2023 Annual Conference & Exposition
Baltimore Convention Center, MD | June 25 - 28, 2023
The Harbor of Engineering Education for 130 Years
ASEE
Paper ID #39214

Expanding, Improving, and Completing a Multi-College Interdisciplinary B.S. Data Science Program with Concentrations

Dr. Karl D. Schubert, University of Arkansas

Dr. Karl D. Schubert is a Professor of Practice and serves as the Associate Director for the Data Science Program for the University of Arkansas.

Lee Shoultz
Shantel Romer, University of Arkansas

Expanding, Improving, and Completing a Multi-College Interdisciplinary B.S. Data Science Program with Concentrations

Abstract

This paper describes the roll-out, continuous improvement, and completion of a multi-college interdisciplinary undergraduate program in data science that culminates in a Bachelor of Science degree with its first graduates. The success of graduate-level data science programs across the nation has influenced the creation of undergraduate degrees in data science. There has been tremendous interest by employers in having a workforce that has the skills and techniques, theoretical and practical experience at the undergraduate level to meet their increasing growth and dependency on data to be successful and competitive. This paper presents the results and experiences and continuous improvement of the 36 months of a newly created interdisciplinary program that uniquely integrates business, arts and sciences, and engineering disciplines while meeting the needs of industry stakeholders. The iterative development process, the curriculum, the launch of the program, the first-year data science program, and added years' courses completing the degree, the full-year Practicum and first graduates are described as well as new concentrations developed along the way. Lessons learned are summarized to aid other universities contemplating similar programs. Finally, we discuss student internship readiness, study abroad opportunities, student post-graduation employment offers, industry partner feedback, and the future.

Introduction

The need for data scientists continues to grow as "…use cases push innovation to meet the rising demand for data-driven business outcomes." Interviewing a group of experts, Datamation identified nine data predictions for 2022 [1]:

1. Addressing growing data quality concerns.
2. Investing in remote worker cybersecurity and threat detection.
3. The growth of natural language processing.
4. Commodifying the Internet of Things for genuine business needs.
5. Leaning on AI for network monitoring.
6. Data fabrics aiding the management of unstructured data.
7. Tech workers demanding new benefits.
8. Shifting cloud security landscape.
9. Localization meets globalization in data compliance.

Interviewing another group of experts, a year later, Datamation identified six data predictions for 2023 [2]:

1. AI Boom Fuels Data Science Growth
2. Machine Learning Growth to Remain Strong
3. More MLOps in Data Science
4. Data Science in Cloud Management
5. The Rise of Bioinformatics
6. Neural Radiance Fields

These data areas cross market segments and potentially affect all companies – regardless of size. Yet, based on published reports, the demand continues to outstrip the supply of data scientists. "The demand for data science is growing fast, yet recruiters are struggling to find data scientists that can help them move their digital ambitions forward." [3] ZD Net reports that "recruiters are noting a 295% increase in the number of data science-related tasks recruiters were setting for candidates in the interview process during 2021." The same report notes that demand for Python-based skills increased by 154% over the same period.

The Bureau of Labor Statistics has estimated employment for Data Scientists and in the past year has changed the category from "Data scientists and mathematical science occupations" to "Data Scientists." In the space of one year, the estimates on each category have significantly increased (Table 1).

Quick Facts: Data Scientists [4,5]			
	2021	**2022**	**%Change**
Median Pay	$98,000	$100,910	3%
Typical Entry-Level Education	Bachelors	Bachelors	N/A
Work Experience in a Related Occupation	None	None	N/A
On-the-job Training	None	None	N/A
Number of Jobs	63,000	113,300	80%
Job Outlook Over Ten Years	31%	36%	3%
Employment Change Over Ten Years	20,000	40,500	202%

Table 1. Quick Facts on Data Science Jobs in the United States [4,5]

The Bureau of Labor Statistics [4] projected the employment change for Data Scientists to be much faster than the average and notes that "[e]mployment of data scientists is projected to grow 36 percent from 2021 to 2031, much faster than the average for all occupations."

As has been previously pointed out, "[e]mployers have recognized that data science professionals will be a critical resource to their operational excellence, as well as for the future of their innovation ecosystems." [5]. "The ability to research, organize, analyze and extract insights from raw data is a critical skill to a company's overall success." [6]

Similarly, the National Academy of Sciences, Engineering, and Medicine have published an interim report on the undergraduate perspective envisioning the data science discipline [7] and a full report on the opportunities and options for data science undergraduates. [8]: "Imagine it is now 2040. Students born in 2018 are graduating from college. It is more than 30 years since billions of autonomous sensors and devices started continuously delivering data to cloud-based databases, which record the states and activities of vehicles, buildings, customers, patients, and citizens…. Thus, it is not farfetched to expect academic institutions to envision the data-driven world of 2040 as they shape the future undergraduate experience."

In the next section, we present updated background since our previous paper that exemplifies both the recognized need, and the evolution of solutions to meet those needs. Then, because the program is in its third year of operation, we provide a discussion of our continuous improvement process, the evolution from "start-up" into "operational" mode, and the additions and changes to the program evolving it to its current state. We then make some suggestions about the future of the program.

Background and Update

The Data Science Program has changed and evolved, and grown, over the past two years. Where we were earlier anticipating growth and offering all four years of courses in sequence [7], our growth has exceeded original expectations and it was necessary to offer all four years' courses over three years due to the number of students transferring from other majors and other colleges and universities into the Program. We also established and are growing the Honors program with the objective of an Honors course in every semester of every year. We expect our first graduates in May 2023 – three years after program inception.

Two factors have been important to our success: applying continuous improvement practices and realizing that, in a sense, the program is in start-up mode (as in an entrepreneurial start-up) and therefore we need to be nimble and willing to evolve the program as we improve it and expand it.

As we have grown, we have also seen an increase in the number of students transferring from other majors within the institution and from other colleges and universities. These include students who have courses that *may* satisfy some courses in our curriculum, particularly the introductory programming courses (Python, R, Object Oriented Programming). This has motivated us to develop a course equivalency list which benefits the students and our academic advisors.

More on these topics in the next sections.

Program Continuous Improvement

Continuous improvement should be a part of every program and course and having the benefit of designing the program with a clean sheet gave us the opportunity to integrate continuous improvement from the start, beginning with our courses. To assist, we created a slide format to summarize what we learned and that we also use to discuss our new insights with our Data Science Advisory Council. For example, Figure 1, below, is an example for the course for students not ready for Calculus I. Our faculty completes them every year for their course, and we use the feedback to improve the course and the student experience.

DASC 1011 – Success in Data Science Studies

Instructor(s): Ms. Lee Shoultz, Dr. Karl Schubert

What went well:

1. Established a sense of belonging to the Data Science program
2. Introduction to case studies
3. Interview etiquette, resume building & STEM Career Fair prep most successful class taught

What did not go well:

1. Spacing of assignments and length of time to complete the assignments
2. Too many topics from University Perspectives
3. Student interaction was lacking during class

Continuous Improvement for next year:

1. We are adding a textbook *Teach Yourself how to Learn* by Saundra McGuire
2. Adding more case studies and information about the data science program + concentrations
3. Reevaluate how many things we utilize from University Perspectives

General Observations about the students:

1. Students wanted as much feedback as possible about their performance and grades
2. Time management

Figure 1. Continuous Improvement Learnings for DASC 1011 – Success in Data Science Studies

Another course technique we use is a course retrospective as a course assignment. It is not anonymous; however, the grading is quite straightforward, and we have shown the students that we proactively use the results to improve that class and as possible to make improvements to the next class in their sequence. The questions for the retrospective are:

1. What went well in the class this semester?
    a. What can we do to not "lose the formula"?

2. What did not go so well this semester?
    a. What can we do to improve it?

The grading for the course retrospective is very straightforward:

- 100% - thoughtful input
- 75%. - some thoughtful input
- 50% - any input
- 0% – no input

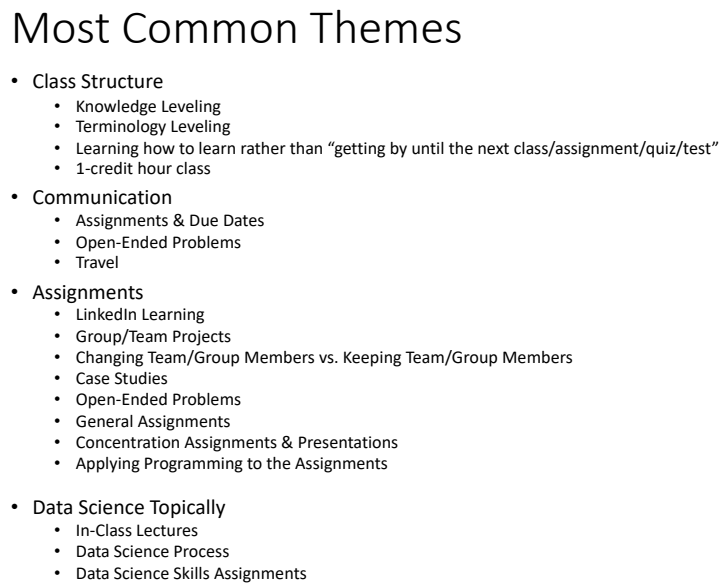


Figure 2: Summary Feedback from DASC 1001 Introduction to Data Science Retrospective

The students did not have any problem giving thoughtful, constructive feedback on both the positive and the negative. Figure 2 is an example summary of the feedback from DASC 1001 - Introduction to Data Science taught in the Fall of 2022. Each were discussed in the first day of the following Spring semester class with explanations and improvements that would be made to address them, as appropriate. In some cases, the feedback conflicted – such as the LinkedIn Learning courses on being a good team member, how to be a team leader, and project management. The remaining time of the first day of class covers the syllabus and expectations for the class.

During our Fall Faculty Retreat and our yearly Data Science Advisory Council (DSAC) meeting, the instructors for each course presented their course retrospective and the improvements they planned to make the next time they taught the course. The feedback from the DSAC was overwhelmingly positive, with one major corporation chief technology officer saying, "I have never seen this level of transparency in a degree program, and I am on other major university advisory boards."

This has also been especially interesting with our first cohort as they have been taking every first- and second-year course for the first time it has been taught and, realistically, we knew that there would be improvements to make – real-time and for the next time the course was taught. Discussing this with the students at the beginning of each course and letting them know that we would adjust as needed was helpful and appreciated. The third-year and fourth-year courses were taught in the second and third years as there were several students in the first cohort who had transferred from other majors and we accelerated introducing the courses to make a "6-semester plan" feasible for them to be able to graduate in either 4 or 5 years from when they began at the university.

At the same time as we have been applying continuous improvement, the program has been growing significantly and that has fueled evolution and expansion, which are discussed in the next section.

Evolution and Expansion of the Program

As is to be expected in any new degree program, we have evolved, expanded, and grown, and will necessarily need to evolve again as we go into our fourth year in the Fall of 2023. The projections are for 80-100 incoming first-year students, taking the overall program enrollment to 220-240 students. This influx of students has obvious ramifications for class size, instructor resources, classrooms, labs, etc. Each will be discussed in turn.

Currently, ~40% of our students are Honors students, so an obvious evolution to deal with the expanding student population is to create an Honors section of the first-year introductory courses ("Introduction to Data Science" and "Role of Data Science in Today's World) and encourage and advise the Honors students to enroll in the Honors sections. This will allow us to manage the open-ended team-based projects to a reasonable number for active learning and in-class presentations of project results and is expected to be implemented for Fall 2023. We found that this year with seventy-two students and eighteen teams there was not sufficient in-class time for all teams to present. Our solution was to have all teams record their presentations and to offer teams to volunteer to present for extra credit. While this did a least provide some in-class project report interactions, our goal is to have all teams have the opportunity to present as developing visualization and communication skills is a principal element of our degree program. Overall, our objective is to have at least one Data Science Honors course pers semester and with these additions we will have offerings in all but the second year of the four-year program.

For the introductory programming courses ("Programming Languages for Data Science" and "Object Oriented Programming for Data Science"), the instructors are familiar with having large lecture classes and then adding lab sections to keep the lab enrollment to ~25-30 per section for the best student experience.

One additional observation from our second (most previous) year teaching the first-year courses is the importance of instructor enthusiasm and qualifications. While many believe that teaching first-year courses is easy and doable by nearly anyone, such is not the case. Having an instructor in a first-year class who is unenthusiastic about the topic, the course, and the students in the course, have a negative impact on student retention – in the major at a minimum. It also reduces the actual *learning* by the students as they quickly realize that the instructor is uninterested in their class. The learning from a program point of view is to ensure that the prospective instructor(s) are qualified – they know the topic – and are *enthusiastic* about teaching the course(s) to these very enthusiastic first-year students and that the program administrators follow-up to ensure that these come through with the students. This is certainly important for all courses, but it is especially important for first-year courses as we help the students learn their major of choice and transition from high school to college. It is just as important to recruit, not assign, instructors in colleges and universities as it is to recruit, not assign, people on teams in industry.

The Peer Mentors (Peer Mentoring is discussed later in this paper) are invaluable in providing feedback from the students on how the first-year classes are going, as viewed from the student perspective.

One additional adjustment in our Program was to expand our two introductory courses, *Introduction to Data Science* and *Role of Data Science in Today's World*, from one- and two-credit hour courses, respectively, to three-credit hours each. They were originally designed that way but were reduced to fit in another course whose content has been folded into other courses. This has resulted in minor changes to the first two years of the 8-semester plan. At the same time, this change has allowed us to expand the *Introduction* and *Role* courses to the appropriate number of hours to meet the learning objectives.

With our program's growth, one positive aspect is that we have achieved the critical mass to have an RSO – a "Registered Student Organization." Students from our first and second cohort spent last Spring's semester organizing and creating the RSO and have had their first meetings this year – with a substantial number of the total student population attending. It is early in their life and the future looks particularly good for them in working together to find interesting data science topics to do together.

Our degree being a *program* rather than a degree in a *department* is a strong aspect for us and for our students. Being a collaboration of the Colleges of Engineering, Business, and Arts & Sciences, we can draw from all three colleges for our faculty and instructors. They, in turn, bring the work they do into the classroom, so the students benefit from seeing real data science applications, assignments, tools, approaches, case studies, etc., from their instructor's point of view. This does, on the other hand, create interesting challenges for course scheduling. The process involves working through scheduling courses the same as if the instructors are in the same department and then negotiating with the departments and faculty involved, some adjustments due to in-department conflicts, and creating a final schedule with which will work for all. Then, there are curriculum changes in those departments and scheduling changes and even courses that were required, core (though not data science) courses change in sequencing, number of credit hours, and instructors, requiring a "back to the start" approach and even potentially requiring catalog changes to the 8-semester plan to accommodate those changes.

One additional challenge, particularly with the combination of being a new program and with a high percentage of Honors students is the number of transfer students (from other majors) and number of students with a substantial number of AP and dual-credit hours when the join the program. For them, we have created what we call the "6-semester plan" which has (in most cases) the courses from years two, three, and four put into years two and three for these students so they can graduate in six semesters from the time they enter the program. Each student is a "special case" for everything except the core data science courses. These are prescribed in a particular way to minimize (though not eliminate) the effects of taking courses as co-reqs rather than in pre-req sequence. In some cases, this also means the students are required to do supplemental studies (such as completing DataCamp introductory programming courses over the summer) to be able to take a class requiring the use of that language while they are also in the course learning that language.

We have also expanded the number of domain concentrations from ten to twelve, adding Cybersecurity Data Analytics and Music Industry Data Analytics.

Cybersecurity Data Analytics

- ✓ Help organizations assess, detect, analyze threats, while securing & protecting data & data-driven systems
  - ✓ Master technical strategies, tools, techniques to secure data and information in the enterprise
  - ✓ Understand & apply cybersecurity, crime, tort, & privacy law to the management of data & systems
  - ✓ Understand disclosure, notification, breach, & other privacy & transparency obligations under state, federal, & international law
  - ✓ Detect & identify common malicious software and attack protocols in order to assist organizations with continuously monitoring & analyzing use patterns & network traffic to more quickly identify threats
  - ✓ Apply critical thinking to creatively & systematically solve problems & meet challenges of the ever-evolving cybersecurity threats
  - ✓ Understand the state of cybersecurity nationally & globally
  - ✓ Apply data & cybersecurity management techniques to data science.

Music Industry Data Analytics

- ✓ Understand basic history, law, and economics of the music industry
- ✓ Analyze current music industry trends
- ✓ Design promotional strategies for music and artists
- ✓ Use data to make informed music industry decisions
- ✓ Understand historical and current diversity, equity, and inclusion issues in the music industry

ABET Readiness

We also adjusted our original six outcomes to better match those expected for ABET accreditation. The figure below shows the mapping of the outcomes – which were already close enough that we have not needed to make any adjustments to our course sequencing, course design, or program design. We first adjusted the wording to match more closely the ABET wording [10] and then placed them side-by-side to verify they met the objective as shown in Figure 3.

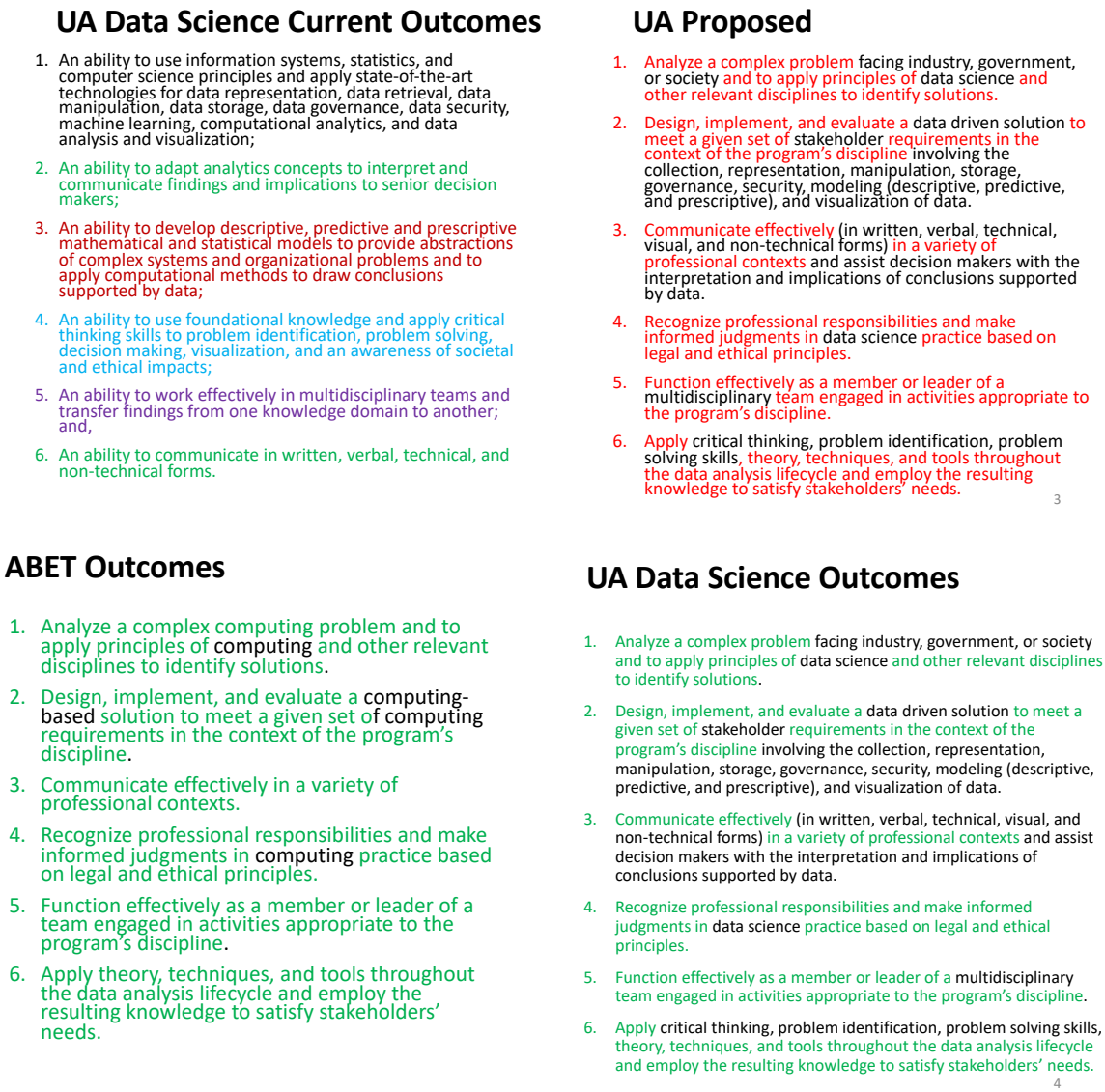


Figure 3. Adjusting Data Science Program Outcomes to Match ABET Outcomes

Our ABET-ready outcomes, then are as follows:

The Data Science major will prepare students for a successful career in data science with an amalgamation of capabilities:

1. Design, implement, and evaluate a data driven solution to meet a given set of stakeholder requirements in the context of the program's discipline involving the collection, representation, manipulation, storage, governance, security, modeling (descriptive, predictive, and prescriptive), and visualization of data.

2. Analyze a complex problem facing industry, government, or society and to apply principles of data science and other relevant disciplines to identify solutions.

3. Recognize professional responsibilities and make informed judgments in data science practice based on legal and ethical principles.

4. Apply critical thinking, problem identification, problem solving skills, theory, techniques, and tools throughout the data analysis lifecycle and employ the resulting knowledge to satisfy stakeholders' needs.

5. Function effectively as a member or leader of a multidisciplinary team engaged in activities appropriate to the program's discipline.

6. Communicate effectively (in written, verbal, technical, visual, and non-technical forms) in a variety of professional contexts and assist decision makers with the interpretation and implications of conclusions supported by data.

Course Resourcing

Because only one of our faculty/instructors is full-time and in the program administration while teaching two courses each semester, beyond scheduling the faculty are real-life and career events that require additional planning: sabbaticals, retirements, department administrative changes, departures, and more. A recent example was a faculty member who had committed to developing one of the junior-level data science courses, asserted he was on schedule, then two weeks before classes were to start (and one week before we were to review the unit and lesson plans), he resigned to take a position at another University and left before classes started. We were able to work with his former department to have our students (who needed this course to graduate this academic year on their 6-semester plan) to substitute a course. We appreciated their willingness to collaborate with us, allowing the students to continue their course. The downside was that while the course was similar (it was Cloud Computing & Big Data) in learning outcomes (to 80%) but it was not the same approach. This course used Java; ours is to use Python. This course used high-performance computing clusters for their "cloud"; ours is to use Amazon Web Services, Microsoft Azure, and Google Cloud Services. This course relied on individual assignments; our is to have a combination of individual and team-based projects including open-ended case studies. There were other differences, but these were the key ones. Nevertheless, in programs resourced as this one, allowances sometimes must be made for the unexpected especially when you are not likely to know that there are in-process actions that you would know about if it were organized as a department rather than a program.

Graduate Teaching Assistants

Another challenge is recruiting graduate teaching assistants (GTAs). The three colleges collaborating have differing views on the need for and use of GTAs and agreed when the program was developed that having GTAs for the courses would benefit everyone: students, instructors/faculty, and the GTAs, too. Recruiting GTAs is another matter as we do not yet have (though are planning to develop) an M.S. or Ph.D. in Data Science. Our approach has been to

send our requirements to all three colleges collaborating on the program. Over the past two plus years, we have realized that we must be even more specific than would appear to be necessary due to the rigorous nature of the program and the amount of math, statistics, and programming languages used. Our most recently revised requirements now include:

- Earned a bachelor's degree from an accredited academic institution
- Admitted or are currently a student at the University of Arkansas in either the College of Engineering, Walton College of Business, or the Fulbright College of Arts & Sciences
- Must be in good academic & judicial standing with the University of Arkansas
- Have experience in studying high level math such as Calculus I, Calculus II, Calculus III, Linear Algebra, Discrete Math, Differential Equations, 3000-level or higher statistics course or equivalent
- Must have programming experience in Python and/or R, Tableau and/or Power BI

We call those GTAs who qualify for our Statistical Learning and/or Optimization courses "unicorns" as the requirements are so high. So far, we have been fortunate enough to find unicorns in Astronomical Sciences (Ph.D. student), Mathematics and Statistics (Ph.D. student), and Industrial Engineering (dual M.S. student). This will be more straightforward once we have our own graduate program but even then, there are advantages to having GTAs from diverse backgrounds and studying in different program similar to our instructors and faculty.

Governance

Our original organizational team was a combination of interested and college-represented faculty and faculty who would teach in the program. As we have evolved and expanded, it made sense to move from "start-up" to "expansion and growth" (as noted earlier) and it was clear that we needed to add governance to our structure and increased focus on course evolution. To address these, we created the DASC Undergraduate Studies Committee and the DASC Technical / Curriculum Advisory Workgroup. These are shown with our other committees in the figure, Figure 4, below.

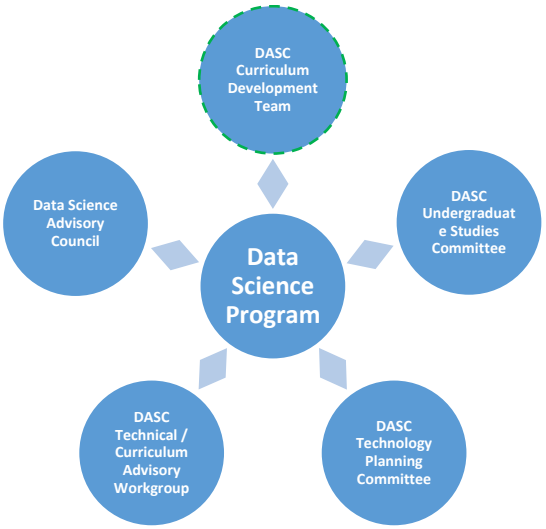


Figure 4. Data Science Program Committees

The Data Science Program is governed by the faculty teaching in the program and additional members who serve in an advisory capacity, including the original curriculum team. The Undergraduate Studies Committee has a member from each of the collaborating colleges, a Committee Chair from one of the collaborating colleges (separate from the other members) and the Data Science Program Associate Director who also serves as the Committee Secretary. All members are voting members and they serve as the primary academic change, improvement, and governance recommending body. Items brought before the Undergraduate Studies Committee are reviewed and voted to be rejected, changed, or recommended to be approved by the Data Science Faculty. Following the first 18 months of operation, the Data Science Faculty voted to have the recommendations from the Undergraduate Studies Committee be placed on the consent agenda for the Data Science Faculty meetings, allowing for discussion in the meeting if anyone desired, but history has shown that that committee was appropriately representing the faculty.

The DASC Technical / Curriculum Advisory Workgroup was created to allow for continual involvement with industry practitioners and our course instructors/faculty. This helps with continual improvement, bringing industry case studies and guest lecturers into the Program, and helps guide us in staying current (and hopefully to some degree ahead) of industry needs. One or more industry practitioners are matched with each course in the curriculum that matches their expertise and interests. This approach has the potential to develop internship opportunities for our students even more directly. This Workgroup was instantiated two years ago and is developing more slowly than originally planned but that has been because it was a bit ahead of our developing and offering the courses. Over the next year, we expect its activities to increase.

As we have grown, so has our need for team study areas and gathering points for our students. Our Fulbright College of Arts & Sciences had space being vacated in a new-ish building in an excellent location on campus and made that space available to us. We designed it with our students in mind: three team rooms, six-person capacity each, with video conferencing capabilities and wirelessly-connectable large screen display; five more team areas, six-person capacity each, with wirelessly-connectable large screen display; a lounge area where they can relax and visit, as shown in Figure 5, below. The tables and chairs are all movable so the open space can be rearranged to have larger group meetings. In fact, it is where our RSO has their meetings. We also have an administrative area that has three administrative offices, a printer room, and a conference room. The administrative offices have wirelessly-connectable large screen displays, one has video conferencing capability, and all have meeting tables for student advising and small meetings.

Figure 5. Data Science Student Team Rooms, Team Spaces, and Lounge Area

Student Internship Readiness

Our program was designed using high impact practices, such as open-ended problems, case studies, individual and team assignments, real-world/real-data problems, authoring reports, preparing, and delivering presentations, and partnerships with industry. We start these in the first year, first semester, and continue them throughout their years of study. While they are not thrilled about these, they find that putting these experiences on their resumes has resulted in even rising sophomores receiving offers for data science internships. Returning from their internships, they say: "The first thing they did was put me on a team and assign me an open-ended problem. Then, they had me write something up about it and create and give a presentation on it to our management team." Our employers report remarkable success with our interns, and many go back for a second internship or continue a part-time internship through fall and spring and into the next summer. We have been fortunate in that nearly everyone in the program who has wanted an internship has received one or more offers; all industry internships have been paid and most had a housing allowance and reimbursed moving expenses. The internships have been with consultancies, government agencies (federal and state), protein companies, banks, consumer packaged goods companies, credit card processors, and internet search companies. The locations have included many within Arkansas (we have major corporations in the state), Washington, D.C., Omaha, NE, Wilmington MA, Seattle, WA, Kansas City, MO, Plano, TX, Houston, TX, and Louisville, CO, to name a few.

Study Abroad

As our student population has grown, we have had a significant increase in interest by students to study abroad in a way that continues to contribute to their degree progress. As B.S. Data Science programs are relatively new, especially ones that are not mostly re-branded computer science degrees, research was necessary to identify a program that would match ours well. We found one and it was one – or to be exact, it found us – that is already a study abroad partner for our University. After completing a course and 8-semester plan equivalency evaluation (more on that, in the next section), we concluded that this could be an exceptionally good match for us and our students. We made a site visit to meet with the faculty and to review each other's programs and courses in more detail and developed a proposal that is in its final stages of evaluation (as of this writing). We believe that it is possible that the program could be available for AY23-24. Student feedback has been overwhelmingly positive and hopeful. We also have a University of Arkansas Center in Rome, Italy that invited us to stop by on our overseas trip and as a result, we began development of a study abroad program that would also line-up with our 8-semester plan and could potentially be available AY24-25. Both would provide an excellent broadening experience for our students while allowing them to advance in their degree plans.

Course Equivalencies

As the demand for Data Scientists continues to grow, we have been helping other institutions across (the state) to develop their data science programs. This initiative is part of the broader efforts for a state-wide data science academic ecosystem that will allow students who are unable to attend four-year colleges, to have access to data science certificates and associate degrees at 2-year colleges within the state who have opted-in to the program. The institutions we are working

with are a part of our NSF EPSCoR Education Theme 2+2 program. [11] This program's vision is to create a "start anywhere, finish anywhere ecosystem" that enables potential students to start at 2-year colleges in their area and have the ability to transfer to any of the four-year institutions in the state who have opted-in to the program to complete a 4-year Bachelors of Science degree in Data Science. This will be achieved through providing course equivalency cross-checks between established courses within the University of Arkansas' Data Science Program and the other institutions' courses.

The course equivalency program uses courses established at the university and compares them to those that have been designed and developed at other institutions. For example, to be approved as equivalent, there needs to be at least an 80% similarity of course syllabi, learning objectives, and lesson plans. The cross-checks are first done in an Excel spreadsheet by the Graduate Research Assistant and the Undergraduate Research Assistants in the Data Science program who have taken the courses. This allows students to give input on if the major and most beneficial topics are integrated into the collaborating institution's syllabi. After this, the spreadsheet is reviewed with the Associate Director of the Data Science Program. Once they certify the course equivalencies, it is then submitted an Associate Dean in the College of Engineering, and subsequently reviewed for approval by the Provost and Legal to complete approval of a Memo of Agreement (MOA) authorizing the program and describing the mechanics of operation.

Below are the 2-year plans of the University and NorthArk College. NorthArk's 2-year plan will result in an associate degree in data science.

In the 2-year plan, students at both institutions are advised to be enrolled in a closely-matched and sequenced series of courses. The programs are similar with minor differences. Students at the University will have covered Economics, Multivariable Math, and Probability and Statistics, whereas, at NorthArk, students will have completed more of the required electives and humanities. Nonetheless, students who will graduate from NorthArk will be able to transfer to the University to complete a bachelor's degree in data science within the 4-year period.

University of Arkansas			**NorthArk College**	
Year 1: Semester 1 (14 credits)			**Year 1: Semester 1 (16 credits)**	
MATH 2254	Calculus I		**ART or DRAM or ENGL or FA or FL or MUS or PHIL**	Elective (3 credit hours)
ENGL 1013	Composition I		**ENGL 1013**	English Composition I
DASC 1003	Intro to Data Science		**DVSC 1003**	Introduction to Data Science
DASC 1104	Programming Languages for Data Science (Python and R)		**(CIS 1603 & CIS 2011) OR DVSC 1104**	Introduction to Programming & R for Data Science OR Programming Languages for Data Science
			ANTH or ECON or GEOG of HIST or PLSC or PSYC	Elective (3 credit hours)

University of Arkansas			NorthArk College	
Year 1: Semester 2 (17 credits)			**Year 1: Semester 2 (16 credits)**	
MATH 2564	Calculus II		**MAT 2204**	Analytical Geometry and Calculus I
ECON 2143	Gen ED, Basic Economics: Theory and Practice		**ANTH or ECON or GEOG of HIST or PLSC or PSYC or SOC**	Elective (3 credit hours)
ENGL 1033	Gen ED, Technical Composition II		**ENGL 1023**	English Composition II
DASC 1204	Intro to Object Oriented Programming for DASC (JAVA)		**ART or DRAM or ENGL or FA or FL or MUS or PHIL**	Elective (3 credit hours)
DASC 1223	Role of Data Science in Today's World		**DVSC 1013**	Intermediate Data Science

University of Arkansas			NorthArk College	
Year 2: Semester 1 (16 credits)			**Year 2: Semester 1 (17 credits)**	
DASC 2594	Multivariable Math for Data Scientists		**MAT 2304**	Analytical Geometry and Calculus II
INEG 2323	Probability & Stochastics Processes for Industrial Engineers		**BIOL or CHEM or GEOL or PHSC or PSYC or SOC**	Elective (4 credit hours)
DASC 2213	Data Visualization & Communication		**DVSC 2213**	Data Visualization & Communication
DASC 2113	Principles & Techniques of Data Science		**DVSC 2113**	Principles & Techniques of Data Science
GNED NNN3	Gen Ed, History or Government		**HIST 2003 or HIS 2013 or PLSC 2003**	U.S. History I or U.S. History II or American National Government

University of Arkansas			NorthArk College	
Year 2: Semester 2 (17 credits)			**Year 2: Semester 2 (13 credits)**	
SEVI 2053	Business Foundations		**CIS 2203**	Data Structures & Algorithms
INEG 2314	Statistics for Industrial Engineers		**SPCH 2303 or Social Sciences or Humanities Elective**	Public Speaking or Social Sciences or Humanities Elective
DASC 2203	Data Management & Data Base		**DVSC 2203**	Data Management & Database
GNED NNN4	Gen Ed, Science		**BIOL or CHEM or GEOL or PHSC or PSYC or SOC**	Elective (4 credit hours)
RRRR NNN3	[Required Concentration Course]			
Total	64 credits		**Total**	62 credits

Table 2. Data Science 8-semester Plan 2+2 Plan for Course Equivalency and Sequencing

Current Findings and The Future

In this section, we present the results of the first two and a half years of operation of the program. This includes results of recruiting students, students and internships, students advancing to the second year, how our handling of orientation and advising have evolved, and where we are headed next.

Recruiting

The process of recruiting at the University of Arkansas is conducted through both central administration and individual colleges on campus. Even though the data science program is multidisciplinary, and sits within multiple colleges, the program's recruiting is administered through the College of Engineering. While this approach has been beneficial during the start-up phase of our program, it can be perceived as problematic as we continue to move our program further. The reason why this approach can be viewed as problematic is that students who may have an interest in data science, they may initially be recruited into a business or arts & sciences program and be unaware of the Data Science program as a result whereas a student recruited by the College of Engineering will be more likely to be made aware of the Data Science Program (or may already be aware).

Our long-term goal is to bring our recruiting in-house by hiring a recruiter for the program specifically. By hiring our own recruiting team, we will be better suited to recruit a wider audience of students who may have an interest in data science but are not necessarily drawn to a STEM based degree program.

In the meantime, our plan is to execute a targeted approach to the high school administrators to which our current students attended. Our plan is to mail a box that contains both promotional items and information about our degree program to not only thank them for sending us their best students, but to also raise awareness of our degree program. Most of the high schools in which we anticipate targeting, are in the local quad-state area.

Cohorts

Our revised estimate for AY2022-2023 was approximately 160 for total population for the cohort and it ended-up being approximately 150. Interestingly, our *original* estimate when developing the program was approximately seventy-five students.

The demographic information for cohorts 1, 2, and 3 are shown in Tables 3,4, and 5, below. In May of 2023, the program will have its first group of data science graduates. These three incredible students will represent a significant milestone and achievement, not only for themselves personally, but also for being the first students to graduate from this groundbreaking program. They have received, and accepted, job offers and some of the next graduating classes have already received employment offers for after they graduate in May 2025.

While it is too early to project actual numbers for the incoming AY2023-2024 cohort, early indications suggest that our next cohort will be approximately one hundred new first-year

students which will lead to an overall total student population of around 240. Although we have not done a significant amount of recruiting up to this point, we would like to focus our future recruiting efforts on attracting more women to our program.

Fall 2020 "11th Day" Enrollment			Spring 2021 "11th Day" Enrollment	
DTSCFR = 23	74% M, 25% F		DTSCFR = 21	72% M, 38% F
DTSCBS = 9	66.6% M, 33.3% F		DTSCBS = 14	79%M, 21% F
Total = 32	72% M, 28% F		Total = 35	74% M, 26% F

Table 3. Eleventh Day Enrollment Demographics for Fall 2020 and Spring 2021. DTSCFR are first-year students and DTSCBS are second year and above students.

Fall 2021 "11th Day" Enrollment			Spring 2022 "11th Day" Enrollment	
DTSCFR = 46	63% M, 37% F		DTSCFR = 54	61% M, 39% F
DTSCBS = 38	71% M, 29% F		DTSCBS = 30	63%M, 37% F
Total = 84	66.6% M, 33.3% F		Total = 84	62% M, 38% F

Table 4. Eleventh Day Enrollment Demographics for Fall 2021 and Spring 2022. DTSCFR are first-year students and DTSCBS are second year and above students.

Fall 2022 "11th Day" Enrollment			Spring 2023 "11th Day" Enrollment	
DTSCFR = 91	61% M, 39% F		DTSCFR = 93	60% M, 40% F
DTSCBS = 57	70% M, 30% F		DTSCBS = 51	64%M, 36% F
Total = 148	65% M, 35% F		Total = 143	62% M, 38% F

Table 5. Eleventh Day Enrollment Demographics for Fall 2022 and Spring 2023. DTSCFR are first-year students and DTSCBS are second year and above students.

Orientation

Summer 2021 brought much uncertainty as the pandemic was still in full swing. Therefore, campus took the hybrid approach in offering both virtual and in-person orientation. Regardless of your choice in how to participate, students met with their advisor virtually for approximately 30 minutes the week before or for in-person on the day of their orientation session. Virtual students were additionally provided with an online series of videos which described the information that would be discussed during in-person sessions such as living and dining, getting involved, and how to register for classes just to name a few.

Our main objective was to ensure that each student, regardless of their preferred attendance method, received as much of the same attention as possible. Therefore, we gave each student a welcome package which included information about our program such as the curriculum and laptop requirements, and a few promotional goodies to make the students feel a part of our program. If a student participated in virtual orientation, we sent the packages to their home in advance of their session.

During the in-person sessions of orientation, our information sessions covered topics such as the 8-semester curriculum plan, math pre-requisites, how students could incorporate internships in

their degree plan, our laptop requirement, and the Data Science Honors program. Additionally, in-person attendees received the benefit of being able to ask questions in real time. While orientation sessions went smoothly regardless of the method, the number one benefit for the student to attend in-person was the ability to ask questions in the moment.

Summer 2022, the University of Arkansas returned to exclusively in-person overnight stay orientation sessions. Each student spent 2-3 hours with our academic program orientation session, which included the same presentation as mentioned above with in-person advising and registration for classes. We continued to provide a welcome package to all students that attended orientation. During Summer 2022, we had the fortune of having a current student work for us part-time which allowed us the opportunity to expose incoming students to an existing data science student and so they could learn more about what they can expect from the program and faculty. This led to an increase in participant engagement during our orientation sessions and helped establish a sense of community for both incoming students and parents. It is our intention to continue this approach during future orientation sessions.

Moving forward, it is our goal to continue to make our program as transparent as possible during summer orientation sessions. We plan to accomplish this by continuing to emphasize math placement options, the rigorous nature of our program curriculum, and job placement opportunities.

Advising

Academic advising is a process in which we have struggled to balance throughout the tenure of the program. During Fall 2021, we saw a need to streamline our advising process as our student population continued to grow. Therefore, we created an advising worksheet for each of the twelve concentrations and a generic worksheet for students who have not yet selected a concentration.

Prior to an academic advising appointment, the respective advisor will update the worksheet based on the courses the student is currently enrolled in and/or their course history. Then the advisor will fill in the suggested courses for the upcoming semester, either before or during the advising appointment, and discuss the recommended courses with the student for the upcoming semester. The advisor will update the electronic version in real time and give the student the paper copy for them to take home and register for the courses discussed. Additionally at the University of Arkansas, we add the academic advising notes to that student's success profile which can be used as a reference for the student and for other administrators across campus. While this process is not fool proof, as students are always subject to not taking their advisors advice, we feel as though it is a more productive way of making notes as we return to all in-person advising.

Success in Data Science Studies

During Summer 2021, we came to the realization that we would have a sizable number of students not Calculus I ready, and therefore unable to enter the Data Science course sequence.

The question became “how do we make these students feel like they are active participants in the Data Science program if they are not able to take the first set of introductory courses?”. Our solution for this question was to create a course, much like University Perspectives (an introductory course that first-year students are required to take through other colleges on campus), called *Success in Data Science Studies*. The course, modeled conceptually after the First-Year Engineering Course *Success in Engineering Studies,* covers a variety of topics, and included guest speakers from both the university and industry. Examples of the topics covered in the course include:

- Campus Resources
- Excel 101
- Interview Etiquette & Resume Building
- Data Science Case Studies
- Core Concepts of Data Science
- How to Work in a Team Environment

Secondarily, guest speakers who visited our class taught on a variety of subject matters including:

- Academic Coaching
- How to Write in Academia
- University of Arkansas Library Resources
- History of the University of Arkansas
- History of the University of Arkansas Alma Mater
- Data Science Career Development
- Diversity & Inclusion

The assignments throughout the course were both individual and team projects. This helps prepare them for the introductory Data Science courses which also include both individual and team-based projects. At the end of the semester, we had the students submit a retrospective on the course, asking them the following questions:

1. What do you think went well this semester?
2. What was your favorite of the topics we covered?
3. What do you think did not go well this semester?
4. What ideas do you have for improving the course for next year?

While most of the students had positive feedback about the course, we used some of their constructive criticism to adjust for Fall 2022, including but not limited to: incorporating more data science topics such as case studies, learning about the concentrations, and an introduction to either R or Python programming. Additionally, with the continuing noticeable gap in college level academic preparation, we incorporated the book *Teach Yourself How to Learn* by Saundra McGuire for our students to have better guidance on how to study and prepare for classes. Adding this text proved to be an asset as many students expressed that they were unsure of how

to prepare for class and study for exams. Furthermore, during Fall 2022 students expressed that they genuinely enjoyed the introduction we assigned the class in the form of a DataCamp tutorial for the programming languages. Therefore, we are working to add additional tutorials, guest lectures, and assignments for the upcoming academic term.

First-Year Data Science Program (FYDS)

The First-Year Data Science (FYDS) program is designed to help first-year students build a solid foundation for their data science education. All students take a common set of data science foundation courses. These classes will help students understand the foundations of data science and guide them to decide on a concentration at the end of their first academic year. First-Year Data Science is more than just a set of classes. It also provides:

- Orientation
- Peer Mentoring
- Academic Advising
- Career Advising
- Academic Skills Development

What is Peer Mentoring? Peer Mentoring is a form of mentorship that takes place between a person who has lived through a specific experience and a person who is new to that experience. In the case of Peer Mentoring in the FYDS program, a student's Peer Mentor will be an upper-level, undergraduate Data Science student who will have experienced many of the challenges, obstacles, and experiences that await the first-year student. A student's Peer Mentor is a resource for the student in the realms of FYDS, the University of Arkansas, college-level academics, and the overall transition to college life. Peer Mentor meetings are 30 minutes once a week and consist of specific topics, typically ones that have proven to be useful for most first-year students. These topics are geared towards helping students better transition not only academically to a university setting, but also professionally and socially. We encourage students to take charge of the conversations with their mentor. There are many things the students can discuss with their mentor, but it is up to the student to make the most of their peer mentorship experience.

Peer Mentoring is a required component of the first-year data science courses. As such, the weekly meetings with the mentors count as a grade. Failure to meet with the mentor will result in the loss of points for that week. We place a lot of value in the Peer Mentoring program, as such we expect students to treat this opportunity with sincerity and respect. It may be a mandatory process, but it is one that we have seen work within other colleges, when utilized effectively by students.

Academic year 2021-2022 provided us with lessons with regards to the hiring of our peer mentors. First, because our inaugural cohort of students was small, we had a limited pool of people that we could hire. We only had four applicants and we hired all four of those students. However, throughout the summer our incoming cohort grew, and we saw a need to hire a fifth peer mentor at the last minute. With the hiring of our fifth mentor, it meant that each peer mentor

would have 10-13 mentees as opposed to 15-17 mentees each. Even with the reduction in mentees, the peer mentors have shared with us that they would like for us to consider having even fewer mentees per mentor in the future.

After the peer mentors are hired, they participate in an all-day training during the week leading up to the first day of classes. The training is focused on campus resources, role playing for situations that may arise when talking with students and going over the weekly discussion topics for the semester. By going through this process, it helps all mentors be prepared for the uncomfortable conversations that may come up with their mentees such as relationship hardships, roommate problems, and signs of depression.

Looking towards the next academic year, our goal is to hire enough mentors so that each student has about ten mentees to supervise throughout the semester. This will not only help the mentors give adequate attention to each mentee it will also allow them to better manage their time with the other commitments they may have.

FYDS-Decision Day

Each Spring semester, our students participate in an event we call "Decision Day." Decision Day is the event in which students officially declare their concentration of choice. We do our best to ensure the event is a fun day for the students. It is our hope that as our program continues to progress, that students will come to anticipate and get excited about Decision Day. To make the event entertaining, we provide noisemakers to the students to utilize during the festivities, and after everyone makes their declaration to their classmates and our faculty, we provide lunch and give them a Data Science t-shirt. Overall, we want Decision Day to be a memorable and fun experience for each of our students as they continue their path towards earning their data science degree. The results of Decision Day for our first two cohorts are shown in Figures 6 and 7, below.

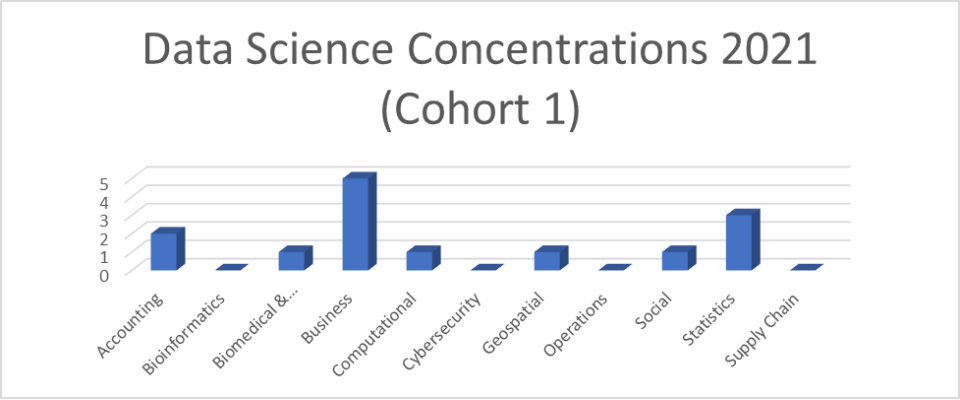


Figure 6. Spring 2021 Data Science decision day concentration selections.

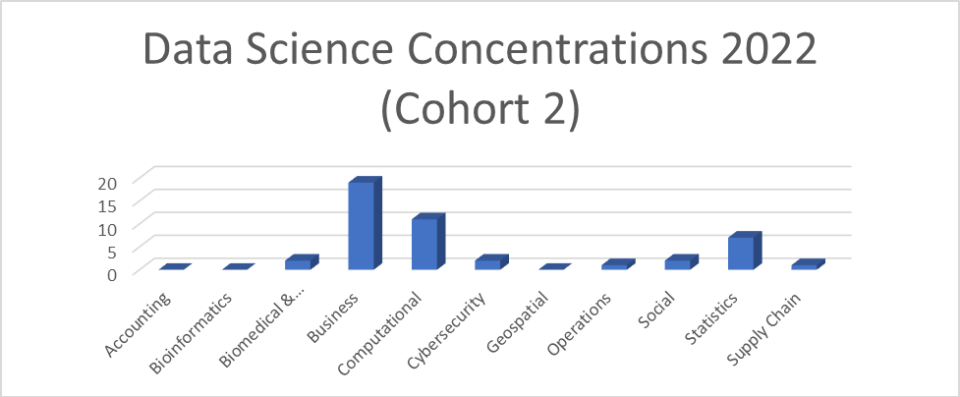


Figure 7. Spring 2022 Data Science decision day concentration selections.

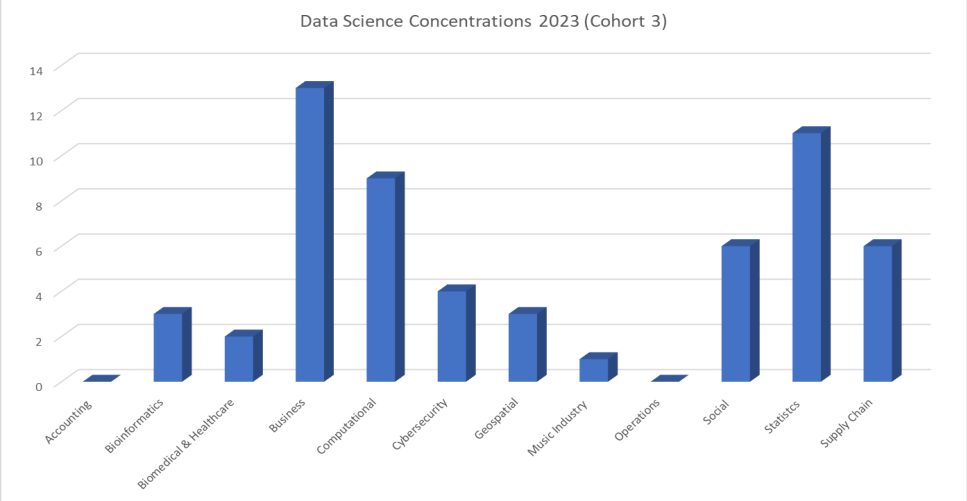
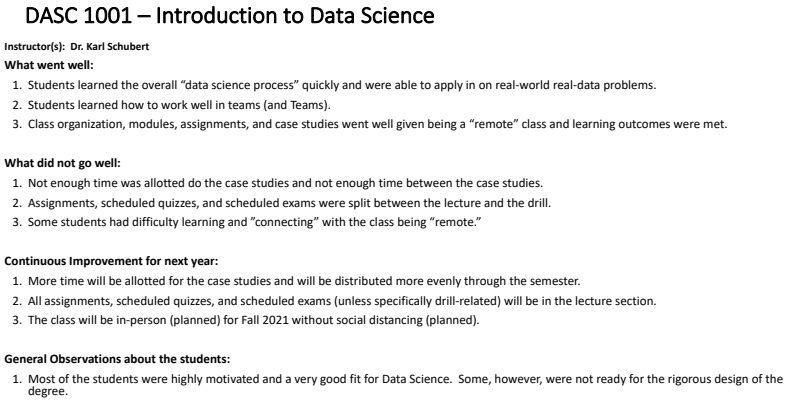


Figure 8. Spring 2023 Data Science decision day concentration selections.

Decision Day 2023 had a wider variety of concentrations being selected and we expect that to continue as our incoming classes grow.

The Future

After the nearly third academic year, we have confirmed and further learned that much of what we designed works well and with iterative / continuous improvement can evolve to improve it even further. The use of the Agile retrospective approach has worked well, and our faculty have embraced it which fits well with the ABET continuous improvement expectations. For example, each year, we combine the feedback from the University's student course evaluation surveys (anonymous), student class surveys (not anonymous but extra credit is provided generally), and the instructor's and teaching assistant's observations to develop an iterative improvement plan for the next offering of the course. An example is shown in Figure 8, below.



DASC 1001 – Introduction to Data Science

Instructor(s): Dr. Karl Schubert

What went well:

1. Students learned the overall "data science process" quickly and were able to apply in on real-world real-data problems.
2. Students learned how to work well in teams (and Teams).
3. Class organization, modules, assignments, and case studies went well given being a "remote" class and learning outcomes were met.

What did not go well:

1. Not enough time was allotted do the case studies and not enough time between the case studies.
2. Assignments, scheduled quizzes, and scheduled exams were split between the lecture and the drill.
3. Some students had difficulty learning and "connecting" with the class being "remote."

Continuous Improvement for next year:

1. More time will be allotted for the case studies and will be distributed more evenly through the semester.
2. All assignments, scheduled quizzes, and scheduled exams (unless specifically drill-related) will be in the lecture section.
3. The class will be in-person (planned) for Fall 2021 without social distancing (planned).

General Observations about the students:

1. Most of the students were highly motivated and a very good fit for Data Science. Some, however, were not ready for the rigorous design of the degree.

Figure 8. Iterative/Continuous Improvement Retrospective for DASC 1001

This process is based on an iterative/continuous improvement process the lead author had been using for decades in industry for tech industry product development. It is straightforward, easy to explain, and easy to deploy. The retrospective for DASC 1001 was for the first time the class was taught: Fall 2020. This academic year, it was taught for the third time and is significantly improved. Yet, the expectation that there will be more to improve on as is nearly always the case. We have established a culture for this program that such improvements are a true positive – not a negative reaction. That has helped us improve every course in a timely manner.

With this academic year, we will have taught all our classes (excepting the Cloud Computing & Big Data class) and have benefitted from the first-time-offered classes having been smaller as we evolve the classes through teaching them and with the classes growing each year. That said, the challenge now at hand in the first-year courses, soon to be followed by the second-year courses, is that we have reach optimal course and lab sizes and now need to go to multiple sections, increasing the faculty load, each year for at least the next three years.

The timing for iterative improvement, evolution, and expansion nicely lines-up with our plans for ABET accreditation and preparation. This academic year, we are focusing on completing the ABET-style syllabi and identifying the data to be collected for assessments and continuous improvements, the latter of which are already being collected. The remaining timeline and activities are shown in Figure 9, below, culminating in an expected first ABET review in Fall 2026 in concert with the timing for the College of Engineering, optimizing preparedness, time, resources, and cost.

Figure 9. ABET Accreditation Timeline

The statewide collaborative effort funded by the NSF EPSCOR grant [11] to create a data science ecosystem with four-year university "hubs" and two-year college "spokes" and a vision for "start anywhere, finish anywhere" 2+2 pathways is progressing well. We expect the first template and Memo of Agreement (MOA) to be completed and approved in this academic year. This template and MOA will be the basis for the "opted-in" two-year institutions and we are collaborating with them and with the state's department of education to allow all following the template to have expedited approval of their A.S. programs and for the MOA to be the basis agreements. We have also just received approval to bring the two-year colleges who are not in our "university system" (that is, the state's private universities) into our STEM Prep Program [12], completing the approach of bootstrapping them into the state-wide data science ecosystem. The first of the data science courses, *Introduction to Data Science*, was made available Summer 2022 and the second course, *Role of Data Science in Today's World*, will be made available starting Summer 2023.

Early in January 2023, the Data Science Program Administration held a two-day Program strategy and planning meeting with a subset of our faculty and our Data Science Advisory Council. The purpose was to develop a five-year vision and a longer-term vision based on a stated requirement by the industries in our state for an order of magnitude increase in the number of data scientists we graduate across the state. While this provides quite the challenge, it is one that provides an outstanding opportunity to contribute to the educational pipeline, to basic and applied research, and create to a dynamic economic engine by the partnering with industry, academic, and government in data science.

References

[1] Hiter, S. 2022. Data Science & Analytics Predictions, Trends, & Forecasts. Datamation. Data Science & Analytics Predictions for 2022 | Datamation. Datamation, accessed February 9, 2022.

[2] Robb, D. 2023. 6 Top Data Science Predictions for 2023. Datamation. 6 Top Data Science Predictions for 2023 | Datamation. Datamation, accessed February 6, 2023.

[3] Hughes, O. Employers are desperate for data scientists as demand booms. ZD Net. Employers are desperate for data scientists as demand booms | ZDNet, accessed February 9, 2022.

[4] U.S. Bureau of Labor Statistics. 2023. Occupational Outlook Handbook: Data Scientists, Data Scientists : Occupational Outlook Handbook: : U.S. Bureau of Labor Statistics (bls.gov), accessed February 6, 2023.

[5] U.S. Bureau of Labor Statistics. 2022. Occupational Outlook Handbook: Data scientists and mathematical science occupations, Data for Occupations Not Covered in Detail : Occupational Outlook Handbook: : U.S. Bureau of Labor Statistics (bls.gov), accessed February 9, 2022.

[6] Indeed.com. 2022. FAQ: Why Are Data Science Jobs in High Demand? FAQ: Why Are Data Science Jobs in High Demand? | Indeed.com, accessed February 6, 2023.

[7] National Academies of Sciences, Engineering, and Medicine. 2018. Envisioning the Data Science Discipline: The Undergraduate Perspective: Interim Report. Washington, DC: The National Academies Press. https://doi.org/10.17226/24886.

[8] National Academies of Sciences, Engineering, and Medicine. 2018. Data Science for Undergraduates: Opportunities and Options. Washington, DC: The National Academies Press. https://doi.org/10.17226/25104, p. 9.

[9] Schubert, K. D., & Rossetti, M. D. 2021. Creating a Multi-College Interdisciplinary B.S. Data Science Program with Concentrations. Paper presented at 2021 ASEE Virtual Annual Conference Content Access, Virtual Conference. https://doi.org/10.18260/1-2--36868.

[10] ABET. 2020. Diving into Data Science. https://www.abet.org/diving-into-data-science.

[11] NSF. 2020. DART – Data Analytics that are Robust and Trusted," Award No. OIA-1946391.

[12] University of Arkansas STEM Prep Program, Home | STEM Prep | University of Arkansas (uark.edu), accessed April 18, 2023.