August 2017

Roundtable
HIGHLIGHTS

Roundtable on Data Science Postsecondary Education

Meeting #3 - May 1, 2017

The third Roundtable on Data Science Postsecondary Education met on May 1, 2017, at the Pew Research Center in Washington, D.C. Stakeholders from data science education programs, funding organizations, government agencies, professional societies, foundations, and industry convened to discuss data science training in the workplace. 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**.

DATA SCIENCE TRAINING IN THE WORKPLACE: GOVERNMENT

Practicing Data Science in the Government

Ron Prevost, U.S. Census Bureau

Prevost explained that data produced by the U.S. Census Bureau are expected to be unbiased, statistically accurate, delivered quickly at low cost, useful to determine causality, reproducible, transparent, and protected. While striving to meet these expectations, statistical agencies confront many challenges, including greater than expected costs and lower than expected response rates for surveys, complex information requests, competition among data products and questions of product validity, new data sources and methodologies, and policy requirements.

The Census Bureau hopes to supplement survey data with data that have been repurposed from other sources. However, this data integration needs to be transparent and reliable, utilize quality measures, and ideally incorporate model-based estimation and data source acquisition and integration processes. In his view, the Census Bureau could advance this paradigm shift by taking several critical steps, including (1) consolidating business processes and systems and generalized solutions, (2) supplementing current business processes with new processes, (3) developing new products, (4) building new capabilities, and (5) optimizing current business processes. Prevost noted that institutional, bud-

getary, and security barriers can limit this type of large-scale transformation. For example, to address this transformation thoroughly, staff in information technology departments would have to learn new processes for managing, curating, and using data and metadata; to ensure that new software complies with government security protocols; and to organize, explore, and test real data in a collaborative environment often referred to as a "sandbox."

Many federal agencies are also exploring how increased opportunities for interdisciplinary teamwork and professional development could better equip employees for work that requires new computing techniques and new methodologies. The Census Bureau is evaluating program use cases to determine which skill sets will be needed by both current and future employees, as well as how projects will be funded. According to Prevost, current knowledge gaps include data science, business/data analytics, reproducibility, software design and engineering, data storage and retrieval models, and operations research. After extensive investigation, the Census Bureau created a catalogue of 80 programs (with a total of 600 courses) located in or near Washington, D.C., or available online, that offer degrees, certificates, or short courses in the needed content areas. Massive Open Online Courses (MOOCs) may be a cost-effective alternative or complement to these more traditional training programs because they address specific agency needs quickly and flexibly.

Nicholas Horton, Amherst College, emphasized the need to train employees to understand the unique advantages and disadvantages of using different types of data in their work and encouraged the Census Bureau's emerging emphasis on fusion of found and designed survey data. He noted that issues of data ethics and cybersecurity are crucial areas for employee training. Jeffrey Ullman, Stanford University, suggested that there is a disconnect between training and real-world problems that could be eliminated with further development of core computer science skills. Prevost agreed that skills gaps exist but noted that the program use cases explored thus far were focused more on training for research analysts than for information technology specialists.

Patrick Perry, New York University, asked for clarification on what is driving the transition to model-based estimation and inference and whether new training is necessary to apply this type of methodology. Prevost responded that the Census Bureau sought new approaches to improve legacy products, given declining survey response rates and questions about bias in these products. He added that while the Census Bureau does provide training in big data and statistics to its employees, much of the current training is in project and budget management. Jordan Sellers, Howard University, suggested that the Census Bureau take the lead in establishing a professional development policy; however, Prevost noted that a formal, standardized training policy may not be effective because staff training evolves around mission-critical activities and rapidly changing technologies. Victoria Stodden, University of Illinois, Urbana-Champaign, asked how people can track the provenance of Census Bureau data sets. Prevost stated that all Census Bureau data products undergo numerous quality measurements related to collection methods, variance, and benchmarks. To learn more about Census Bureau data products, and how they compare to other data products, he recommended researchers visit the **Federal Statistical Research Data Centers** located throughout the United States.

Training Government Employees in Data Science
Drew Zachary, U.S. Department of Commerce

Zachary noted that developing creative training initiatives is essential for federal agency managers who have limited funding or authority to offer education programs or hire new staff. When evaluating how to bring together the right set of data science skills, two models are useful for employees and managers to consider: (1) a "unicorn" model, in which one employee has all of the skills needed to complete a task, or (2) an "X-men" model, in which people with diverse skills work together to complete a task.

The **Commerce Data Academy** is an internal upskilling data science education initiative that relies on the Commerce Data Service, as well as extra-governmental instructors from organizations including General Assembly and Data Society, to train Department of Commerce colleagues in data science, data engineering, and web development skills. After training more than 1,500 Department of Commerce employees in 35 courses (both online and in-person) over the past year and a half, the Commerce Data Academy now invites employees from other federal agencies to enroll in its courses on an as-needed basis.

Initiated by the Office of Science and Technology Policy during the Obama Administration, **Fellows in Innovation** reaches programs across government, representing 400 fellows and 30 agency divisions. Zachary noted that this program allows data professionals to apply their often underutilized technical skills to a policy problem, as well as to transfer these

data science skills to their teammates. For example, a team used machine learning, digital mapping, and sentiment analysis to help understand neighborhood data and explore opportunities for economic development in high-poverty communities.

Supported by the General Services Administration, the Federal Data Cabinet creates a "community of practice" for data professionals in government to share best practices and success stories, as well as to discuss challenges faced throughout the data life cycle. One of the working groups within the Federal Data Cabinet, the Data Talent Working Group, plans to create a decision guide to help hiring managers and team leaders assemble teams and choose effective training models to best meet project needs.

Natassja Linzau, National Academies (formerly of the Department of Commerce), emphasized that all of the "teachers" in the Commerce Data Academy are sharing their time and expertise without additional compensation, and the "students" do not pay any fees to take their courses. Ullman wondered if MOOCs could be used in the Commerce Data Academy in the future and whether there may be introductory courses that could be added to the list of offerings. Zachary and Linzau noted that many of their course materials and recordings are available on the Commerce Data Academy website so that anyone who is interested can use them to learn. They also plan to explore using MOOCs as a way to enhance future course offerings.

Louis Gross, University of Tennessee, Knoxville, wondered how a decision is made regarding whether to train employees or to hire consultants to solve particular problems. He suggested that agencies learn how to better use their talent pools, highlighting the Fellows in Innovation program as a good model, and Zachary noted that the Federal Data Cabinet could also serve as a repository for this information. Prevost added that mentorship programs could also be expanded to address this issue. In response to a question from Rebecca Nugent, Carnegie Mellon University, Zachary noted that sustained funding of the program is a concern, as is relating the benefits of the program to fellows' supervisors.

DATA SCIENCE TRAINING IN THE WORKPLACE: BUSINESS

The Technology Sector

Emily Plachy, IBM

Plachy defined data scientists as "pioneers" who solve problems by relying on quantitative training, effective communication skills, business acumen, and various data and analytics tools and programming languages. She noted that data science continues to evolve in response to the era of cognitive computing. Data scientists now need skills in hybrid analytics, streaming data, artificial intelligence, application program interface-based analytical services, and Cloud-based solutions. Data scientists often expect their employers to help them build upon their technical and business skills to keep pace with the evolving field (Figure 1), so Plachy suggested that it may be useful for employers to establish a certification roadmap. Offering workplace data science training not only improves employee performance, but may also increase employee retention, according to Plachy. IBM created a Data Science Profession to encourage data scientists to continue to train and develop their skills; it uses "open badges" that contain metadata representing "skill tags" and accomplishments, both to signal and verify employees' skills and to improve social connections among colleagues.

Data science education opportunities for IBM employees include the following:

- Data Science Bootcamp—New data science employees can develop awareness of various data

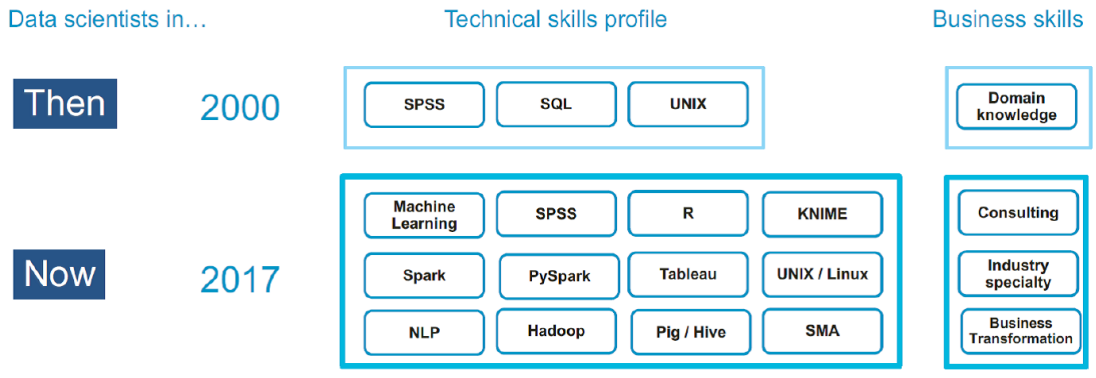

FIGURE 1 *Data scientists will need to expand their technical and business skill sets as the field continues to evolve. SOURCE: Martin Fleming, Chief Analytics Officer, IBM Corp., included in Emily Plachy's presentation to the Roundtable.*

science concepts and form networks with other practitioners over 8 days.

- Data Science Experience—Scientists collaborate in sandboxes, using data analytics to solve problems.
- Big Data and Analytics University—Participants enroll in virtual data science courses at one of three expertise levels.
- Analytics Product Course—Short courses provide overviews of available IBM products.
- Development Activities—Employees select topics for monthly fundamentals courses.
- Analytics Education Series—Employees select from more than 30 1-hour videos of IBM expert lectures on topics such as natural language processing or spatio-temporal analytics.
- Cognitive Academy—Data scientists receive training in areas such as data visualization or machine learning.
- *Analytics Across the Enterprise: How IBM Realizes Business Value from Big Data and Analytics*—Textbook includes 32 case studies of problems solved using data analytics.

Plachy described the Chief Analytics Office, IBM's version of the "X-men" model introduced by Zachary, in which 50-75 people with diverse technical skills form teams to solve business problems within IBM. She said that many recent hires at IBM have knowledge gaps in cognitive computing and the skills to better harness unstructured data to solve business problems; they could benefit from stronger quantitative foundations, better communication skills, and more curiosity and patience. Since data science will continue to evolve, it is unlikely that the conversation about knowledge gaps in data science education will ever end, and IBM may add apprenticeship programs in the future. Plachy suggested that it would be helpful if stakeholders created a public education system for data science where organizations could share ideas for workplace training.

Kristin Tolle, Microsoft, added that, in her organization, experimental design is a major knowledge gap among recent hires. Plachy agreed with Tolle about the importance of training in that area and noted that IBM hires experimental physicists to help colleagues with experimental design and also teaches design of experiments in a six sigma course. In response to a question from Ullman about gaps in current computer science degree programs, Plachy responded that she would like to see more preparation in artificial intelligence, deep learning, and natural language processing.

The Consulting Perspective

Ashley Lanier and Ashley Campana, Booz Allen Hamilton

Lanier and Campana noted that the need to fill knowledge gaps in employee education is not a problem unique to the field of data science. At Booz Allen Hamilton, while employees without data science training need to learn how to use tools efficiently and to analyze and share data, employees with data science specialties need to learn "consulting skills" such as communicating, storytelling, working with clients, working in a team, understanding an audience, and choosing the right approaches.

Because there are unique infrastructure constraints in upskilling employees in consulting firms, Booz Allen Hamilton offers a variety of education programs to its employees, all of which include essential training in teamwork and presentation skills:

- Data Science Bowl—Approximately 2,000 teams from around the world participate in this 90-day online hackathon for social good.
- Tech Tank—Similar to a master's certificate program, with a math and a computer science track, 160 hours of training over 12 months are offered to employees with a scientific background and within 2 years of hire, upon nomination from a supervisor. In addition to technical training, participants receive training (based on personality test results) in communication skills and mentoring. Participants pitch to leaders acting as clients and work on a real problem during an apprenticeship.
- Internship Program ("Summer Games")—Approximately 300 undergraduate interns work on STEM-focused problems and pitch to Booz Allen Hamilton leadership over a 9-week session.
- Data Science 5K Challenge—Similar to Tech Tank, except that training is delivered by an external vendor instead of by Booz Allen Hamilton leadership. This allows more people to participate at the right level and helps the company to increase the total number of data scientists on staff.

Booz Allen Hamilton also offers a data science book club, Yammer groups, bi-monthly Hackathons, a distinguished speaker series, occasional boot camps, and a workshop series as additional, flexible ways for employees to become more engaged in data science. In response to a question from Stodden

about additional data science problems that Booz Allen Hamilton interns and employees have helped clients better understand or solve, Lanier and Campana highlighted the following projects: (1) applying analytics to cardiology to assess heart function, (2) using data analytics to increase adoption rates at animal shelters, (3) employing network analysis to better understand human trafficking in the United States, and (4) using data analytics and technology to help houses go off the electric grid. Gross asked if the education programs at Booz Allen Hamilton have been formally assessed and whether those results have been published. Booz Allen Hamilton tracks billable hours, promotion, and retention of its Tech Tank participants to demonstrate the program's value, but that information is not shared externally. These assessments have also revealed that participants are more incentivized by the opportunity to make a difference solving real problems using real data sets than by the opportunity to earn social media "badges" or prize points for their work. Gross suggested that Booz Allen Hamilton publish future assessment results, as doing so could aid the larger data science community in its development of training.

Nugent encouraged increased collaboration between companies and universities, especially in terms of student skill assessments, so that companies are hiring the best-suited employees. In response to a question from Deborah Nolan, University of California, Berkeley, about the timeline for skill cultivation, Lanier noted that new hires start developing communication, leadership, and presentation skills immediately. Doing so also helps determine with which projects new employees should be aligned. In response to a question from Ullman about gaps in current computer science degree programs, Lanier responded that she would like to see more preparation in machine learning and presentation skills. William Finzer, Concord Consortium, asked if the emerging field of data science education research could address challenges in employee training. Lanier noted that Booz Allen Hamilton currently utilizes research collaboration sessions, rapid innovation workshops, and design thinking exercises to facilitate internal problem solving.

DATA SCIENCE TRAINING IN THE WORKPLACE: EXECUTIVE EDUCATION

Executive Education Online

Brian Caffo, Johns Hopkins University

Caffo described **Johns Hopkins' Data Science Specialization**, delivered via Coursera, which includes the following courses: The Data Scientist's Toolbox, R Programming, Getting and Cleaning Data, Exploratory Data Analysis, Reproducible Research, Statistical Inference, Regression Models, Practical Machine Learning, Developing Data Products, and a Capstone Project done in collaboration with industry.

He explained that the program is unique in that it attempts to offer a complete data science curriculum through a large amount of bundled content; it provides all course notes on GitHub in R markdown and uses R almost exclusively; it utilizes Statistics with Interactive R Learning (**Swirl**); it allows free course textbook downloads via Leanpub; and it offers a LinkedIn space for alumni to connect upon completion.

Because Caffo and his colleagues found that industry managers often have fewer technical skills than their junior-level employees, they realized the urgent need for a specific training program to equip executives with the right skills to manage their teams. Johns Hopkins adapted the Data Science Specialization to create the Executive Data Science Specialization, which provides an overview of data science management. The Executive Data Science Coursera curriculum includes four content courses designed to be completed in only 1 week each:

1. A Crash Course in Data Science—High-level overview of statistics by example, machine learning, software engineering for data science, outputs of data science experiments, definitions of success, and the data science toolbox.
2. Building a Data Science Team—Overview of differences between types of data scientists and data engineers and how they can work together effectively.
3. Managing Data Analysis—Overview of types of questions asked by data scientists, qualities that make a sound question, exploratory analyses, inference, prediction, interpretation, modeling, and communication.
4. Data Science in Real Life—Ideal goals for data analysis, including clean data pulls, carefully designed experiments, and clear results, and strategies for when decisions are unclear or data products are ineffective.

Similar to the original program, the executive program emphasizes active learning and offers a Capstone Project (in partnership with Zillow and incorporating Swirl) upon completion of the coursework. Over the past year, 2,020 people completed the

Capstone Project in the executive program, with 99 percent awarding it positive ratings.

Ullman asked if the courses cover explainability of models, and Caffo responded that they circle around the topic of explainability by discussing knowledge creation, simple models, parsimony, and interpretability. In response to a question from Perry about the student demographics in the executive program, Caffo noted that the content is designed specifically with managers in mind; however, he cannot confirm whether managers are actually enrolling. Kathleen McKeown, Columbia University, inquired about the cost of the executive program, and Caffo noted that although the course videos and materials can be viewed for free, students have to pay to receive the certification upon completion. Mary Moynihan, Cape Cod Community College, mentioned that high-cost, for-credit online courses typically have only a 30 percent completion rate and wondered if this is the best way to train people in data science. Because completion rate is not necessarily an accurate indicator of engagement and learning in free or low-cost online courses and MOOCs, Caffo suggested that these programs may need to be evaluated differently from high-cost online courses.

Horton suggested that professional development is needed for faculty who wish to deliver online course content effectively. Prevost added that there also needs to be an incentive for an employee to complete an executive course, whether it be a component of a performance review or a monetary award. Horton posed a related question: How do we encourage people who do not have any incentive for further coursework? He noted that community colleges could play a role in training because of their low-cost, flexible offerings.

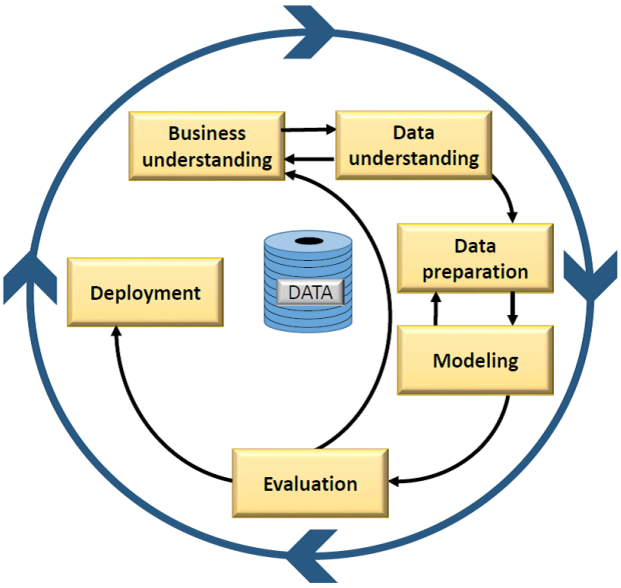

FIGURE 2 *The New York University course Data Mining for Business Intelligence follows an iterative data science process that emphasizes the formulation of a problem that can be addressed through data. SOURCE: Foster Provost and Tom Fawcett,* Data Science for Business, *2013.*

Executive Education in Business Schools

Claudia Perlich, Dstillery and New York University

Perlich described a course she offers at New York University titled Data Mining for Business Intelligence that offers two tracks for Masters in Business Administration students: the technical and the managerial. The technical track is offered in collaboration with the Center for Data Science and the computer science department and is taught solely in Python, while the managerial track often enrolls students without any programming skills but who wish to learn how to manage data science. This course introduces data science (1) terminology, (2) methods (e.g., supervised and unsupervised learning, model evaluation, data processing), (3) applications (e.g., case studies, Weka), and (4) management (e.g., deployment, hiring, interviewing, and proposal evaluation). This content is delivered via weekly lectures, guest speakers, homework assignments, a final exam, and a final team project. The course project requires students to identify a problem, find data, solve the problem, demonstrate business value, submit a written report, and present to the class—all of which are important data science skills (Figure 2).

In Perlich's view, students often have difficulty recognizing a predictive modeling problem, understanding the value of good baselines, translating a model into action, using precise language, and budgeting time for data preparation. However, upon completion of the managerial track, students are expected to be able to do the following:

- Approach business problems thoughtfully using data analytics to improve performance and know how to hire a data scientist;
- Understand that data preparation takes time but is necessary;
- Recognize that not all problems are data science problems;
- Think backwards from a problem, not forwards from the data;
- Know the basics of data mining processes, algorithms, and systems; and
- Have hands-on experience mining data.

In response to a question from Jessica Utts, University of California, Irvine, about helping students gain skills desired by employers, Perlich noted that it is incredibly difficult but necessary to teach communication skills and teamwork. John Abowd, Census Bureau, expressed concern about offering two separate tracks for the course, since the managerial-track students may not receive the same critical assessment experience as the technical-track students. Perlich responded that it would be difficult to cater to two audiences if faculty delivered this content via a single course, and she added that even if the tracks did not exist, students would likely self-select a course that best meets their knowledge and needs based on the syllabus content. While she sees value in offering a course without programming, she shares the concern about an overall decrease in technical content in data science curricula. In response to a question from Ullman about the course's attention to explainability of models, Perlich acknowledged that although both tracks discuss this topic, she is unconvinced that such discussions of transparency and explainability truly address issues of fairness and bias in data science.

OPEN DISCUSSION

Funding and Scaling Innovative Data Science Education in the U.S. Government

In response to a question from Alok Choudhary, Northwestern University, about data science education within the government, McKeown noted that training and retention are particularly important when hiring is constrained. Mark Krzysko, Department of Defense, said that government employees have unique challenges to becoming data literate, sharing data, and communicating across departments. He reiterated that the government has to work with what it has without overusing skilled employees. Krzysko noted that the government would benefit from an authoritative source of information on how data science can be used to help solve problems. Antonio Ortega, University of Southern California, highlighted USAFacts.org as an open-access repository that collects data from local, state, and federal government. He suggested that this be used as an entry point to address government data challenges. Horton suggested an independent statistical system that maintains high-quality data used to make better decisions in a non-partisan way.

Ullman noted that cost should not be considered a barrier to education given the accessibility to MOOC video content; using such material, faculty can create short courses at low cost. Zachary responded that while data training can be free, it is still completely inaccessible to many, and continuing to offer training only to those with access broadens inequalities in our communities. Krzysko and Abowd observed that hiring opportunities in the public sector are more limited than in the private sector and commended Zachary's creative efforts in addressing training challenges for the Department of Commerce workforce. Prevost suggested that organizations ask themselves what will be needed to upskill core employees, as well as those around them, and develop a "product-training-process" cycle that can be implemented whenever a new problem related to staff training surfaces. Abowd remarked that, when it is possible to hire new employees with different skill sets, the government needs help creating job descriptions that attract appropriate candidates. McKeown encouraged organizations in the public sector to weigh the benefits and drawbacks of hiring new employees versus upskilling current employees and added that recruiting and retaining individuals in government jobs that pay less than industry jobs can be challenging. Abowd mentioned that interns could meet specific data science needs, albeit with short-term availability.

Stodden highlighted the potential role of OpenGov, which leads the government transparency movement, in solving data science problems or in helping to build pipelines for data scientists to do public service for government agencies. She also suggested forming a group for data scientists, modeled after the Peace Corps or Teach for America, in which they can help communities or organizations solve large problems.

Using Sandboxes Across Organizations to Better Facilitate Progress

David Levermore, University of Maryland, College Park, said that lack of access to data is problematic for faculty trying to create hands-on projects for students. He noted how helpful it would be if industry and government made their data available to universities for student coursework. Caffo explained that simply granting open access for faculty to use data in their classes is insufficient; faculty have to analyze and prepare the data first to align it with their curricular goals, which is a time-intensive process. Plachy remarked that some small IBM data sets are shared in a sandbox for public use, but Laura Haas, IBM, responded that data licensing can be challenging; most companies do not want to risk legal action for accidentally releasing copyrighted data, which creates a substantial barrier to sharing data with universities. She posited that government agencies may face simi-

lar obstacles to data sharing since some data assumed to be open could actually include copyrighted material. Abowd suggested that faculty check the **Census Bureau data application program interface** for data they could freely use within the classroom.

Eric Kolaczyk, Boston University, explained that sandboxes spanning multiple organizations offer a holistic experience of working iteratively with experts in a team; the use of data repositories alone is inadequate. Creating successful sandbox experiences can be expensive, require energy and time, and depend on established relationships among stakeholders. Kolaczyk also wondered about the possibility of scaling sandbox experiences. Plachy referred to IBM's free beta version of the Data Science Experience because it provides teams a place to store data and collaborate. Kolaczyk suggested that this would be an even more useful platform if users had access to IBM data and IBM team members.

Abowd noted that the General Services Administration tried to execute sandboxes with GovCloud, but satisfying the agency-specific security requirements and completing the associated paperwork created implementation challenges. Kolaczyk reiterated that sandboxes are most useful when users have access to people, not just data. Abowd noted that government sandboxes will remain accessible to government employees for the time being, but he would like to see multi-organizational sandboxes offered in the future. Krzysko added that operational rules and infrastructure do not yet exist in the government to support such an endeavor; however, a recent pilot program giving federally funded research and development centers access to data and a dissemination guide indicates good progress.

Bridging Gaps in Knowledge and Perspectives Through Teamwork and Communication

Stodden wondered what makes communication skills for data scientists unique. Patrick Riley, Google, responded that while people working in traditional technical fields talk predominantly with other technical people, data scientists need to be able to explain difficult concepts to non-technical audiences. Perlich agreed that students have to learn to frame problems clearly for non-technical audiences. Riley suggested that students would benefit from practice exercises in which they have to present summaries of analyses to varied audiences. Levermore reiterated that the need for strong communication is not a new phenomenon; he suggested looking to the past when computational sciences was a new field and expanding those ideas to fit the even larger data science revolution. Gross said that methods for how to communicate scientific ideas to others could be integrated into any data science curriculum. He also suggested that educators focus on creating teams of varied backgrounds and perspectives, not just diverse knowledge levels. He pointed to educational approaches that can help reduce unconscious bias and teach others to speak effectively with one another, both of which are useful skills for teams composed of technical and non-technical members. Andrew Zieffler, University of Minnesota, cautioned that definitions of "teamwork" and recommendations for team sizes vary in the literature across disciplines, institutions, and organizations and need to be researched carefully by faculty designing curricula. A university that teaches broad teamwork skills best prepares students for diverse work environments, and Zieffler explained that one way to do this is to give students problems that are impossible to solve individually. Choudhary added that it is important to involve students in experiential learning and to bring technical and non-technical people together to define and refine problems. McKeown noted the value of exposing students to the unique vocabulary and approaches in varied disciplines so as to prepare them to work more cooperatively in interdisciplinary teams. David Culler, University of California, Berkeley, added that liberal arts skills (e.g., critical thinking, abstraction) aid in developing better data scientists.

Horton referenced a **software engineering course** at the University of California, Berkeley, as a model of teaching cross-disciplinary teamwork in which students used technology to solve important problems for non-profit organizations. Culler believes that current students often want to be producers of knowledge instead of consumers of knowledge; they just need the right tools and experiences to make a difference. Perlich added that while there is no shortage of good will, there is a crucial lack of project management, especially in volunteer programs attracting data scientists. She thinks that a model to ensure that people with the right skill sets are brought together and that volunteers are doing work related to their areas of expertise is needed. Catherine Cramer, Hall of Science, discussed the early intervention program "Big Data for Little Kids," which works with young children from immigrant families to improve access to STEM education. Noting the value of community partnerships, Zachary added that it is challenging to translate technical capacity to a specific need. She noted that inequalities may continue to grow if data scientists do not engage with the community's problems.

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, 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 (BMSA), the Computer Science and Telecommunications Board (CSTB), and the Board on Science Education (BOSE). Roundtable meetings will take place approximately four times per year. Please address any questions or comments to Ben Wender at bwender@nas.edu.

ROUNDTABLE MEMBERS PRESENT: Eric Kolaczyk, Boston University, Co-Chair; Kathleen McKeown, Columbia University, Co-Chair; John Abowd, U.S. Census Bureau; Ron Brachman (via webcast),Cornell University; Brian Caffo, Johns Hopkins University; Alok Choudhary, Northwestern University; Alfred Hero, University of Michigan; Nicholas Horton, Amherst College; Charles Isbell, Georgia Institute of Technology; Mark Krzysko, U.S. Department of Defense; Chris Mentzel, Gordon and Betty Moore Foundation; Deborah Nolan, University of California, Berkeley; Antonio Ortega, University of Southern California; Claudia Perlich, Dstillery and New York University; Patrick Perry, New York University; Victoria Stodden, University of Illinois, Urbana-Champaign; Mark Tygert (via webcast), Facebook Artificial Intelligence Research; Jeffrey Ullman, Stanford University; and Jessica Utts, University of California, Irvine.

GUESTS PRESENT: Quincy Brown, American Association for the Advancement of Science, Ashley Campana, Booz Allen Hamilton; Catherine Cramer, New York Hall of Science; David Culler, University of California, Berkeley; Ying Ding, Indiana University; Renee Dopplick, Association for Computing Machinery; E. Thomas Ewing, Virginia Tech; William Finzer, Concord Consortium; Louis Gross, University of Tennessee, Knoxville; Laura Haas, IBM; Ryan Seth Jones, Middle Tennessee State University; Brian Kotz, Montgomery College; Ashley Lanier, Booz Allen Hamilton; David Levermore, University of Maryland, College Park; Andrew McCallum, University of Massachusetts Amherst; Mary Moynihan, Cape Cod Community College; Rebecca Nugent, Carnegie Mellon University; Emily Plachy, IBM; Ron Prevost, U.S. Census Bureau; Lee Rainie, Pew Research Center; Hridesh Rajan, Iowa State University; Patrick Riley, Google; Rob Rutenbar, University of Illinois, Urbana-Champaign; Jordan Sellers, Howard University; Kristin Tolle, Microsoft; Ken Wilkins, National Institutes of Health; Drew Zachary, U.S. Department of Commerce; and Andrew Zieffler, University of Minnesota.

STAFF PRESENT: Linda Casola, Janel Dear, Natassja Linzau, 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