students that would not normally consider a STEM education and work. The specific agreed upon principles, drawing from the findings expressed in the National Academies study [2], included:

- Designing the curriculum in a way that admitted students would be expected to meet the university-wide minimum requirements in mathematics, statistics and computing.
- Strong reliance on project topics of broadest societal interest, with a focus on those that raise awareness of social issues.
- Preference for courses that coach learning from examples that extend to life skills.
- Providing strong student support through advising, tutoring, non-traditional learning assistants, and cohort-based programs.

The intention of these principles is to provide opportunity, learning methods and support to students that may not have advanced educational offerings and pathways in their secondary education. The expectation is to enroll and graduate students that provide a better reflection of our community than is currently present in most STEM based programs.

We share here some of the admission criteria created as the result of these principles. The intent of the University is to offer admission to applicants whose credentials indicate a strong likelihood for success in the BS DS curriculum. Students considered for the program should meet general university admission standards that include, for example, grade point averages in the upper 50% of regional high-school averages and credits from a reasonable number of Math, Science, English and History/Social Studies courses. Another important concern was the ability of the students who transfer to our University from 2-year institutions to be able to complete all courses required in BS DS major in additional two years. Transfer students are more likely to come from lower income families, are more ethnically diverse and may face significant socio-economic adversities [6]. Therefore, we developed a program structure that is highly flexible and allows selection of a wide variety of elective courses, many of which may have been taken prior to the university transfer. In addition to stated accommodations, the Admissions Advisory Committee may make exceptions to the minimum criteria for applicants who are judged to have potential or talent not revealed by test scores and academic performance.

The BS in Data Science will require students to complete general education requirements and complete 55 credit hours of study within the major. There will be no pre-requisites beyond those courses required for general education. The program will not seek specialized accreditation; accordingly, barriers to degree progression that might otherwise be associated with such programs will be avoided.

IV. Data Science Pedagogy

One of the most innovative concepts in the developed curriculum is the extensive use of studio-based education. Studio teaching is an approach that minimizes the time the instructor spends delivering lectures. The features of studio classrooms include team-based, collaborative and cooperative learning, challenge-driven projects and hands-on experiences in a dynamic environment that emphasizes personalized intellectual development and self-regulated learning. Data Science education, with so many different but related areas of study, seems to be particularly well suited for studio-based instruction.

We developed the core component of the curriculum in the form of Data Science studios, a series of six credit hour (approximately 90 contact hours of instruction) courses, one in each of the four academic years. The goal of the studio courses is to engage students as active learners while solving ethically complex, real-world challenges. The challenge problems come with the publicly available data sets that exemplify societal challenges. In the initial design of Studio I, the first year course, data sets will come from sociology and urban science, exemplifying income inequality. Students will be asked to evaluate relevant literature, interpret social concepts (race, gender, education), formulate research questions and define data driven analysis. In addition, Studio I will cover introductory principles of data science (acquisition of data, testing hypotheses, interpret patterns), statistics (models for distributive and aggregate properties, charts, histograms, sampling, distribution mean), computing (basic programming skills in Python), and ethical considerations (Organization for Economic Co-operation and Development privacy principles and the relationship to the data sets in use [7]). Subsequent studio courses will further increase the sophistication of data manipulation techniques and data volume (from structured to unstructured and big data), the complexity of social research questions, and advance statistical, computational and ethical analysis methods and concepts.

The open-ended studio environment assists students in developing critical reasoning, combine many existing tools and techniques into analysis workflows, develop data cleaning and "wrangling" skills, set up problem-appropriate sound analytical methods and communicate their results effectively. This is our approach to ensuring that program graduates will receive solid data acumen skills. Regular critiques ensure that students avoid conducting flawed or incomplete analyses and ensure valid and reproducible outcomes and conclusions.

The studios will be co-taught by instructors from at least two traditional campus departments, typically one from statistics/mathematics/computing and the other from a social science discipline (sociology, urban studies, public policy, criminal justice, etc). As the studio courses were being developed, we included a collaborator from Philosophy with expertise in ethics and the disparate impacts to ensure the implications of data science are properly represented in the studios, as opposed to adding a separate course in data ethics.

In addition to the data science studios, program requirements include 18 credit hours of coursework in computing, 15 credit hours of mathematical and statistical foundations, and a selection of courses necessary to complete the university's general education requirements. Nearly 40 credit hours are left for elective coursework, allowing us to develop multiple specialization tracks and concentrations in the future. In addition, the structure of curriculum and timing of course offerings enable transfer students with an Associates degree to complete the entire set of major courses in two years.

V. The Handoff

The Data Science program will commence in Fall 2020. Upon the approval of the proposed undergraduate curriculum, the University announced the creation of the new School of Data Science. The School is a transdisciplinary unit emerging on campus by drawing on joint faculty appointments from four different colleges and many departments. Currently, faculty members are actively developing studio courses, while the administration is preparing for the admission of its first students. We are also working with Business-Higher Education Forum on the expansion of digital and data skills into a broader population of undergraduate students to provide foundational skills for the new economy [8].

Our enrollment analysis indicates significant interest from current students, incoming freshmen, and community college transfer students. We have been cautioned to carefully manage incoming student demand as it may quickly overwhelm capacity to manage. Expected rapid enrollment growth mandates significant efforts to ensure effective quality improvement loops, sustainable management of faculty resources, and the evolution of the School of Data Science as the "academic home" for the new mold of students and faculty.

The program will rely on current faculty in the College of Liberal Arts & Sciences and the College of Computing and Informatics. Some of them have had their research areas and teaching interest migrate into the field of Data Science over the last 10 years. The others are new hires who specialize in Data Science. Current interest in joint and affiliated appointments within the School of Data Science exceeds the requirements for the proposed curriculum and currently expected enrollment.

VI. Summary

In the process of program launch, we gained significant insights into championing an innovative collaborative program clearly crossing traditional domains of academic ownership and comfort. In discussions with many peer institutions, the lack of clear institutional champions and squabbles between faculty members from various disciplines appear to be the most significant barriers to effective curriculum innovation in undergraduate data science.

This paper provides a short overview of some of the processes we utilized in the development of the new bachelor's degree program in Data Science. Our university context and determined leadership incentivized program development and minimized the potential for cross-disciplinary disagreements. Participants in the development of the BS Data Science program described the experience as one of the most collegial and innovative efforts of their careers. Less than two years will have elapsed between the first university-level meeting, which launched the initiative, and the expected day when the first data science majors will walk into the classroom. Within the two years period, the second year has been entirely dedicated to pursuing approvals for the launch of the program at all levels, from departments and colleges, to the university and the university system, the accrediting agencies and government entities.

The final challenge, as we prepare this manuscript, relates to the impact of the COVID-19 pandemic. Uncertainties in enrollment, university funding and even the feasibility of organizing campus instruction in Fall 2020 seem to be causing some anxiety among the members of the implementation team. Even in this uncertainty, faculty, enrolling students, and staff are all collaborating to launch a successful program.

Acknowledgment

The authors would like to express sincere thanks to the members of the ad hoc committee charged with the development of the curriculum for the undergraduate degree in data science at the University of North Carolina at Charlotte. They initiated many of the discussions and proposed many of the solutions included in this manuscript. We also thank the staff members of the School of Data Science for the continuing support throughout the undergraduate curriculum definition process and launch of the program.

References

[1] Institite for Advanced Analytics, "Graduate Programs in Analytics and Data Science," NC State University, 2020. [Online]. Available: https://analytics.ncsu.edu/?page_id=4184 [Accessed May 14 2020].

[2] National Academies of Sciences, Engineering, and Medicine. "Data science for undergraduates: Opportunities and options," The National Academies Press, 2018. http://nap.edu/25104, DOI 10.17226/25104.

[3] J W. Markov, S. Brangaza, B. Taska, "The Quant Crunch: How the Demand for Data Science Skills is Disrupting the Job Market," Burning Glass Technologies, 2017.

[4] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big Data: The Next Frontier for Innovation, Competition, and Productivity," McKinsey & Co., New York, 2011.

[5] Professional Science Master's, PSM, 2020. [Online] https://www.professionalsciencemasters.org/, [Accessed on April 6, 2020].

[6] D'Amico, M.M., Dika, S.L., Elling, T.W. *et al.* Early Integration and Other Outcomes for Community College Transfer Students. *Res High Educ* **55,** 370–399 (2014). https://doi.org/10.1007/s11162-013-9316-5

[7] Organisation for Economic Co-operation and Development, "The OECD Privacy Framework," OECD, 2013. [Online] https://www.oecd.org/ [Accessed on May 14, 2020].

[8] Business and High Education Forum, "BHEF Releases New Report Illustrating the Foundational Skills Needed for the Digital Economy," BHEF.[Onlne] https://www.bhef.com/ [Accessed on April 6, 2020].