

December 2017

Roundtable on Data Science Postsecondary Education

Meeting #4 - October 20, 2017

The fourth Roundtable on Data Science Postsecondary Education met on October 20, 2017, at Northwestern University in Evanston, Illinois. Stakeholders from data science education programs, government agencies, professional societies, foundations, and industry convened to discuss alternative mechanisms in data science education. 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. Watch meeting videos or download presentations at nas.edu/DSERT.

STANFORD UNIVERSITY'S CERTIFICATE PROGRAMS

Jeffrey Ullman, Stanford University

Ullman shared the history of Stanford University's professional certificate programs. In the 1960s, the School of Engineering broadcasted recorded lectures through the Stanford Instructional Television Network (SITN), and couriers delivered lecture notes to and collected homework assignments from local industry participants. Employers paid twice the tuition rate for their employees to complete a Master of Science in Engineering through the SITN. The SITN eventually became the [Stanford Center for Professional Development](#) (SCPD), which now offers a variety of courses and certificates worldwide via the internet. Participants do not have to apply to or enroll in the university to participate in SCPD programs.

He noted that although a graduate certificate is not equivalent to a diploma, it does hold more weight than a statement of completion since all certificate coursework is graded. The statistics department introduced the Data Mining and Applications certificate (3 courses) in 2009, and the computer science department followed in 2010 with the Mining of Massive Data Sets certificate (4 courses). Ullman emphasized the initial popularity of both programs but noted a decrease in enrollment since 2013, likely owing to the availability of more certificate programs in other disciplines of interest (e.g., artificial intelligence, cybersecurity). However, there has been a 50 percent increase in the

total number of graduate certificates awarded across Stanford University's departments since 2009.

Ullman turned to a discussion of two of Stanford's approaches to data science. Although the computer science department does not offer a data science degree, students can complete a data science specialization at both the undergraduate and graduate levels. In conjunction with the Institute for Computational and Mathematical Engineering, the statistics department offers a Master of Science in Statistics: Data Science. Another difference, according to Ullman, is that computer scientists utilize algorithms to solve problems, while statisticians validate the soundness of solutions. He challenged Drew Conway's **Data Science Venn Diagram**, noting that it fails to acknowledge the value of computer science's understanding and implementation of algorithms, and he displayed his own version of the Venn diagram that removes mathematics and statistics from the core of data science.

Ron Brachman, Cornell Tech, asked if matriculated Stanford graduate students are eligible to participate in certificate programs. Ullman noted that while it is possible, students are prohibited from cross-counting courses. Victoria Stodden, University of Illinois, Urbana-Champaign, asked Ullman about the role of university administration in sustaining the certificate model. He explained that individual faculty members propose content for certificate courses and emphasized that curriculum change occurs via bottom-up approaches. Challenging Ullman's version of the Venn diagram, Kathy McKeown, Columbia University, emphasized not only how much computer science and statistics overlap, but also how important statistics and mathematics are to the study and practice of data science.

BOSTON UNIVERSITY'S STATISTICS PRACTICUM

Eric Kolaczyk, Boston University

Kolaczyk described Boston University's commitment to developing students' data science skills, achieved through complementary top-down and bottom-up approaches to curricular innovation. In 2015, a **Master of Science in Statistical Practice (MSSP)** emerged, attracting a broad audience of quantitative students and producing holistically trained statisticians who have the foundational knowledge to work in an integrated data science environment. Participants enroll

in eight courses and complete both a written portfolio and a two-semester statistics practicum.

He explained the primary motivations for developing the MSSP: (1) hiring organizations were increasingly demanding both degree completion and experience from their applicants; (2) employers wanted to hire people with both technical and communication skills; and (3) faculty were becoming dissatisfied with current course content. Thus, this revised statistics curriculum is practice-centric and requires the integration of diverse skills (Figure 1). Instead of adjusting existing infrastructure, MSSP faculty created a new organizational principle integrating practice and pedagogy and adopted a cohort-based system. The MSSP practicum's success is dependent upon a steady stream of real-world problems that are right-sized for student group work on various time scales, according to Kolaczyk.

The practicum is taught by a team of faculty members, fellows, and teaching assistants. Each class includes assigned readings, quizzes, discussion, and group work on topics such as data manipulation, visualization, modeling, and analysis; inquiry and interpretation; process management, workflow, and reproducibility; and communication. Statistical consulting is also available on walk-in, limited, and collaborative levels as part of the practicum. For their final projects, students collect data from various sources and must deliver a presentation, report, code, and data products—many of these projects focus on issues in the City of Boston (e.g., service quality, homelessness).

Kolaczyk described a number of challenges both in the curriculum and for the instructor, including balancing pedagogy and practice; emphasizing process; regulating project scope and timing; increasing students' independence; and elevating standards, goals, and accountability. He reiterated that the MSSP is a practice-centric, results-driven curriculum, and fac-

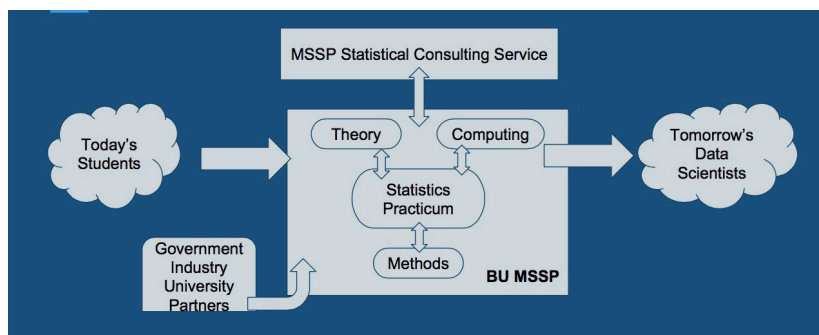


FIGURE 1 Boston University's practice-centric approach to the Master of Science in Statistical Practice. SOURCE: Eric Kolaczyk, Boston University, presentation to the Roundtable.

ulty must be prepared to modify course plans when necessary.

In response to a question from a participant, Kolaczyk explained that students analyze scenarios as a way to discuss issues of ethics, fairness, and misuse. Katy Börner, Indiana University, suggested the use of an online learning system platform to deliver course content and reveal learning analytics. Kolaczyk noted that although it would be possible to move theory education to an online learning environment, this could create even more challenges in tracking students' results in the practicum. In response to a question from James Frew, University of California, Santa Barbara, Kolaczyk described that students work in small groups for nearly every component of the practicum, which provides good preparation for future workplace experiences. Kolaczyk relies on a strategy of "benevolent guidance" to arrange students in balanced groups. Another participant inquired about the program's success in workforce placement, and Kolaczyk noted that, anecdotally, students are being hired in a variety of data science positions and applying the skill sets developed in the program. In response to a suggestion from Karl Schmitt, Valparaiso University, Kolaczyk remarked that he hopes to disseminate educational materials from the program soon.

In a response to a question from Stodden about the role of university administration, Kolaczyk noted that university-wide changes can be challenging, especially for institutions that contain multiple schools. It is also important to recognize that infrastructure and hierarchies vary from campus to campus, which creates additional challenges when trying to scale programs. Al Hero, University of Michigan, suggested the creation of institutes or cross-school units to serve as brokers, identify commonalities, and resolve differences among departments. Schmitt commented that because data science is and can be done well at liberal arts colleges, which are already set up to be integrative, many new technology hires come from institutions other than R1 research universities. He added that top-down approaches have been successful at Valparaiso, specifically, where co-teaching arrangements have been widely supported.

CORNELL TECH AND THE JACOBS TECHNION-CORNELL INSTITUTE

Ron Brachman, Cornell Tech

Brachman highlighted New York City's 2010 call to universities to start or expand an applied science/engineering campus. Cornell University and the Technion-Israel Institute of Technology won the

challenge, receiving \$100 million in capital and a plot of city-owned land on which to build. Brachman emphasized that this new university, Cornell Tech, was developed from the ground up. Structured to offer practically oriented graduate degree programs, Cornell Tech opened in 2012 and, with an additional \$130 million gift from the founder of Qualcomm, created the Jacobs Technion-Cornell Institute to focus specifically on application domain areas.

Brachman explained that Cornell Tech relies on a team-oriented "studio" environment that focuses on real problems to better prepare graduate students to enter (and lead) the digital technology economy. Interaction between academia and industry is key, and, because many graduates are hired locally, Cornell Tech has a direct impact on New York City's economy. He later added that Cornell Tech also employs a job placement staff, who contribute to the high placement rate of the graduates. A self-described "start-up company," Cornell Tech has graduated 324 students to date and currently has 250 master's students, 50 doctoral students, and 30 faculty on campus. According to Brachman, Cornell Tech plans to have 2,000 students and 200 faculty on campus by 2043.

Cornell Tech currently offers **seven master's degree programs**. Brachman confirmed that although the coursework is structured differently, these degrees are equivalent to those conferred by Cornell University. All seven programs incorporate team-based, project-based learning, starting in the first semester—external companies with real problems ask the students to develop and manage products. This integrated studio education makes up approximately one-third of the total coursework for students enrolled in 1-year programs and includes alternative educational activities such as 24-hour project sprints, weekly critique sessions with external practitioners, open studios, and opportunities to win monetary start-up awards. With such personalized attention and variety of rich experiences, he noted that this model could be challenging to scale.

Brachman commented that Cornell University does not yet have a specific degree or certification in data science. However, a cross-campus data science task force has emerged to evaluate current offerings and propose a new integrative structure for the future of "engaged data science." At Cornell Tech, specifically, faculty and administration are considering how the studio curriculum could integrate existing data science coursework into a degree or certificate program. In response to a question from a participant, Brachman remarked that, depending on the program,

some incoming students come to Cornell Tech directly after receiving their undergraduate degrees, while others enroll after some amount of work experience. McKeown asked about the level of interest from external companies to engage more than once in Cornell Tech's studio projects and whether there are any intellectual property issues with the data they share. Brachman noted that companies continue to return, and the project list continues to grow. He added that companies are required to have a representative participate actively with the students, and they must agree that any work done by the students will become open source. Ullman raised a concern about students' open source data because venture capitalists who support their start-up companies may want to control that data. Brachman agreed that this issue warrants further discussion.

AMERICAN STATISTICAL ASSOCIATION DATAFEST

Andrew Bray, Reed College

Bray described the [American Statistical Association's DataFest](#) as a weekend-long competition held each Spring on numerous campuses across the United States, Canada, and Germany. All participating host institutions must adhere to specified terms of use. During the competition, three to five undergraduate students—typically from the disciplines of computer science, statistics, engineering, business, social sciences, and natural sciences—work together to extract meaning from a complex data set (e.g., 10 million records from Expedia). The data set is not “revealed” until the first evening of the competition. The time that follows includes team time (all work must be done on site), support from on-site consultants, and optional workshops (similar to just-in-time teaching experiences). During the final evening of the competition, each team gives a 5- to 10-minute oral presentation of its findings to a panel of judges who are practicing data scientists from academia, industry, or the public sector. Two thousand students participated in 2017, and awards were given for best data visualization, best use of external data, and best insight.

Bray explained that DataFest gives students a sense of what it means to be statisticians as well as to be part of a community built around data. It is an opportunity for the students to practice the skills they have been taught in the classroom and to make connections with local data professionals. Students build technical and communication skills, learn to better generate and scope questions, and develop content for future job interviews. For faculty, DataFest can also

be helpful in revealing some of the knowledge gaps that exist in the current academic curricula.

Bray acknowledged that there are a number of organizational challenges associated with DataFest. It can be difficult to find data that is of interest to the students, is not too specialized, is sharable, has multiple angles of inquiry, and is of appropriate size to be manipulated on a modern laptop. He cautioned that events such as DataFest could encourage irresponsible or ill-conceived analysis, so consultants interact with students throughout the event in an effort to combat such behavior. In the future, Bray hopes DataFest will coordinate a national competition, diversify data, and continue to increase student participation.

Brachman agreed that DataFest provides an excellent opportunity to expose students to the differences between responsible and irresponsible analysis, and Stodden suggested that examples from previous years' competitions be used as models, eliminating the need to embarrass any current participants who may be engaging in faulty analysis. Bray added that this topic could be integrated into a future DataFest workshop. McKeown asked Bray how DataFest prevents the exposure of private information. Bray commented that they utilize de-anonymization and limit the number of covariates, as well as engage in lengthy discussion with participating companies' legal teams, but privacy continues to be challenging. He added that students will occasionally have to sign a non-disclosure agreement prior to participating in DataFest, or the company may lock up the data immediately after the competition concludes. David Ziganto, Metis, asked if DataFest has considered using synthetic data instead to help avoid such privacy issues. Bray explained that the original intent was to give students as authentic of an experience as possible with data as it exists in the wild, but he agreed with Ziganto that it is possible that even richer experiences could be had with synthetic data.

Hero asked what skills students need in order to participate in DataFest. Bray responded that students with some experience in a computational environment will be able to engage with their team and complete the challenge. He added that many students have worked in R, while some have experience with Java, Python, MATLAB, or Stata, depending on their home disciplines. In response to a question from Ullman, Bray noted that industry representatives from companies that provided the data set do attend DataFest and occasionally will engage in follow-up work with a team that shared interesting findings. Börner remarked that a non-profit organization

without the resources to finance such a competition would benefit greatly from having students work on its data problems and offer solutions. Bray said that, historically, non-profit organizations have not had the infrastructure to engage in DataFest, but he corroborated the value of having students do work with diverse data that could make a societal difference. Stodden asked about DataFest's level of integration with industry and wondered whether data science is being perceived as a scientific practice or as industry training. Bray explained that industry partners may support DataFest financially, and judges sometimes privilege a presentation with an actionable solution over one that is very scientific in nature.

BOOT CAMPS

David Ziganto, Metis

Founded in 2013, Metis offered its first **boot camp** in New York and now has locations in California, Illinois, and Washington. Ziganto explained that Metis's boot camp is the only one of its kind in the United States with endorsement from the Accrediting Council for Continuing Education and Training, though he hopes others will follow suit so as to improve the overall reputation of the boot camp model. In addition to a 12-week boot camp, Metis provides corporate training and online and evening courses in data science. He clarified that a boot camp is meant to bridge the gap between academia and industry and serve as a complement to other learning mechanisms. The boot camp model adjusts in real time to industry's demands for particular skills and technologies, while providing a fully immersive experience for participants.

Boot camp participants learn a combination of theoretical concepts and applications, including how to ask a solvable question, scope projects, collaborate and communicate with diverse groups, and use emerging tools and technologies. Ziganto added that Metis's boot camp allows students to work throughout the full data science pipeline on five projects (including posing the question and gathering the data), whereas in traditional academic settings, students typically enter the pipeline only when it is time to explore and clean the data (Figure 2). In his view, the boot camp approach to the data science pipeline gives participants practice being data scientists before entering the workforce as entry-level data scientists.

Ziganto described boot camp participants in three ways: (1) fresh graduates without a portfolio; (2) career changers with a strong programming background, weaker math skills, and no portfolio; or

(3) career changers with a strong analytical background, little programming experience, and no portfolio. Approximately 50 percent of participants have bachelor's degrees, while 49 percent have advanced degrees; 71 percent have industry experience, while 29 percent have experience in academia. He reiterated that a boot camp is meant to supplement hackathons, online courses, and advanced degrees, and he explained that successful boot camps have rigorous admission criteria, a rapidly evolving curriculum (partially influenced by employer feedback), instructors with industry experience, student-driven portfolio projects, and links to the data science community. For students enrolled in its boot camp, Metis provides career services to help with résumé development and networking. And for students who are not yet prepared to meet the admission criteria to enroll in a boot camp, Metis provides guidance for skill building.

Stodden cautioned about the use of the term "boot camp" to describe such a program, concerned that the label can stifle curiosity or lead people to believe that data science is exclusive. She suggested the use of the phrase "quick start" or "jump start" instead. Another participant remarked that the term "boot camp" has an entirely different connotation—that of a remedial program intended to get participants up-to-speed on missing qualifications. Ziganto reiterated that Metis adopted the commonly used title because it symbolizes the fully immersive experience in which participants engage to refine their skills and build their portfolios. The participant added that both remedial and finishing programs serve equally important purposes, so it is important to clarify to participants what type of program is being offered when the term "boot camp" is used. Schmitt asked whether boot camps are targeted to particular sectors of industry, and Ziganto responded the boot camps focus on more sustainable data science fundamentals, while the sector-specific needs are addressed in Metis's corporate training programs. Ullman noted that algorithms, not models, solve data science problems, but Ziganto explained that the end goal in the data science process is to have an approximation, and a model is an approximation of reality.

INFORMAL DATA SCIENCE EDUCATION

Stephen Uzzo, New York Hall of Science

Uzzo pointed out that while science practice has transformed, there has not been an equivalent revolution in science education. In his view, our ability to gather data has outstripped our ability to analyze it; new tools and techniques emerge rapidly; and data science pervades the science, technology, engineer-

ing, and mathematics learning ecosystem. Because data science problems are complex and interdisciplinary, data science has also transformed many other sectors of society. Yet, according to Uzzo, data science is generally not taught in any depth in the public school system, if at all, which ultimately threatens the pace of society’s technological progress. This gap between data-driven science and technology practice and the understanding of science and big data for lifelong learning can be closed with big data literacy programs in informal educational settings, he explained. He noted that the abilities to adapt, innovate, collaborate, and analyze are essential in a data-driven society.

Uzzo explained that approximately 95 percent of learning happens outside of a classroom, reinforcing the need for more informal science programming as well as for new technologies to access such educational opportunities (e.g., computational tools for visualization technologies). **The New York Hall of Science** (NYSCI) focuses on providing this needed data literacy to the public by offering knowledge when, how, and where the public can best engage with it. Science centers and museums exist, according to Uzzo, because people learn better by doing, embodying abstract ideas, and engaging with phenomena (such as big data) through sight, touch, and creation. Core principles of museum experience and exhibit design include (1) placing people and play at the center, (2) envisioning visitors as creators, (3) introducing worthy problems with divergent solutions, and (4) issuing an open invitation to participate. NYSCI strives to create immersive experiences

and share complex ideas to increase public interest and skills in science, which can be challenging for an audience of learners of various ages.

Catherine Cramer, New York Hall of Science

Cramer explained that NYSCI is situated in Corona, Queens, a community that is largely Spanish-speaking and includes 60,000 students—the largest school district in New York City. To support data literacy, the museum engages with local families, provides exhibits, offers public experiences, helps visitors understand new tools, organizes out-of-school programs, and hosts conferences. Cramer provided an overview of some of NYSCI’s recent and upcoming activities:

- *Connections: The Nature of Networks*. Large floor exhibit, displayed 2004–2014.
- *Network Science for the Next Generation*. 3-year program pairing high school students from New York City and Boston with graduate students to create and present network science research projects.
- *Network Science in Education*. Hosts of annual international symposiums and teacher workshops, as well as authors of “**Network Literacy: Essential Concepts and Core Ideas**.”
- *Big Data Fest*. 2015 event in which 40 organizations provided data activities for the public.
- *Northeast Big Data Innovation Hub*. Effort to generate a collaborative inquiry process and a framework of principles for big data literacy.
- *Estuary Science Complexity*. Plans to develop a new science center that focuses on the data-dependent field of estuary science.
- *Mobile City Science*. Program for students at New York’s International High School who recently immigrated to the United States. Students used GoPro videos to map their community, identify problems, gather evidence, and propose solutions.
- *Big Data for Little Kids*. A current workshop designed to understand how 5- to 8-year-olds define, collect, represent, and interpret data, as well as how their caregivers engage with them in data inquiry activities such as variation, measurement error, data aggregation, interpretation, and prediction via a “make-your-own museum exhibit.”

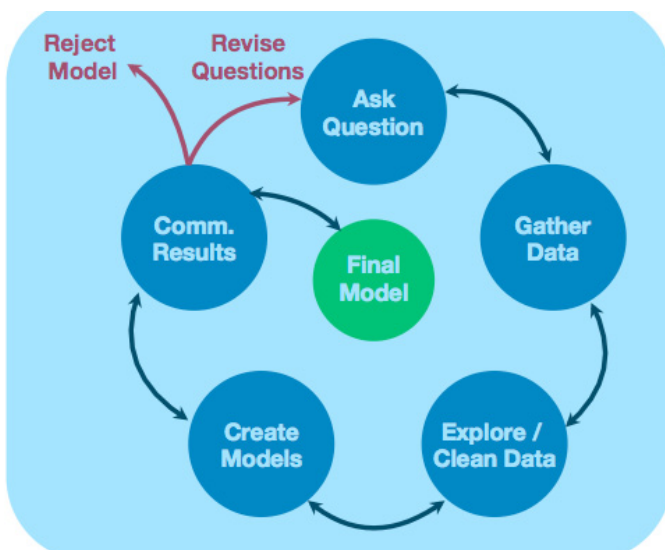


FIGURE 2 Metis’s comprehensive boot camp model of the data science pipeline. SOURCE: David Ziganto, Metis, presentation to the Roundtable.

- *DataDive Exhibit*. Playful and personally meaningful experiences with data that help visitors understand patterns, algorithms, and machine learning processes.

Katy Börner, Indiana University

Börner shared her work in defining, measuring, and improving data visualization literacy—a combination of literacy, visual literacy, and data literacy that allows one to read, make, and explain data visualizations—which is critical for success in our data-intensive global society. In a study of 1,000 children and their caregivers who regularly visit a science museum, she found that most were unable to name, read, or interpret common data visualizations. She emphasized the need to bring more “macroscopes” to public spaces to help people make sense of large-scale data streams, identify patterns and outliers, and observe trends (Figure 3). She explained that macroscopes are not static instruments but rather continuously evolving bundles of software packages. She added that with numerous types of questions, varying experiences and knowledge of users, and different levels of abstractions, it can be challenging to create such toolkits.

One way to scale this education is through massive open online courses (MOOCs). Since 2012, students from 100 countries have participated with over 350 faculty in [Indiana University’s Information Visualization MOOC](#). Participants look at different workflows, run different types of analyses and visualizations, and learn to work collaboratively through algorithms to develop an actionable visualization. She also described a new project under way (joint among the National Science Foundation, Indiana University, the Science Museum of Minnesota, NYSCI, and the Center of Science and Industry in Columbus, Ohio) titled *Data Visualization Literacy: Research and Tools that Advance Public Understanding of Scientific Data*. At the Science Museum of Minnesota, for example, a sports exhibit is available for children to explore and construct data visualizations after capturing their own race data and characteristics in a scatter plot.

Hero said that he has witnessed a decline in both data and visual literacy among high school students; he wondered how to reverse these trends and how to engage more students in science. Cramer noted that Network Science for the Next Generation students, for example, had little science training or interest in college prior to the program but became open-minded about their futures after the program. Even field trips to science museums can increase student interest in science, she added. Börner commented

that many high schools are actively teaching visualization skills, and the global population of the Information Visualization MOOC has not demonstrated the decline in literacy skills that Hero described. She suggested that if U.S. schools continue to “teach to the test,” data visualization questions can be added to those tests to increase data visualization literacy. Uzzo added that the [Next Generation Science Standards](#) for K–12 students emphasize modeling, suggesting that graphic literacy may be developed before high school begins. He also noted that network science is a field that appeals to students because of its focus on investigation; students can capture their interests from Harry Potter to human cells.

McKeown asked about strategies for diverse participation in informal settings, and Cramer noted that it took much work to engage her local community in NYSCI. Free museum entrance days and homework-help hours attract local families to the museum, which now has approximately 1,200 children visiting on a regular basis. Uzzo added that it is a challenge to appeal to and engage a wide age group in a single exhibit; however, many exhibits interest both adults and children when they simultaneously offer objects for children to manipulate and complex ideas for adults to ponder. He highlighted the importance of scaling the intellectual capacity of every space, especially since adults often accompany children to a museum. Börner supported intergenerational teaching and learning that happens outside of a classroom setting, in which people of different ages and experiences share knowledge with each other.



FIGURE 3 While microscopes and telescopes only reveal the infinitely small and infinitely large, respectively, macroscopes allow one to study the infinitely complex. SOURCE: Katy Börner, Indiana University, presentation to the Roundtable.

SMALL GROUP DISCUSSIONS AND CONCLUDING CONVERSATIONS

Following a set of small-group discussions, Börner and McKeown shared considerations raised by their group for scaling data science programs. They noted the importance of trying to reach as many people as possible through varied methods of both formal and informal education. They suggested that libraries and museums serve as distribution systems for information that is not as readily accessible in rural areas as it is in urban environments. They cautioned about educational inequalities that exist owing to the economic circumstances of individuals or the resources of educational institutions. They lauded the value of experimentation and personalization in curriculum design. However, they noted that strategies that work in one setting may be difficult or inappropriate to scale in another. Lastly, they encouraged the development of top-down structures for program development.

Mark Krzysko, Department of Defense, asked the Roundtable to consider carefully the definition of scale and the purpose of data science training. He encouraged start-up style thinking across campuses and emphasized that it is individuals who can bring cultural change to organizations and institutions. Kolaczyk agreed that cultural change is key, especially given that the term “data science” is so broad and relevant academic spaces are no longer so well defined. Bray noted the importance of protecting what academia has done well—teach durable skills that outlast changing technologies. As the gap between theory and practice begins to close, however, and undergraduate programs and opportunities change drastically, he wondered whether master’s programs will still be needed. Deborah Nolan, University of California, Berkeley, remarked that master’s-level programs offer a deeper dive into the theory and methods previously learned and will adapt accordingly as the undergraduate programs change. She emphasized the need for undergraduate faculty to continue to focus on teaching fundamentals instead of emerging technologies. Börner suggested defining and surveying “timely knowledge” and “forever knowledge” for certain courses, as well as “theory” and “practice,” to provide guidance for developing new curriculum. Hero noted that while there are universities conducting these learning analytics, and balancing their use with student privacy considerations, others still rely on intuition and anecdotal evidence for course development.

On behalf of her discussion group, Nina Mishra, Amazon, discussed the values and challenges of project-oriented curricula. Her group emphasized the importance of understanding the purpose for incorporating student projects into a curriculum: is the goal to prepare the students for industry jobs or to teach them how to use data to gain deep insight? She suggested that faculty avoid tailoring projects too closely to industry today, as most employers want to hire “big thinkers” who can solve tomorrow’s problems. Students may be most successful if the project allows them to develop the analytic skills needed to work on future data projects. Mishra also explained that there is a spectrum of projects that serve different purposes, and those that require low student-to-faculty ratios may be difficult to scale. She also highlighted that it can be difficult to access company data for student projects, which can be a concern because students may not be as excited by the alternate option of working with public data. Mishra wondered if collaborating more closely with industry or searching for new resources could alleviate this constraint. Lastly, Mishra noted the importance of carefully scoping the project problem with students so that there is a concrete question to be answered. This, in addition to active engagement from participating companies, can improve project outcomes. Hridayesh Rajan, Iowa State University, suggested that because access to alumni networks and industrial partners (and thus projects) is limited on some campuses, it would be helpful if an open resource of projects were available to all institutions.

Kolaczyk noted that it is important to balance what industry wants and what students need. For example, when students persist about learning how to use a particular software package, it is the responsibility of the faculty to shift their mindsets by explaining that all of the technologies will change and that there are multiple languages in which to communicate data. According to Kolaczyk, some of this cultural change can happen through facilitated group-based self-learning. Ziganto noted that there are only so many “big thinkers,” and people who can make smaller changes are also essential—what is most valuable to employers is a student with the right fundamental knowledge to be able to learn quickly and adapt to new situations.

ABOUT THE ROUNDTABLE: The Roundtable on Data Science Postsecondary Education is supported by the Gordon and Betty Moore Foundation, the National Institutes of Health Big Data to Knowledge program, the National Academy of Sciences W. K. Kellogg Foundation Fund, the Association for Computing Machinery, and the American Statistical Association. 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.

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 fourth meeting on October 20, 2017. 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.

ROUNDTABLE MEMBERS PRESENT: Eric Kolaczyk, Boston University, Co-Chair; Kathleen McKeown, Columbia University, Co-Chair; Ron Brachman, Cornell Tech; Alok Choudhary, Northwestern University; James Frew, University of California, Santa Barbara; Alfred Hero, University of Michigan; Nicholas Horton (via webcast), Amherst College; Mark Krzysko, U.S. Department of Defense; Chris Mentzel (via webcast), Gordon and Betty Moore Foundation; Nina Mishra, Amazon; Deborah Nolan, University of California, Berkeley; Antonio Ortega, University of Southern California; Victoria Stodden, University of Illinois, Urbana-Champaign; Mark Tygert (via webcast), Facebook Artificial Intelligence Research; and Jeffrey Ullman, Stanford University.

GUESTS PRESENT: Katy Börner, Indiana University; Andrew Bray, Reed College; Catherine Cramer, New York Hall of Science; Abhijith Gopakumar, Northwestern University; Dan Nicolae, University of Chicago; Michelle Paulsen, Northwestern University; Hridesh Rajan, Iowa State University; Karl Schmitt, Valparaiso University; Stephen Uzzo, New York Hall of Science; Nicholas Wagner, Northwestern University; and David Ziganto, Metis.

STAFF PRESENT: Linda Casola (via webcast), Janki Patel, Michelle Schwalbe, and Ben Wender.

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