



December 2018

## Roundtable HIGHLIGHTS

### Roundtable on Data Science Postsecondary Education

Meeting #8 - September 17, 2018

The eighth Roundtable on Data Science Postsecondary Education was held on September 17, 2018, at the Georgia Institute of Technology in Atlanta, Georgia. Stakeholders from data science education programs, government agencies, professional societies, foundations, and industry convened to discuss existing efforts in computing, statistics, and mathematics societies to improve core fields' engagement with underrepresented populations and to learn about several new programs focused on broadening participation in data science. 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](https://nas.edu/DSERT).

Welcoming Roundtable participants, co-chair Eric Kolaczyk, Boston University, described a profound lack of representation of women and minorities in science, technology, engineering, and mathematics (STEM) fields. He noted that tremendous challenges and opportunities exist to improve equity and diversity in STEM education programs and workplaces. He suggested that the emergence of data science, with its focus on new paradigms, has the potential to create a watershed moment to better engage women and minorities in STEM fields and beyond. The presentations and discussions that followed detailed best practices and possible strategies for creating opportunities in STEM for underrepresented populations.

#### THE CONSTELLATIONS CENTER FOR EQUITY IN COMPUTING

Kamau Bobb, Georgia Institute of Technology

Bobb, along with Charles Isbell, Georgia Institute of Technology (Georgia Tech), developed the [Constellations Center for Equity in Computing](#) in an attempt to address some of the structural challenges that students, particularly students of color, experience both in the city of Atlanta and throughout the nation. Despite the fact that computer science skills are central to decision making in a modern digital economy, Bobb noted a dearth of computer science educators in both the K–12 and postsecondary spheres—for example, Georgia has more than 528,000 students enrolled in public high schools but only 93 teachers certified to teach computer science. With low teacher pay, limited professional development opportunities, and industry pull for recent college graduates, this educator shortage will likely continue even as student interest in computer science education increases, he explained. In response, the Constellations Center built a structural tool to deliver computer science content through a hybrid infrastructure: skills are delivered online and classroom teachers facilitate learning. This model has the potential to increase *equitable* access to computer science skills for minority and low-income students.

Bobb described inequities in access to computing education and their impacts on undergraduate enrollments of underrepresented minorities. While Atlanta's population is more than 50 percent African American, only 3 African American students are enrolled in advanced placement computer science courses in local public high schools, and this population is similarly underrepresented in Georgia Tech's College of Computing, according to Bobb.

This year, three fellows from the Constellations Center are going into six public high schools in Atlanta to teach Advanced Placement Computer Science Principles to the students, while the classroom teacher observes. In the future, a virtual course will take the place of the fellow, and the classroom teacher will facilitate. Scale is the most challenging aspect of this model since it is impossible to deploy fellows to all schools; however, Bobb noted that the Constellations Center's work continues to receive support from the National Science Foundation (NSF) and various independent organizations.

Victoria Stodden, University of Illinois, Urbana-Champaign, asked about next steps for research and resources as well as how this problem of access relates to data science specifically. Bobb responded that the dominant problem for students of color is access to higher education in general; by prioritizing access to computational skills in particular, students will be exposed to data science and able to pursue any number of computational-type fields in college and beyond. Jeffrey Ullman, Stanford University, said that the number of students who took the 2018 Advanced Placement Computer Science Exam had increased substantially, including those in rural areas. He wondered if the problem of access is being resolved across the country. Bobb replied that while increases in the numbers and types of students taking the exam are important achievements, there is still much progress to be made in terms of the numbers and types of students passing the exam. He explained that the subject matter is still not being deployed at even a minimal level in many parts of the United States. Renata Rawlings-Goss, South Big Data Hub, asked how teachers are selected for participation in the hybrid program. Bobb said that his team currently asks local principals to suggest teachers with the interest and the aptitude. Another avenue involves identifying teachers who lead courses in Georgia's [Career, Technical, and Agricultural Education](#) infrastructure's computer science and information technology pathway. During a later discussion, Uri Treisman, University of Texas, Austin, posed Bobb's hybrid approach as a public policy question: Is it a public good, and, if so, who should pay for it?

## **PANEL PRESENTATIONS ON EXISTING PROFESSIONAL SOCIETY EFFORTS TO INCREASE DIVERSITY**

### **Student-Centered Interventions to Retain Women, Underrepresented Minorities, and Persons with Disabilities in Computing**

*Ayanna Howard, Georgia Institute of Technology and Computing Research Association*

Before beginning her presentation, Howard mentioned that the [Computing Research Association—Women](#) (CRA–W) will soon change its name and mission statement to include all underrepresented populations, including persons with disabilities. She showed a brief video of CRA's 2018 [graduate cohort for underrepresented minorities and persons with disabilities](#) (URMD). The 2018 URMD cohort enrolled 90 people from 60 institutions, all of whom were sponsored. Howard noted that CRA will host another URMD cohort in March 2019 and a graduate cohort specifically for women in April 2019. All of CRA's programs rely on the cohort model, which incorporate opportunities for participants to learn both from one another and from senior-level mentors. CRA offers two undergraduate programs—the [Distributed Research Experience](#) and the [Collaborative Research Experience](#)—both of which include student research, stipends, and mentorship. CRA also hosts [Discipline-Specific Workshops](#), [Distinguished Lecture Series](#), and [Virtual Undergraduate Town Hall](#) events.

**Celebrating Women in Statistics and Data Science: Goals, Creation, Implementation, and Outcomes**  
*Dalene Stangl, Carnegie Mellon University and American Statistical Association Committee on Women in Statistics*

Motivated by the words of Susan Ambrose and Barbara Lazarus at Carnegie Mellon in 1992—that traditional pedagogical approaches emphasizing male patterns of behavior have restricted teaching and learning for women—Stangl and a team of women in STEM at Duke University committed to “disrupting the hierarchy.” In particular, Stangl's participation in the [Grace Hopper Conference on Women and Computing](#), which today attracts more than 20,000 female participants annually, illuminated the different educational and professional experiences of men and women. With the help of a \$10,000 grant from the American Statistical Association (ASA), Stangl initiated “Celebrating Women in Statistics and Data Science,” which gives women a “place to learn, understand, and voice what they value whether it agrees with or goes against a mainstream work culture.” The group hosted its first [Women in Statistics and Data Science conference](#) in 2014; ASA has now taken over hosting this annual conference, offering technical presentations, professional development, and networking opportunities for those new to the field and those with more experience.

## Collaboration, Cohorts, and Comfort Zones: The Three Cs of Community

*Ami Radunskeya, Pomona College and Association for Women in Mathematics*

Radunskeya said that although women have made progress in terms of representation in mathematics, more work is needed. The [Association for Women in Mathematics](#) (AWM) supports women and girls all along the pipeline through programming and with the assistance of 200 volunteers. AWM's programs for middle and high school girls include essay contests, mathematics days, and mentorship, and more than 200 AWM student chapters are located on college campuses across the country, she explained. For women who are more advanced in their careers, AWM offers travel grants, semiannual conferences, workshops, prizes, and distinguished lectureships. AWM also partners with NSF's [ADVANCE program](#) on career advancement for women through research-focused networks. AWM's goal, according to Radunskeya, is to increase recognition of women at all levels with tiered mentoring and supportive collaboration—20 collaboration networks have already been established. Radunskeya has also been involved for 20 years with the [Enhancing Diversity in Graduate Education](#) (EDGE) program, a comprehensive mentoring program to encourage women to stay in graduate mathematics programs. EDGE offers a summer immersion program, online mentoring, “difficult dialogues” sessions, support for research and travel, summer symposia, and regional clusters, and participants have become leaders in their fields across the country. Radunskeya reiterated the value of a cohort program such as EDGE in forming large networks of women.

## PANEL DISCUSSION

### Initiatives for Data Science

In response to a question about whether the initiatives presented by the panelists could be duplicated for data science, Radunskeya remarked that they should be replicated for data science because data science is computing, modeling, and solving problems. She added that early collaborations with industry would be particularly useful in data science, along with mentorship opportunities. She also mentioned that the NSF [Inclusion across the Nation of Communities of Learners of Underrepresented Discoverers in Engineering and Science](#) (INCLUDES) initiative is designed to enhance U.S. leadership in STEM discoveries and innovations by focusing on broadening participation in these fields at scale. Across these various programs, Radunskeya noted

that mentoring is repeatedly described as essential. Howard said that many of the existing initiatives could be duplicated in data science programs.

### Investment and Research Strategies

Panelists were asked what initiatives they would like to implement if resources were unlimited. Stangl said that the social stratification problems in elementary and high schools should be addressed first. Howard emphasized the need for time resources, in addition to financial resources, especially at the K–12 levels and for students at under resourced postsecondary institutions. Radunskeya noted the value of dedicating time and financial resources to middle school programs, camps, 1-day events, and other partnerships to motivate children to study STEM and emphasized the importance of respecting the people involved in organizing such programs. She also suggested that funding be allocated to research experiences for undergraduates that intentionally engage underrepresented minorities.

Stodden wondered if a more established research agenda would help prioritize issues of diversity and accessibility. Treisman noted that high-quality research already exists (e.g., [a 2015 \*Frontiers in Psychology\* paper](#) on the underrepresentation of women in STEM fields) and that the next step is for such research to inform classroom practice. He emphasized the need for a systemic, institution-wide approach to issues of inequity and injustice instead of simply having a few people offer useful programming. Radunskeya agreed that excellent research is available but that many practitioners often do not understand the research, how to implement it, or how to engage faculty in relevant professional development experiences. Brandeis Marshall, Spelman College, agreed with Treisman about the need to overcome the isolation of programs and singular points of advocacy. She explained that initiatives should be discussed with all relevant audiences (e.g., parents of elementary and high school students should be invited into conversations about the value of computational thinking for their children). She also suggested an investment beyond the academic institutional environment—for example, if the media shows women of color and people with disabilities working in computational fields, participation may increase in those areas.

### Underrepresented Populations in STEM

In response to a question about issues facing women in STEM today, Stangl stated that many of the problems that existed 25 years ago remain widespread. She said that structural changes (e.g., more flexible teaching) would better accommodate the different



ways in which individuals learn, thus broadening participation in STEM communities. Treisman said that, in the past 20 years, some fields have experienced a dramatic increase in the numbers of women earning PhDs (e.g., molecular biology), while others remain lacking (e.g., physics). He wondered whether cultural features of disciplines have generated the dramatic shifts and if lessons learned could be leveraged for other fields. Stangl noted that the field of life sciences seems to have more growth for women than non-life sciences, and she added that the higher the percentage of women in departments, the lower the pay. Radunskaya observed that the breakdown of mathematics publications suggests that women are more interested in areas of study that allow for collaboration. This bodes well for gender diversity in data science, she continued, because data science is an inherently collaborative field. She also emphasized the need to abandon the myth that a field such as mathematics requires innate ability, as that is another deterrent to broad participation. Emily Fox, University of Washington, added that stratification exists even within fields (e.g., the number of women studying computational neuroscience is abysmal, while in neuroscience female representation is strong). She speculated that women often enter emerging fields later than men, which may contribute to their initial underrepresentation.

Referring to Stangl's comments about structural issues in the education system, Fox asked the panelists what universities could be doing on a regular basis to address these inequities. Howard said that a university's response should depend upon its demographic of interest. She encouraged academic institutions to offer family leave and to include bias training in faculty hiring initiatives. She also emphasized the need to engage students at the undergraduate level in computing by offering various flavors of introductory computer science courses. Radunskaya noted that the EDGE program model could be used in both undergraduate and graduate programs, and she championed the "Uri Treisman model" for cohort building at the undergraduate level. Stangl said that flipped classrooms are effective for undergraduates, though such an innovation may be difficult at large public universities with fewer resources per student. Ron Brachman, Cornell Tech, wondered if inclusive strategies for persons with disabilities are different from those for members of other underrepresented populations. Howard said that being an underrepresented person is a shared quality. In terms of best educational practices, she emphasized accessibility at the postsecondary level, and she encouraged participants to sign the "Computer Science for All"

[Accessibility Pledge](#) to make computer science materials more accessible.

## OPEN DISCUSSION

### Diversity in the Professoriate

Isbell described [Diversifying Future Leadership in the Professoriate](#) (FLIP), a consortium working to change the process for graduate school admittance into computing programs and to improve representation of minorities in the professoriate. Kathy McKeown, Columbia University, said that representation in the professoriate influences institutional changes and mentorship opportunities, and she supported Isbell's commitment to fixing problems of underrepresentation in the professoriate. Over the past 5 years, Columbia increased the number of women in leadership positions, specifically in the School of Engineering, which led to changes in hiring processes. In response to a question from Brachman about the proportion of underrepresented graduate students who become faculty, Isbell noted that biases drive decisions about whether students are motivated to pursue faculty positions. He cited that 65 percent of faculty at the top 10 institutions have earned their PhDs from those top 10 institutions, while 100 percent of African American faculty at the top 10 institutions have earned their PhDs from those top 10 institutions. Isbell explained that the way in which people are pushed down the pipeline is flawed, and risk-averse faculty hiring practices place underrepresented minorities at a disadvantage.

### Data Science for K–12 and Postsecondary Students

Nicholas Horton, Amherst College, wondered about the extent to which data science initiatives could be implemented at the K–12 levels in order to improve participation at the postsecondary level. Rawlings-Goss said that there is abundant opportunity to be involved in the K–12 space. She said that this conversation should involve data scientists who can help create a people-driven solution informed by data. McKeown suggested asking capstone students to work on some of the data-rich city problems that Bobb discussed, and Rawlings-Goss proposed including high school students in these capstone experiences as a way to introduce them to data science and as an opportunity for mentorship. Rachel Levy, Mathematical Association of America, noted that the newness of data science in K–12 could allow teachers to reimagine themselves as mathematics doers and statistical thinkers and to help them empower their students to develop computational thinking skills. She added that engaging with students who have different kinds of learning abilities could reveal new,

more accessible strategies for teaching all students more effectively. Marshall said that understanding how to get students involved in computational thinking is an ongoing conversation throughout the community. She emphasized the value of considering all individuals, not just underrepresented individuals, in order to take socially responsible actions in education. Treisman noted that high school students comprise 25 percent of the student body at 40 percent of the community colleges in Texas. No standard offering exists for such dual enrollment in mathematics, he continued, and that creates a space for the introduction of data science.

## **KEEPING DATA SCIENCE BROAD**

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

Rawlings-Goss explained that the four big data hubs are part of an NSF initiative to bring together academic, industry, and government researchers and practitioners in the data science space for the benefit of U.S. economic and social well-being. She noted that 563 data science programs exist at the undergraduate and graduate levels in academic institutions across the United States. The South Big Data Hub's "Keeping Data Science Broad" project encompassed three webinars and a workshop with 60 participants. Workshop participants included faculty from historically minority-serving institutions, community colleges, and 4-year liberal arts schools interested in creating data science programs, as well as representatives from government and industry. The project's first webinar featured speakers from campuses that already have data science programs and focused on structural topics such as whether to require prerequisites, tips to build successful programs, and strategies to train different types of data scientists. The second webinar focused on alternative avenues to data science education such as industry programs, museum experiences, and academic programs outside of traditional STEM disciplines.

A [written consensus report](#) including challenge topics, visions for the future, top asks, and next steps for introducing data science emerged from the project's activities. The report was released in January 2018, followed by the project's final webinar. Challenges and visions discussed in the report were categorized in terms of (1) access to data; (2) assessment and evaluation; (3) curriculum; (4) data literacy; (5) diversity, inclusion, and equity; (6) ethics; (7) faculty, staffing, and collaboration; and (8) the pipeline to higher education. Rawlings-Goss commented that discussions on diversity, inclusion, and equity, in particular, revealed that a one-size-fits-all solution does

not apply in all academic communities. The report revealed that implicit bias training for faculty, staff, and institutions; culturally relevant, high-quality curricula; and respect for the role that 2-year institutions, minority-serving institutions, and K–12 schools play in program development are essential.

Rawlings-Goss also described the DataUp program (hosted by the South Big Data Hub in partnership with The Carpentries), which offers introductory "train the trainers" workshops. In these workshops, participants develop skills that will be useful for training their academic colleagues. The South Big Data Hub hosts the Data Science for Social Good program as well, in which graduate students work with undergraduates on local problems. Rawlings-Goss hopes that these programs will increase inclusivity and diversity in the field of data science.

Kolaczyk asked about follow-up and quality control measures for the DataUp program, and Rawlings-Goss explained that because the program is only in its first year, evaluation is still evolving. She said that upfront training reduces quality control issues and added that assessment will be conducted to understand how this program ultimately affects the trainers' institutions. Treisman turned the focus of the conversation to how to scale such efforts, and he emphasized the need to think beyond simply spreading programs. Instead, a scaling framework used in mathematics that could be helpful for data science programs includes four dimensions: spread, depth, and ownership of the program, as well as normative changes in policy and practice. He emphasized the criticality of a shift in ownership. He also highlighted the notion that "diversity," "inclusion," and "broadening participation" mean different things to different people, which can make achieving the desired social justice implications of educational programming difficult. Accordingly, Rawlings-Goss wondered about whether the DataUp program should focus on spreading to more institutions or increasing the depth at institutions currently involved.

## **THE DSX PROJECT: A FIRST LOOK AT DATA SCIENCE EDUCATION ON SPELMAN AND MOREHOUSE CAMPUSES**

**Brandeis Marshall, Spelman College**

Marshall described the Data Science Extension (DSX) Project, which is funded by NSF, as a 3-year targeted infusion project between Spelman and Morehouse Colleges—both of which are private, minority-serving, baccalaureate, liberal arts institutions that have mathematics and computer science departments but

do not have statistics departments. DSX focuses on faculty and their impacts on students through curriculum, with the objectives of (1) sharing the power of data in context and (2) increasing access to and participation in data science practices for Spelman and Morehouse students. Marshall explained that DSX embeds one or two data science concepts into the existing curriculum. Faculty meet for 2 weeks in the summer and monthly throughout the year, giving them the time and space to consider the connection between their disciplines and data science. Faculty training revolves around interdisciplinarity, competency-building, and knowledge transfer to students.

Marshall explained that the project is challenging because it requires in-house faculty development in data science (which is especially difficult at small institutions where faculty are already overburdened), technological and computing infrastructure, availability of relevant course offerings for students, and sustainability planning of courses. She described the project's benefits for students as exposure to data science in the core, cognate, and elective courses during sophomore, junior, and senior years; applicability to a variety of disciplines; and incentives to examine career opportunities in data science.

Ullman asked if data processing could be infused into any of the courses, and Marshall replied that it depends on the course. Some courses might integrate units on data ethics or data storytelling, while others are more hands-on (e.g., an environmental science course integrated a unit on data processing). She noted that faculty do not make any assumptions about their students' previous experiences with coding or computing, and she added that many tools exist to help with those aspects of data science. In response to a question from Kolaczyk, Marshall explained that this model is still a pilot, so it will continue to be evaluated and the measures of success will remain varied (e.g., whether the faculty training is effective, how many students are impacted, whether the infused content is valuable in the course). In response to a question from Jessica Utts, University of California, Irvine, Marshall said that the materials from the project's faculty retreats will be posted publicly in mid-2019.

McKeown asked which fields Spelman and Morehouse students pursue after completing their undergraduate degrees. Marshall said that approximately 50 percent attend graduate school. In response to a question from Kolaczyk about the management style of the project, Marshall explained that faculty par-

ticipants use the Piazza platform during the summer retreats and rely on email, meetings, seminars, and small-group conversations during the academic term. Rawlings-Goss asked about differences in the infusion process at Spelman and Morehouse, and Marshall said that she has not yet observed any differences but will continue to process the data. Treisman asked if Marshall has surveyed the faculty and students on their levels of comfort with data science tools such as Python, and Marshall noted that the vast majority of incoming students at Spelman arrive without any computational knowledge.

## **HISPANICS AND NATIVE AMERICANS IN COMPUTER SCIENCE: PATTERNS, PRESSURES, AND PROGRAMS**

**Lydia Tapia, University of New Mexico**

Tapia showed a series of graphs from [CRA's Taulbee Survey](#) to demonstrate that Hispanic and Native American students continue to be significantly underrepresented in computer science bachelor's, master's, and doctoral programs. In addition, the rate of change for the degree production is not matching the rate of change for the population of the Hispanic community. She noted that 15 years ago, fewer and fewer Hispanic students were in each stage of the computer science pipeline (i.e., bachelor's degree through full professorship). Ten years ago there were only small gains, and, 5 years ago, both small increases and decreases were evident at various stages of the pipeline. Overall, she stated that this demonstrates that not enough progress has been made for underrepresented populations in computer science.

Tapia provided an overview of her educational path to becoming a faculty member at the University of New Mexico, which began with a supercomputing challenge in high school, included an internship at Sandia National Laboratories with continued mentorship, and concluded with a doctoral degree in computer science from Texas A&M University. She noted that members of underrepresented groups often lack technology resources, endure pressures to stay close to home after high school and to contribute to the family (sometimes financially), lack an understanding of graduate school, and experience difficulties with travel. With all of these challenges, Tapia continued, intervention programs throughout the pipeline, starting as early as kindergarten, are crucial. Successful programs at the K–12 levels include the [New Mexico Supercomputing Challenge](#), which is an expo of student computing projects with mentorship opportunities; the [Tapia Lab Demos](#); and the [New Mexico](#)



[CS4All program](#), which trains students in dual enrollment classes and trains teachers who will be working with those students. At the undergraduate level, the [NASA Swarmathon](#) and the [Robot Guru](#) both engage underrepresented students in computer science. The CRA–W URMD grad cohort, discussed by Howard, is beneficial for graduate students, Tapia added. At the faculty level, career mentoring workshops and proposal writing workshops are two methods to improve retention in the field. Moving forward, it is important to consider additional ways to increase participation at all stages of the pipeline.

Marshall asked Tapia why a drain exists at every level of the pipeline for a variety of demographics. Tapia replied that while encouragement to pursue a bachelor's degree may be strong in Hispanic communities, for example, motivation to attend graduate school is lower since such a degree is not required to secure employment. This explains the first significant drop in the numbers on the pipeline, although the larger drop occurs at the PhD level. Tapia believes that mentoring is the best way to overcome that gap. Kolaczyk asked if mentoring could be expanded to better address local cultural considerations, and Tapia responded that although families may be reluctant to listen to advice from a stranger, it could be valuable for students' mentors to interact with their families. Treisman later cautioned against stereotyping students and viewing them through a deficit-focused lens, as many students came from families who support their educational aspirations.

Radunskaya noted that minorities often carry a larger faculty service burden, and she wondered how successful Tapia has been at convincing her colleagues to support her in these endeavors. While she has supportive colleagues, Tapia acknowledged that this is a challenge and that sometimes service is downplayed relative to research in tenure reviews. McKeown commended Tapia for highlighting this pervasive struggle in the tenure process, and she wondered if it is time to rethink the notion of "credit" in academia. Levy agreed that work that is not "publishable" often does not receive the credit it is due and suggested that the Roundtable continue a conversation about new approaches to recognizing important work. Treisman offered his support for such a conversation.

## **SMALL GROUP DISCUSSIONS AND CONCLUDING CONVERSATIONS**

Participants divided into two groups to discuss issues of diversity and access in greater detail. The

first group focused its discussion on leveraging programs that have been successful for underrepresented populations in other fields. On behalf of his group, Kolaczyk reported that because "data science" is an amorphous term, it can prove challenging to recruit undergraduates into the field. Thus, he continued, it is imperative that students know what data science opportunities exist and which skill sets are needed for particular career paths. He added that it could be possible to leverage the efforts of a field such as neuroscience, which has been successful in recruiting and retaining women, by explaining that data science overlaps with that particular field. He also described the University of California, Berkeley's introductory data science course, Data8, which requires no prerequisites, enrolls approximately 2,000 students each semester, and is complemented by "connector courses" in various disciplines. Data8 participants are exposed to valuable technical skills even if they choose not to follow a technical career path. Regarding professional organizations, Kolaczyk said that they could promote data science activities of interest to students (e.g., ASA's DataFest) and could fund regional events during which students could learn more about various data science careers and the training they require. He suggested that data science organizations collaborate with organizations that are actively engaging women and minorities in other fields. Kolaczyk's group also discussed the importance of scaling up mentorship opportunities. As an example, ASA works directly with high school counselors to encourage early participation in data science activities. He added that the student chapters of ASA or AWM could host networking events or organize panels of faculty, undergraduates, and graduate students to provide information to high school students.

The second group focused its discussion on promising opportunities for investment if resources were unlimited. On behalf of his group, Brachman said that it is critical to consider the current research and to evaluate the largest potential marginal payoff before making funding decisions. Resources could be dedicated to making improvements along the pipeline by first engaging the media to better portray diverse individuals in technical fields. A partnership with an organization such as the Geena Davis Institute on Gender in Media may present interesting opportunities, he continued. Once people are attracted to the field of data science, Brachman explained, they need to be given technical resources and creative opportunities to learn before they enter high school. He added that students have the potential to become most engaged in data science if they are presented

with choices for both school and extracurricular programming. He noted that PK–12 teachers could be given stipends, perhaps from industry, to participate in data science initiatives and to develop innovative educational programs. At the postsecondary level, scholarships would assist students who do not have the means to complete their degrees, although complications may arise if these funds come from companies that have expectations for the students after graduation. Brachman’s group also discussed the need to balance mentorship and sponsorship as well as the importance of resource allocation toward spon-

sorship and collaboration opportunities. He added that collaboration should be built into curricula to prepare students for the work that awaits them in the field of data science. Overall, he continued, it is important to think about ways to revise curriculum and pedagogy to encourage cross-disciplinary work and incorporate studies of data ethics. This includes increased attention to faculty training in new tools and resources, perhaps during summer institutes. Industry could also be involved in departmental reviews so that faculty can improve the way they train students for future industry jobs.

---

**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 eighth meeting on September 17, 2018. 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; Emily Fox, University of Washington; Lise Getoor, University of California, Santa Cruz; Nicholas Horton (via webcast), Amherst College; Charles Isbell, Georgia Institute of Technology; Mark Krzysko, U.S. Department of Defense; Rachel Levy, Mathematical Association of America; Brandeis Marshall, Spelman College; Antonio Ortega, University of Southern California; Victoria Stodden, University of Illinois, Urbana-Champaign; Uri Treisman, University of Texas, Austin; Jeffrey Ullman, Stanford University; and Jessica Utts, University of California, Irvine.

**GUESTS PRESENT:** Sylvia Ankrah; Ginger Baxter, Emory University Goizueta Business School; Kamau Bobb (via webcast), Georgia Institute of Technology; Cheryl Brown, University of North Carolina at Charlotte; Cynthia Bryant, Georgia Institute of Technology; Gregory Chambers, Independent Contractor; Stephanie Espy, author; Blake Fleisher, Anidata; Lisa Gervin, LG Technical Devices; Ayanna Howard, Georgia Institute of Technology; Tremayne Jackson, Tesla; Vivian Lyon, Plaza Dynamics; Nancy Murray, Emory University; Rebecca Nugent, Carnegie Mellon University; Ami Radunskaya, Pomona College; Renata Rawlings-Goss, South Big Data Hub; Alicia Richhart, Georgia Institute of Technology; Dalene Stangl, Carnegie Mellon University; Kendra Strickland, Georgia Institute of Technology; Lydia Tapia, University of New Mexico; and Charlie Wright, Georgia Institute of Technology.

**STAFF PRESENT:** Linda Casola (via webcast), Janki Patel, 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](http://www.national-academies.org)