



## Roundtable on Data Science Postsecondary Education

Meeting #11 - June 12, 2019

The 11th Roundtable on Data Science Postsecondary Education convened virtually on June 12, 2019. Stakeholders from data science education programs, government agencies, nonprofit organizations, professional societies, research organizations, foundations, and industry discussed current efforts in developing data science curricula and programs at 2-year colleges, opportunities for professional development in data science education, strategies for building partnerships with nearby 4-year and master’s-granting institutions, and techniques for understanding the needs of local employers. This Roundtable Highlights summarizes the presentations and discussions that took place during the meeting. The opinions presented are those of the individual participants and do not necessarily reflect the views of the National Academies or the sponsors.

Welcoming Roundtable members and participants to the meeting, co-chair Kathy McKeown, Columbia University, noted that as tuition increases at 4-year institutions across the United States, enrollment at more affordable 2-year colleges continues to grow. At the same time, demand for employees with data science skills is expanding across industries. In light of these trends, participants explored emerging approaches for integrating data science into 2-year curricula as well as strategies to enable connections between 2-year colleges and other postsecondary institutions.

### SETTING THE LANDSCAPE: TWO-YEAR COLLEGES AND DATA SCIENCE EDUCATION

*Nicholas Horton, Amherst College*

Horton emphasized the important role that 2-year colleges play in the education system and in the development of a diverse and inclusive workforce. More than 6 million students are enrolled at 2-year colleges, representing approximately one-third of the total undergraduate student population in the United States. A large proportion of Pell Grant recipients, U.S. veterans, and historically underrepresented students are enrolled in 2-year colleges, where “open-door policies” provide accessible, affordable pathways—average annual tuition is approximately \$4,000. The committed educators and administrators who are focused on student success and community engagement in 2-year colleges face distinct structural and organizational challenges, Horton explained. He referenced the [December 2018 Roundtable meeting](#) in which speaker D.J. Patil, head of technology at Devoted Health and former Chief Data Scientist in the White House Office of Science and Technology Policy, highlighted the value of 2-year colleges, especially in light of the interdisciplinary nature of data science. Patil explained that his experience at a 2-year college gave him “three gifts”: a love of mathematics, an understanding of how to write in various genres, and confidence to succeed at the postsecondary level. He considered this experience to be a crucial “on-ramp” to his future success.

Horton provided an overview of the 2018 National Academies’ consensus study report [Data Science for Undergraduates: Opportunities and Options](#), which identified 10 components of “data acumen”: (1) mathemati-

cal foundations, (2) computational foundations, (3) statistical foundations, (4) data management and curation, (5) data description and visualization, (6) data modeling and assessment, (7) workflow and reproducibility, (8) communication and teamwork, (9) domain-specific considerations, and (10) ethical problem solving. He noted the need for students at 2-year colleges to develop the appropriate depth of understanding in each of these areas and for faculty to have the time and resources to support such learning opportunities. Select recommendations from the report include ensuring that 2- and 4-year institutions work together on issues related to data science education, training, and workforce development; attracting and retaining students who have varied backgrounds and levels of preparation to data science programs; and remaining flexible and developing incentives as programs evolve. The National Science Foundation (NSF) funded the [Two-Year College Data Science Summit](#) to highlight innovative and effective programs at 2-year colleges, delineate three pathways to serve students' unique needs (i.e., certificate, associate-to-transfer, and associate-to-workforce programs), and identify next steps. Recommendations that emerged from that summit included (1) creating courses with modern and compelling introductions to statistics, (2) ensuring opportunities to engage with realistic problems and real data, (3) reducing barriers to entry, (4) ensuring depth in algorithmic thinking, (5) requiring fluency in computational language, (6) infusing ethics, and (7) fostering active learning.

In closing, Horton reiterated that data science is not just for doctoral-, master's-, or bachelor's-level students and that a 2-year education offers the "only affordable game in town." Major changes in pedagogy and course content are under way in science, technology, engineering, and mathematics (STEM) pathways to ensure student success. With best practices in flux, he continued, the need for professional development and continuing education is increasing—faculty need incentives, time, and resources to prepare to teach data science. He provided a series of framing questions for the remaining sessions of the meeting:

- How do we ensure that data science programs attract and retain students with varied backgrounds?
- How do we ensure that faculty development programs are robust and effective?

- How do we develop curricula that instill data acumen and are responsive to workforce needs?
- How can we build/maintain/grow a 2-year college data science community?
- How do we build effective connections between 2- and 4-year institutions?
- How do we build effective connections between 2-year colleges and industry?

### **PANEL ON INTEGRATING DATA LITERACY INTO COURSEWORK AND DEVELOPING DATA SCIENCE PROGRAMS**

*Randy Kochevar, Oceans of Data Institute, EDC*

Kochevar explained that a dramatic change has occurred during the past several years: many people now learn about the world through streams of data from remote sensors. This presents a challenge for educators who are introducing students to data. Instead of working with personally collected data sets (i.e., dozens of measurements), students can now work with much larger data sets (on the scale of many megabytes). At the same time, visualization skills have become more sophisticated. Kochevar proposed that all K–16 institutions have a responsibility to help students develop the required skills to work with complex data sets. Because little research has been conducted on how to best cultivate these skills and limited awareness exists as to the value of data science educator training, he said that new strategies to "educate the educators" are essential. The Education Development Center's Oceans of Data Institute (ODI) promotes the data literacy of K–16 students by building a research-based learning progression, developing and testing curricula and tools, and acting as a hub to convene diverse stakeholders.

Using the acronym CLIP, Kochevar described big data as Complex (i.e., different types of data collected in different ways), Large (i.e., more data than would be needed to answer a specific question), Interactive (i.e., data visualization tools can be used to compare different data sets), and Professionally collected (i.e., not by students). He explained that by studying how people use data in the real world, it becomes possible to understand the foundational skills that students need to develop. To aid in this effort, ODI has created expert worker profiles, including the [Profile of a Big-Data-Enabled Specialist](#) and the [Profile of the Data Practitioner](#). In closing, Kochevar offered a definition of data literacy in the age of big data: "The data literate individual understands, explains, and documents the utility and limitations of data by becoming a crit-

ical consumer of data, controlling his/her personal data trail, finding meaning, and taking action based on data. S/he can identify, collect, evaluate, analyze, interpret, present, and protect data.”

*Joyce Malyn-Smith, Oceans of Data Institute, EDC*

Malyn-Smith reiterated that ODI works with educators to incorporate data skills into curricula and develops tools for big data career pathways, based on input from industry. ODI’s [tool kit](#) includes (1) expert worker profiles, (2) rubrics to guide assessment, (3) a gap analysis tool for assessing industry value and school capability, (4) a curriculum analysis tool, (5) a course planning tool, and (6) a stackable credentials model. Most recently, ODI established partnerships with four 2-year colleges (Normandale Community College, Bunker Hill Community College, Johnson County Community College, and Sinclair Community College) through an NSF-Advanced Technological Education (ATE) project entitled [Creating Pathways for Big Data Careers](#).

Malyn-Smith described her work at ODI trying to identify and articulate what data-related skills are used in the workplace and how local 2-year colleges could incorporate them into their curricula. By interviewing workers and convening focus groups, ODI developed expert worker profiles (e.g., the data practitioner, the data scientist) to capture the broad range of skills, knowledge, and behaviors that are required to be successful in different roles in the workplace. Educators could use these profiles to identify where and how in their curricula these skills are covered, often leading to curricular modifications. Similarly, students could use the content and vocabulary in these profiles to sharpen their resumes, and employers could use them to evaluate employee performance and create balanced teams. Another method Malyn-Smith described that can help 2-year colleges align their curricula with employer needs is to conduct a gap analysis—asking industry partners about their expectations for employees to complete specific tasks and asking educators how their curricula prepare students to complete these tasks. Comparing responses could help to identify gaps in student training, she added.

*Paul Hansford, Sinclair Community College*

Recalling Sinclair Community College’s 1887 motto—“find the need and endeavor to meet it”—Hansford discussed the institution’s innovative approaches to supporting its students. He noted that “closing the skills gap is part of [Sinclair’s] DNA.” Sinclair serves

30,000 students annually, offers more than 270 programs of study, has the lowest tuition in the state of Ohio, and is the largest workforce education provider in its region. It is among the top 5 percent of the 1,100 2-year colleges in the United States in terms of enrollment size, physical plant, and the variety and complexity of educational programs of study, according to Hansford. He noted that new programs at Sinclair have arisen in response to both market demands and the desire to embed data literacy, analytics, and science into decision making that benefits communities. He added that data should act as the foundation for decision making, not as a substitute for human judgment. He also observed that publicly available data tools are surfacing—for example, a 70-question [data literacy exam](#) and a resource on 17 [character traits of a data-literate person](#).

Sinclair offers three data programs via its Department of Computer Information Systems. A 1-year [technical certificate in data analytics](#) has been available since Fall 2012; the [Data Analytics Associate of Applied Science degree](#) has been available since Fall 2018; and [Data Fundamentals](#), a short-term technical certificate, will be offered in Fall 2019. Similar degree programs and/or certificates are also available across the disciplines of business information systems, geography, allied health, and marketing. He described Sinclair’s partnerships with ODI and NSF, which have helped to increase student interest as well as prompted inquiries from other institutions about replicating programs and acquiring resources. Hansford related that Sinclair’s future goals include learning from students’ field experiences, adjusting courses to meet the needs of the local market, working closely in mentorship with other institutions, and spreading domain-specific certifications across disciplines.

*Michael Harris, Bunker Hill Community College*

Harris described Bunker Hill Community College as a diverse campus, with a student population that is 25 percent African American, 25 percent Hispanic, 25 percent Caucasian, and 25 percent other. He explained that Bunker Hill has a three-phase data analytics program, which was devised based on ODI’s stackable credentials model (mentioned above). Bunker Hill also used ODI’s Profile of the Data Practitioner and its associated heat map to understand core competencies for data practitioners and to determine student learning outcomes.

Harris explained that the first phase of the data analytics program is a [data management certificate](#),

available since Fall 2015. Students receive an introduction to data science and data management, learn to work in groups, and solve real-world problems. This curriculum includes the following five courses: IT Problem Solving, Introduction to Big Data, Statistics, SQL Programming, and Advanced Excel. The [data analytics certificate](#), first offered in Fall 2017, is the second phase of the data analytics program. Students who have completed the data management certificate only have to take four additional courses—Data Analytics and Predictive Analytics, Python Programming, Database Programming, and Operating Systems—to earn the data analytics certificate. Students who start in this second phase of the program would have to complete all nine courses, Harris explained. The third and final phase of the data analytics program, an **associate's degree in data analytics**, will be offered for the first time in Fall 2019. To attain the degree, students must take a total of 10 core data courses (i.e., the 9 previously listed courses and 1 in data visualization), 2 general education science courses (which enable students to [transfer](#) to a STEM program in certain 4-year institutions in Massachusetts), 4 additional general education courses (including 2 English courses), and 3-4 mathematics courses (e.g., pre-calculus, calculus, statistics, and linear algebra).

## TEACHING DATA LITERACY IN THE CONTEXT OF ADVANCING WORKPLACE TECHNOLOGY

*Ann-Claire Anderson, Center for Occupational Research and Development*

Anderson described [Preparing Technicians for the Future of Work](#), a project funded by NSF to enhance STEM programs in advanced technology fields for 2-year colleges. The project was developed in response to several issues: the nature of work is changing rapidly; advanced technologies are eliminating some jobs and creating others; [NSF's 2016 10 Big Ideas](#) emphasize a new research agenda, including the “Future of Work at the Human–Technology Frontier”; and technicians are at the center of much of this “disruption.” The mission of the project is to “enable the NSF-ATE community to collaborate regionally with industry partners, *within and across disciplines*, on the transformation of associate degree programs to prepare U.S. technicians for the future of work,” she continued. As the project team tries to make predictions about the workforce in 2030, it considers the following industry-agreed-upon interconnected technologies: big data, autonomous robots, simulation, system integration, Internet of Things, cybersecurity, cloud computing, additive manufac-

turing, and augmented reality. The project is based on five suppositions: (1) Technology will continue to evolve in a cross-disciplinary way, (2) technicians will need a multidisciplinary skill set, (3) some new skills will emerge that are common across multiple technologies, (4) the core knowledge that all technicians must possess will need to be augmented, and (5) 2-year technical programs will need to adapt their curricula.

Anderson explained that the project team conducts employee interviews, visits industry sites, and convenes employers and educators on both a national and regional level to better understand issues that postsecondary institutions face relative to the future of work. From its conversations with educators and industry representatives, the project team has identified three sets of crosscutting skills: data knowledge and analysis, business knowledge and processes, and advanced digital literacy. She described a number of pathways to develop these skills: 2-year college transfer programs, 1-year certificate programs, degree programs aligned with field specialization, stand-alone courses, microcredentials, advanced coursework for returning professionals, and boot-camps/continuing education. Anderson noted that because many supervisors want to hire people with industry-relevant experience, technicians often go to work immediately after attending a 2-year college instead of pursuing a 4-year degree. Because much of the data science work being performed today will be completed by people with 2-year degrees, she suggested that data analysis be integrated into technical programs and taught in the context of real work—technicians need to be able to manipulate, interpret, compare, contrast, merge, and operate on data to resolve problems, while using Excel and other common software. This requires institutions to revise mathematics prerequisite courses to reflect the changing demands of the skilled workforce.

She concluded that the project continues to examine what the future holds for STEM education at the associate's level. Next steps include interviewing skilled technical workers about new technologies and needed skills, convening educators and chief executive officers who represent a range of technical disciplines, adopting existing competency frameworks (from ODI and the U.S. Department of Labor) to identify specific skills required by industries of the future, developing recommendations for associate degree programs in advanced technology, collaborating with 2-year colleges and companies to facilitate the implementation of recommendations, and facilitating

the ongoing work of regional networks dedicated to training technicians for the future.

## ROUNDTABLE DISCUSSION

Victoria Stodden, University of Illinois at Urbana–Champaign, wondered how 2-year colleges bridge the needs of different student populations—for example, those who transfer to 4-year colleges and those who enter the workforce. Anderson emphasized the distinct differences between the pathways for these populations. She noted that institutional structure and/or state requirements influence a 2-year college’s ability to serve both populations. Malyn-Smith pointed out that when analyzing courses across four 2-year colleges to create the stackable credentials model, the only difference between data science pathways for students who planned to transfer and those who planned to enter the workforce was one mathematics course—this demonstrates that 2-year colleges likely can provide appropriate trajectories for a variety of students. Horton agreed and pointed to California’s system of higher education, in which courses are clearly mapped for students to transfer from a 2- to a 4-year institution. This transition is even smoother for dual enrollees (i.e., high school students taking community college courses), in his view.

Uri Treisman, University of Texas, Austin, described data science programs as “powerful resources for students seeking upward mobility.” He wondered about student enrollment in Sinclair’s and Bunker Hill’s data programs. Harris said that, historically, approximately 70 percent of Bunker Hill students have been students with undergraduate or graduate degrees who were seeking data science skills for industry jobs, and 30 percent have been people seeking first-time degrees. Similarly, Hansford said that 80 percent of Sinclair’s certificate students are people who are retooling. Jeffrey Ullman, Stanford University, and Treisman asked about the role of traditional foundational coursework (e.g., mathematics, statistics, business, and/or computer science) versus more applied courses in these data science programs. Specifically, Ullman wondered whether emerging data science programs deemphasize the study of methods and foundations. Harris explained that after consulting with representatives from industry, Bunker Hill decided to add data science projects to fundamental courses so that students would receive a balanced education. Hansford noted that Sinclair’s data curriculum includes several classes that emphasize fundamental content (e.g., programming, operating systems, mathematics, statistics) as well as additional courses to align with state

requirements. Treisman pointed out that employers play an important role in the survival of institutions; he asked how to manage programs so that they best serve students and meet the demands of both the institutions and industry. Hansford said that Sinclair conducts annual reviews with industry to discuss its curriculum and plans to seek feedback from alumni in the field. Harris said that he meets each semester with a representative from industry and is currently setting up articulation agreements with 4-year institutions.

## BREAKOUT GROUP DISCUSSION: DATA SCIENCE CAREERS AND INDUSTRY PARTNERSHIPS

Horton (moderator) posed a question about the level of skill mastery expected with an associate’s degree in data science. Anderson noted that data continue to be collected from companies and technicians to understand what skills are required by the workforce and what type of on-the-job training is available. For example, critical thinking might be more important for a technician in a particular role than a specific mathematics skill set. According to Malyn-Smith, ODI’s profiles and rubrics are continually revised based on feedback from current practitioners. Horton then asked how to determine which foundational skill sets are better suited for an associate’s degree than a bachelor’s degree. Malyn-Smith noted that although the biotechnology industry originally sought individuals with bachelors’ degrees, as 2-year colleges enhanced their programs, employers found that individuals with associates’ degrees were well suited for many of their jobs.

Horton asked how students make a smooth transition from a 2-year college to the workforce. Shalita Giannini, Milwaukee Area Technical College, noted the importance of integrating hands-on projects and assessments that relate to the real world to best prepare students for the workplace. In response to a question from Mark Tygert, Facebook Artificial Intelligence Research, Horton said that while co-ops are popular at 4-year institutions, they are starting to emerge at 2-year colleges. He noted that capstone projects with realistic expectations also provide valuable training for students. Anderson suggested that students do apprenticeships or internships—strong partnerships are needed between employers and institutions in order for these to be worthwhile experiences. Malyn-Smith agreed and proposed that institutions consult employer advisory boards when designing and revising programs. Anderson added that early recruiting strategies and dual-enrollment opportunities also show promise. Asia Mieczkowska,

University of North Carolina at Chapel Hill, said that aligning workforce needs with broader foundations is important. Malyn-Smith remarked that some simple strategies are being overlooked, such as inviting guest speakers to class or taking students to visit companies. She and Anderson added that having instructors visit employers could also be helpful. Tyler Kloefkorn, National Academies, asked how to foster collaboration between academia and industry. To begin a partnership, Anderson suggested engaging colleagues who have technical connections as well as designing multidisciplinary courses to help build bridges within a community. In response to a question from Jennifer Travis, Lone Star College, Horton said that while buy-in from multiple programs is important, the 2-year data science landscape is heterogeneous and a clear set of best practices for creating these partnerships does not yet exist. Angelika Gulbis, Madison Area Technical College, added that her institution employs a liberal arts internship coordinator who maintains relationships with industry partners. Horton noted the value of scaling and replicating such models while maintaining flexibility.

### **BREAKOUT GROUP DISCUSSION: DATA SCIENCE LITERACY, CURRICULA, CERTIFICATES, AND DEGREES**

Kochevar (moderator) explained that this discussion would focus on how 2-year colleges decide whether to offer degrees or certificates. Jean Wilson, Carroll Community College, proposed consulting local employers—for example, would they hire a student who has a certificate instead of a degree? Nicki Kowalchuk, Milwaukee Community College, noted the difficulty in motivating employers to accept 2-year college graduates for data analyst positions and expressed a broader concern that a 2-year degree may not be sufficient for most employers. Kochevar shared his experience working with Columbia College and regional businesses to develop an internship program to help bridge this gap between 2-year colleges and local employers. He described this as an effective way to evaluate how students fit into the work environment when leaving their degree or certificate programs. Kowalchuk responded that although Milwaukee Community College has established a partnership with Northwestern Mutual and is seeing increases in the employment of 2-year graduates, a master's degree is still preferred by many employers. In response to a question from Kelley Engle, Harrisburg Community College, Hansford responded that businesses are receptive to Sinclair's 1-year certificate program, which primarily serves students with 4-year degrees who are seeking to add a specific skill set.

Treisman noted that Indian River Community College, Alamo College, and Austin Community College have long-term relationships with employers and might have best practices to share (e.g., collocation of facilities at community colleges).

Kochevar asserted that data literacy will eventually be part of every job. He said that students need to develop skills, starting in elementary school, which will allow them to move in and out of the world of mathematics gracefully through quantitative thinking. He asked how best to build data literacy into 2-year college curricula. Linda Grisham, Massachusetts Bay Community College, noted that NSF has promoted data literacy (e.g., through its BioQUEST and QUBES programs), but disciplines remain siloed. She added that faculty need professional development to change their approaches. Hansford proposed that traditional literacy (i.e., reading, writing, and mathematics) be reconfigured to include courses in data visualization, Python, and R. Treisman explained that local employers seek employees with general data savviness and that quantitative literacy and data acumen are becoming increasingly important at 4-year institutions. Harris said that, to prepare students who plan to transfer to 4-year institutions, Bunker Hill will be offering a data visualization course as an elective. Treisman noted that while 4-year institutions are using R, 2-year colleges often have fewer resources to allocate to software modernization. He added that a systems approach, as well as a governing authority, is needed to facilitate the transitions between 2- and 4-year institutions.

### **CASE STUDIES: OPPORTUNITIES AND CHALLENGES**

#### **Adopting Data8 at a 2-Year College** *Ava Meredith, Seattle Central College*

Meredith stated that Seattle Central College surveyed 100 of its students and discovered that approximately 80 percent had heard of data science/data analytics, and approximately 60 percent were interested in taking a data science/data analytics course. Based on student interest and industry needs, the mathematics and information technology faculty at Seattle Central identified the need for a data science curriculum and decided to adopt a version of **Data8**—a popular introductory data science course at the University of California, Berkeley, that combines inferential thinking, computational thinking, and consideration for social issues in data analysis. The course is designed to be accessible to a broad range of students because it does not require prerequisites beyond high school

algebra. Meredith explained that Seattle Central will adopt six goals of the Data8 course: diversity, equity, pedagogical clarity, scalability, depth, and barrier-free entry. Before implementing any new program, she explained that the curriculum should be aligned to students' backgrounds and needs; administrative constraints should be addressed; and the decision to offer an associate's degree, a certificate, or a single class (for transfer or workforce education) should be evaluated.

Core concepts from Data8 will be included in the Seattle Central curriculum, course content will be managed with Jupyter notebooks, and the course language will be Python3, she continued. However, there are a number of areas in which Seattle Central's approach differs. Instead of offering Data8 in its original integrated format, Seattle Central will offer the program as a set of linked courses: Introduction to Data Analytics and Introduction to Statistics. Students will register for both courses concurrently, and faculty will coordinate the coursework. Meredith explained that Seattle Central opted to focus the course on "data analytics" instead of "data science" after research indicated that unlike data science jobs, data analytics jobs do not require a master's degree or a PhD. Software installation will be part of the curriculum. Instead of working with clean data, Seattle Central students will work with imperfect data sets and real Python libraries and will use Github as a code repository and for assignment submissions. Lastly, the curricula will be offered in flexible modalities (e.g., hybrid and eventually online). Students who choose to pursue a certificate in data analytics will take two additional courses: Python and Database and Data Visualization. Meredith described next steps to include piloting both this new data analytics course and a certificate in data analytics in Spring 2020, developing a plan to advertise and attract a diverse student body, collaborating with the social sciences department to create connector modules and to work with its data sets, partnering with other institutions, and identifying faculty training opportunities.

### **DataUP: Increasing the Capacity for Data Science Education**

*Renata Rawlings-Goss, South Big Data Regional Innovation Hub*

Rawlings-Goss described the objective of the South Big Data Hub: to connect industry, government, and academia around larger issues for societal and economic development, such as data science education and workforce. In 2016, the South Big

Data Hub hosted a workshop—Bridging the Data Divide: Partnering with Diverse Schools to Broaden the Pipeline—in which more than 60 people from 2-year colleges, minority-serving institutions, 4-year liberal arts colleges, government, and industry participated. A consensus report, *Keeping Data Science Broad: Negotiating the Digital and Data Divide among Higher-Education Institutions*, emerged in 2018 from this workshop, detailing 13 challenges, 16 visions for the future, 10 tasks, and concrete next steps for data science education. Two of the challenges highlighted in this report centered on how to implement data science curricula at institutions without the necessary technology stack as well as how to design relevant faculty training.

She explained that [DataUp](#), launched in January 2018, addresses these challenges by providing hands-on training for instructor teams at minority-serving institutions, 2-year colleges, and 4-year liberal arts colleges. 2018–2019 DataUp awardees were Spelman College; the University of Puerto Rico, Rio Piedras; the University of the Virgin Islands; Texas A&M, Kingsville; Florida A&M University; Johnson C. Smith University; and Old Dominion University. Faculty (and students, in some cases) teams applied to participate in the year-long program that included a 2-day data science workshop and a train-the-trainers workshop. The train-the-trainers workshop included a partnership with Software Carpentry—upon completion, the teams are certified, supplied with resources, and expected to conduct data science training workshops in their regions. In its effort to democratize data tools, the South Big Data Hub also piloted a project to host a Jupyter Hub. Teams who participated in DataUp were able to use this software during the 2-day workshop to design their curricula.

Rawlings-Goss described possible improvements for the 2020 DataUp Program: (1) Because administrative pressure can constrain community college and tribal college participation, administrators should be included in the process prior to application. (2) Faculty time to participate in external training is limited, so the benefit to the college must be justified, and there must be a clear alignment between the training program and the institution's goals. (3) Decisions about course-level activity do not always reside with instructors, so it is important to identify course- and non-course-related activities that could be counted toward program completion (e.g., boot camps, meet-ups, or student groups). She encouraged Roundtable participants to engage with the South Big Data Hub community by subscribing to

its monthly newsletter, reading the HubBub blog, joining the South Hub Google group, watching the South Big Data Hub YouTube channel, and following @SouthBigDataHub on Twitter.

### **Data Science: A Community College Approach** *Mary Rudis, Pennsylvania State University Harrisburg*

Rudis described her presentation as a “story of hope for greater inclusiveness and diversity for tomorrow’s coders, leaders, data practitioners, researchers, and innovators.” She referenced a recent report from the Association for Computing Machinery, *Lighting the Path from Community College to Computing Careers*, which contains case studies about unique approaches to implementing computer science educational pathways across 2- and 4-year institutions in New Jersey, Kentucky, California, Oregon, and Hawaii. She also encouraged software developers to connect with 2-year colleges to offer support or host professional development.

Rudis noted that the Community College System of New Hampshire was awarded a 2013 Innovation Fund Grant to create an undergraduate certificate in data science at Great Bay Community College and Manchester Community College. The objectives of the grant were to support the needs of private sector companies in greater New England by developing a modern curriculum to create a data-literate workforce; providing a foundational set of coursework that students could apply immediately and transition into a 4-year (or higher) data science/analytics degree; and enhancing existing computer science/computing resources with modern data analytics and visualization tools. First offered in 2015, the [Certificate in Practical Data Science](#) removes barriers to entry (i.e., only college-level composition and reading skills are required), offers a more modern approach to mathematics and models courses for liberal arts majors (e.g., the mathematics elective transfers to the University of New Hampshire), is marketed to high school mathematics students, and presents a schedule appropriate for students who rely on financial aid. The 1-year program includes Pre-Calculus, Elements of Data Science, Introduction to Python or Introduction to C++, Probability and Statistics for Scientists, Data Analysis, Visual Language, and a summer capstone project. Rudis clarified that this is not intended to be a “direct-to-workforce” certificate. Mathematics pathways were a barrier for students to complete the certificate program, so bridge courses (e.g., discrete mathematics) had to be developed to enable students from various tracks to move easily into a data science program. Rudis suggested that institu-

tions take the process of implementing a data science program slowly, despite any external pressure that might exist, and carefully contemplate how courses will be taught and what professional development will be needed. Direct-to-workforce programs are different from transfer programs; course redesign will be necessary to meet the needs of the 21st century workforce, she concluded.

### **Coordination and Collaboration Between 2- and 4-Year Institutions** *Lior Shamir, Lawrence Technological University and Kansas State University*

Shamir asserted that many 4-year institutions are well funded and suggested that 2- and 4-year institutions collaborate so that resources are allocated, shared, and used more effectively and equitably. Opportunities for collaboration include transfer programs; joint faculty training activities; research experiences; shared access to instructors, courses, and retention-driven resources; and integrative data science programs. Owing to the limited number of data science programs at both 2- and 4-year institutions, few data science transfer programs currently exist. However, he suggested that institutions think about the potential for transfer as they design their programs and begin to develop articulation agreements. He also emphasized the need to create a “soft-landing” for transfer students, who are entering a new environment—the need for institutional readiness to offer this support is often underestimated. For example, faculty training is especially important to alleviate stereotypes that 2-year colleges are not as rigorous as 4-year institutions. He also observed that 2-year colleges are often more diverse than 4-year institutions—teaching should be culturally responsive, embedding students’ cultures in the learning process.

Shamir noted that, by definition, data science is a research job (i.e., making discoveries from data), yet research at 2-year colleges is underfunded. One approach to ensure that research is included in students’ training is to incorporate Research Experiences for Undergraduates (REUs); however, some students will not be selected, others do not view themselves as researchers, and many do not have time for such a commitment. As a result, the REU model may not be the best solution for 2-year colleges. Instead, a course-based research experience (CRE) might be better suited to students’ needs, he continued. Community college students can complete the CRE at a partner 4-year institution and transfer the credit toward their associates’ degrees. The CRE includes the use of scientific practices, discovery, broadly relevant or important work, collaboration, and iteration.



This type of experience serves a larger number of students and does not require any extra-curricular involvement.

## ROUNDTABLE DISCUSSION

Rachel Levy, Mathematical Association of America, wondered how specializations arise and progress as well as how they are categorized, especially in the midst of improving the feedback loop among workforce, industry, and academia. Brandeis Marshall, Spelman College, said that because careers are continuously evolving, industry and academia need to communicate about relevant skill sets and options for job titles. Levy commented on the interesting landscape of 2-year colleges, and Treisman remarked that new mathematics pathways allow students to take courses with a combination of computational, statistical, and mathematical thinking. He suggested that the data science community and mathematical societies capitalize on these reforms. Rudis said that 2-year colleges would welcome more leadership in this area, but she wondered whether this reform of mathematics teaching is happening throughout the educational system. If not, students could encounter challenges when transferring from a 2- to a 4-year institution. In response to a question from McKeown about the proportion of the 2-year college population that could face this barrier, Shamir noted that 20 percent of 2-year graduates transfer to 4-year institutions.

Rudis said that much of what informs how courses are taught depends on the expertise and interests of the faculty. Marshall added that instructors matter, especially in terms of representation of marginalized groups. McKeown appreciated the strategies shared by Rudis and Shamir to remove barriers to entry and to embrace students' cultures and communities, respectively. She wondered how to attract students to mathematics who initially might not be interested in the discipline. Shamir highlighted Wright State University's approach in which engineers take mathematics that is relevant to their field. He noted that the K–12 system has a different mission than the higher education system, which can create knowledge gaps in certain academic areas that need to be closed. Rudis highlighted the importance of partnering with local K–12 institutions and beginning to target students as early as 5th grade. Students could attend mathematics camps hosted by community colleges; however, it is difficult to secure funding for such activities. An online participant asked if best practices for engaging students transfer from one 2-year college to another. Shamir replied that each 2-year college is different, so it is important to understand and tailor approaches to each unique system.

Treisman said that the demographics of 2-year colleges are changing. For example, 2-year colleges in many states are moving to joint programs with K–12 to remain fiscally viable, and the mathematical societies are considering how to integrate K–12 standards with postsecondary institution objectives. It is thus becoming easier to introduce ideas about data acumen into the K-12 curriculum. He reiterated that the demand for students' data knowledge is increasing immensely at the 4-year level, and 2-year colleges will need to develop students' data savvy in a coherent way. Gulbis wondered if it is possible to create a national standard for technician education. Anderson said that while it is possible, it is impractical. Two-year colleges prepare students for hundreds of different jobs, so while some essentials could be standardized, once they specialize in later years, there is not a one-size-fits-all approach. Shamir agreed with Anderson and said that much can be done through integrated data science programs.

## BREAKOUT GROUP DISCUSSION: COORDINATING WITH OTHER POSTSECONDARY INSTITUTIONS

Horton (moderator) asked about the typical barriers that a data science student encounters when transitioning from a 2- to a 4-year institution and best practices to ease this transition. Shamir responded that computer programming can be a barrier; however, it is possible to work in data science without mastering computer programming. Horton added that it is important to think about meaningful pathways for students—allowing students to engage with data that are interesting to them can lead to the improvement of algorithmic thinking skills. He also cited considerations for restructuring courses; for example, it is impractical to require computer science before having students work with data, and students cannot be expected to complete an entire series of calculus before being introduced to statistics and modeling. Shamir responded that data science can start with data-driven thinking, and algorithmic thinking can follow later—if algorithmic thinking is a prerequisite, more barriers to entry will be created for students. Jessica Utts, University of California, Irvine, noted that California State University, East Bay, has a [data science track](#) for statistics majors that does not require calculus and instead teaches using randomization-based methods, thus eliminating the barrier of calculus for transfer students. Horton pointed out that calculus is not included in the list of mathematical foundations for data acumen in *Data Science for Undergraduates: Opportunities and Options*. He added that useful levels of mathematical foundations and programming

knowledge may be different depending on the type of program and the type of future job. He contrasted engineering programs, where traditional mathematics and computer science backgrounds are required, with business programs, which have fewer requirements in these areas.

Gulbis noted the importance of liberal arts and social sciences to the data science curriculum and added that companies such as Apple hire individuals with backgrounds in both technology and liberal arts. Doris Dzameshie, AISCITE Institute, advised getting students involved with Github and company hackathons. Horton added that teaching data science across the curriculum is important so as to develop capacity in all students. Shamir commented that it is essential to define what counts as a “foundation” of data science. David Bapst, Texas A&M University, agreed and noted that many STEM PhDs working in industry on data science problems may have little coursework in programming, mathematics, or statistics but have strong skills in using statistics and programming to seek an answer to a particular question. Horton said that many data science projects involve up to 90 percent of time wrangling data; this is equally true for undergraduate students. Bapst added that tools change quickly and unpredictably; data science curricula should be agnostic to the language or tools— which means that the coursework need not be tied to a specific set of instructors—and updated regularly based on feedback from professionals.

John Hamman, Montgomery College, noted that it is challenging for 2-year colleges to align with multiple 4-year institutions. He noted that delaying programming and calculus coursework could make it difficult for students to transfer to a 4-year institution. Treisman emphasized the need for regional processes to negotiate transfer. Horton noted that in California, articulation agreements between 2- and 4-year institutions are structured with an online database of courses; in other states, they are arranged by state legislation. Hamman said that Montgomery College focused its efforts on aligning with programs at specific institutions, emphasizing that both administrators and faculty should be actively involved in developing these relationships. Treisman stated that data are needed to understand the magnitude of the equity problem that exists for students who transfer from a 2- to a 4-year institution. Shamir pointed out that administrators and faculty at 4-year institutions need to be prepared to work with transfer students from 2-year colleges, which requires training. Treisman agreed and noted that students from 2-year colleges can add much diversity to a 4-year institution.

Kathryn Linehan, Montgomery College, described the challenge that arises in transferring course credits from a 2-year college to a 4-year institution. Treisman responded that student success is a necessity to maintain enrollment, and further work on fairer articulation agreements could help to address this equity problem. If a course will not transfer to a 4-year institution, it likely will not survive. John McKenzie, Babson College, noted that there is a Classification of Instructional Programs code for data science.

## **BREAKOUT GROUP DISCUSSION: ENHANCING PROFESSIONAL DEVELOPMENT AND ADOPTING EXTERNAL CONTENT**

Levy (moderator) asked what resources, programs, and activities exist to support 2-year college faculty in teaching data science. Meredith advised that industry be consulted for guidance on this topic. Grisham described the [BioQUEST Curriculum Consortium](#), which has 33 years of project work and resources as well as week-long workshops for high school and college life science faculty. She also cited QUBES, an NSF-supported project aimed at faculty professional development, which is comprised of a community of mathematics and biology educators. She elaborated that the community typically shares methods and resources to help prepare students to use quantitative approaches to address real, complex biological problems. Levy added that QUBES hosts the mathematics modeling hub, and Grisham noted the importance of building a community, as these groups tend not to interact. Karen Coghlan, National Network of the Libraries of Medicine (NNLM), added that NNLM provides free webinars, classes, and materials for teaching and for research data management. In response to a question from JoEllen Green, Fresno City College, Rudis said that RStudio Cloud eliminates the need to install software and enables collaboration. Eric Simoneau, [STATS4STEM.org](#), noted that RStudio Cloud is currently in alpha mode, which can result in dependability issues.

Rawlings-Goss inquired about institutions that have training programs from industry and wondered how those trainings are received, while Meredith considered the cost of training with certain companies as well as the cost to license technology to an institution. Tygert noted that industry is currently investing heavily in education and training because it has the funding that governments and nonprofit organizations typically do not have. He elaborated that while these efforts are focused on developing students’ skills for future careers, there is also a focus on basic science and research and development. Shir-

ley Usry, Hawkes Learning, asked how textbook and web content for students can keep pace with the evolving field of data science. A participant noted that this phenomenon is inevitable in such a dynamic field; it is important to focus on generalizable skills, knowledge, and behavior rather than focusing on specific nuances of a particular piece of software. The participant continued that while specific tools are useful for providing hands-on experience, it can be valuable to expose students to a variety of tools and then key-in on underlying shared principles. Scott Tousley, Splunk, noted the similarly rapid pace of innovation in cybersecurity.

## GUIDED REFLECTION AND NEXT STEPS

*Brian Kotz, Montgomery College, and Uri Treisman, University of Texas, Austin*

Kotz concluded that several organizations have expressed their desire to support or partner with 2-year colleges, thus increasing the visibility of 2-year data science education. Two-year colleges serve a wide range of students: the average Montgomery College student is over 25, all are exclusively commuters, and some only take a course or two. While community colleges can offer nimble customization, funding and resource constraints make it difficult to implement new programs. He cited two key themes from the meeting: the value of high-quality collaboration and independent customization to meet the unique needs of student populations.

Returning to Horton's framing questions for the meeting, Kotz offered the following commentary:

- **Advocating**—Demonstrate how important data science is and how it impacts all aspects of life
- **Advertising**—Meet students face-to-face and raise awareness
- **Managing expectations**—Success means better-informed students with marketable skills
- **Showing what the students can do**—Share student capstone projects externally such as with local government
- **Assessing students and curricula**—Prepare students for larger goals beyond their next job
- **Evolving**—Maintain flexibility, incentives, and resource-sharing
- **Continuing to reflect and discuss**—Remain open to new perspectives and definitions
- **Offering professional development**—Support educators so they can support students

Kotz also elaborated on topics that he would like the data science community to discuss in more depth in the future:

- **Storytelling**—Are students being trained to communicate about data efficiently and effectively?
- **Data analysts and data architects**—Does data science mean “playing in people’s backyards” or “building and forming people’s backyards”?
- **Distance education**—Do collaborative teams and open resources exist?
- **Local government**—Can students serve their communities through rewarding partnerships?
- **Ethics and privacy**—How are these topics being integrated in 2-year programs?

He hopes to see a platform in the data science education community that enables (1) frequent meetings, (2) systemic structural reforms, (3) an improved understanding of the capabilities of 2-year colleges and their data science students, (4) improved communication within and across institutions and between organizations, (5) a welcoming of others, (6) increased equity for students so that their circumstances do not affect their access and opportunity, and (7) the potential for the 2-year college model to be embraced for data science. Doing so will empower students to change their lives and those around them, Kotz asserted.

Treisman thanked participants and noted that many of the practices discussed are worthy of attention. New structures will be needed to allow institutions to coordinate the development of their data-rich programs, and state governance and professional societies will need to play a role in helping to level the playing field for 2-year colleges. He reiterated that “transfer” is not just from 2- to 4-year institutions; it also involves students moving from 4- to 2-year institutions and from high school to community college. Administrators need to think about models for back-office functions to enable these transitions. It is also important to think about the role of traditional academic departments in the evolution of courses that develop data acumen. This discussion should be complemented by policy and additional information about the jobs for which people should prepare, he continued. Data science will continue to evolve quickly, and evidence-based modernization of curricula needs to be supported, he concluded.

---

**ABOUT THE ROUNDTABLE:** The Roundtable on Data Science Postsecondary Education is supported by the Gordon and Betty Moore Foundation, the National Institutes of Health, the National Academy of Sciences W.K. Kellogg Foundation Fund, the Association for Computing Machinery, the American Statistical Association, and the Mathematical Association of America. Within the National Academies, this roundtable is organized by the Committee on Applied and Theoretical Statistics in conjunction with the Board on Mathematical Sciences and Analytics, the Computer Science and Telecommunications Board, and the Board on Science Education. Roundtable meetings take place approximately four times per year. Please address any questions or comments to Ben Wender at [bwender@nas.edu](mailto:bwender@nas.edu).

**DISCLAIMER:** This meeting recap was prepared by the National Academies of Sciences, Engineering, and Medicine as an informal record of issues that were discussed during the Roundtable on Data Science Postsecondary Education at its 11th meeting on June 12, 2019, during which more than 400 people participated virtually. Any views expressed in this publication are those of the participants and do not necessarily reflect the views of the sponsors or the National Academies.

---

## Division on Engineering and Physical Sciences

*The National Academies of*  
SCIENCES • ENGINEERING • MEDICINE

The nation turns to the National Academies of Sciences, Engineering, and Medicine for independent, objective advice on issues that affect people's lives worldwide.

[www.national-academies.org](http://www.national-academies.org)