



August 2018

Roundtable HIGHLIGHTS

Roundtable on Data Science Postsecondary Education

Meeting #7 - June 13, 2018

The seventh Roundtable on Data Science Postsecondary Education was held on June 13, 2018, at the National Academy of Sciences Building in Washington, D.C. Stakeholders from data science education programs, government agencies, professional societies, foundations, and industry convened to explore the content and organization of new and emerging data science Ph.D. programs and to discuss alternatives for structuring Ph.D. programs, including stand-alone degrees, domain-based concentrations, and activities begun under the National Science Foundation's (NSF's) former Integrative Graduate Education and Research Traineeship program. 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.

Welcoming Roundtable participants, co-chair Kathy McKeown, Columbia University, noted that while many universities have focused on the development of undergraduate- and master's-level data science education, fewer Ph.D. programs in data science have been established. She emphasized the value of discussing curriculum requirements, levels of interdisciplinarity, departmental designations, institutional barriers, degree types, and research opportunities when evaluating or developing Ph.D. programs.

THE PH.D. PROGRAM IN DATA SCIENCE AT NEW YORK UNIVERSITY

Vasant Dhar, New York University

Dhar explained that New York University's (NYU's) [Center for Data Science \(CDS\)](#) was created in 2012 with support from representatives across campus. By creating a separate unit, NYU demonstrated its commitment to data science as a distinct area of study that integrates many disciplines. Although NYU ultimately plans to create full professorships in data science, current faculty appointments are joint between data science and another department.

NYU's Ph.D. program in data science admitted its first cohort—4 students—in 2017. From a well-qualified applicant pool of 400, the 2018 cohort includes 15 students who are diverse in geographic region, gender, and academic discipline. While all applicants had uniformly

high quantitative Graduate Record Examination (GRE) scores, admitted applicants had higher verbal GRE scores. He emphasized the added value of strong written and verbal communication skills as well as the ability to conduct scientific inquiry as preparation for data science study. [The Ph.D. curriculum](#) is structured in a way that blends engineering and social science and gives students flexibility and time to develop a thesis topic and find an appropriate advisor. The curriculum requires five core CDS courses—Introduction to Data Science, Probability and Statistics for Data Science, Machine Learning, Big Data, and Inference and Representation—and a multitude of electives from across the university. Over the course of the program, students participate in formal research rotations with faculty, take a qualifying exam and a comprehensive exam, and complete a dissertation. Dhar expects that the curriculum will continue to evolve in the future, in part driven by new faculty developing courses in their areas of expertise.

Daniel Spielman, Yale University, asked how NYU determines whether students need certain courses. Dhar explained that students can take placement exams, but he would prefer to see those decisions made by faculty on a case-by-case basis. In response to a question from Nicholas Horton, Amherst College, Dhar said that the Ph.D. program's five core courses are also offered at the master's level. James Frew, University of California, Santa Barbara, asked about the workplace experience of NYU's Ph.D. students, and Dhar estimated that at least half enter the Ph.D. program directly after completing a bachelor's or master's program. He noted, in response to a follow-up question from an audience participant, that students in NYU's data science Ph.D. program are funded by a combination of university and external fellowships.

Jeffrey Ullman, Stanford University, asked how a Ph.D. in data science compares to a Ph.D. in computer science for a student seeking employment in artificial intelligence. Dhar responded that if such a student is sufficiently motivated, he/she could attain the equivalent training with the Ph.D. in computer science as well; however, the interdisciplinary nature of NYU's data science program gives students a broad exposure across methods and domains and lead to research questions they might not ask in a typical computer science department. Jeffrey Brock, Brown University, posited that the differentiating factor between the Ph.D. programs in data science and computer science could be mathematical foundations. Dhar commented that while some differences exist in the types of mathematical foundations in

each program, more substantial differences can be found in the overall breadth of problem types that one encounters in a data science program, which can lead to methodological innovation. In response to a question from an audience participant, Dhar said that the Ph.D. programs in computer science and data science at NYU require the same total number of credits.

Devavrat Shah, Massachusetts Institute of Technology, wondered how faculty members balance their time developing courses for data science and teaching in their home departments. Dhar noted that currently those types of decisions are negotiated by the provost and the dean, although such processes will likely become formalized in the future. Despite the burden placed on faculty to contribute in both areas, Dhar reiterated the value of collaborating across disciplines and the excitement of working in an emerging field. Charles Isbell, Georgia Institute of Technology, asked how NYU manages culture clashes commonly found in interdisciplinary programs. Dhar replied that CDS has a positive outlook and has thus far avoided such clashes; participants acknowledge the value of interdisciplinarity and appreciate what they can learn from one another. In response to a question from Abani Patra, University at Buffalo, Dhar said that faculty with joint appointments will be reviewed and evaluated for tenure by both the home department and CDS.

YALE'S PH.D. PROGRAM IN STATISTICS AND DATA SCIENCE

Daniel Spielman, Yale University

Spielman explained that Yale's Department of Statistics became the Department of Statistics and Data Science in 2017 and hosts both an undergraduate major and a Ph.D. program. The Ph.D. program is structured in a way that reflects this evolutionary approach. To foster interdisciplinarity, some new faculty hires at Yale are being offered "half slots" in the Department of Statistics in Data Science; although resources and responsibilities come from both the Department of Statistics and Data Science and the faculty member's home department, the faculty member only completes the tenure process in the home department. The Department of Statistics and Data Science also offers secondary faculty appointments, which provide opportunities for collaboration on student data projects without teaching obligations from the department.

Yale's Ph.D. in statistics and data science requires 12 courses, which help to define what it means to

be a data scientist and to create a common culture among students practicing data science. Spielman noted that the Ph.D. program should take students approximately 5 years to complete, 2 of which will be dedicated to coursework. Requirements include a course and a qualifying exam in probability; a course and a qualifying exam in statistics; coursework in computation; studies in practical data analysis²; and a research oral exam. In response to a question from Ullman about whether requiring a qualifying exam in statistics but not computation emphasized data analysis over problem solving, Spielman explained that the coursework requires successful problem solving as does the practical data exam. Although Ph.D. students can choose advisors from other departments, the thesis is supervised at least in part by a member of the [Department of Statistics and Data Science](#).

Yale plans to increase the size of the incoming class of Ph.D. students from four to six and to revise the grant structure for students. Spielman commented that once a truly coherent culture is developed in the Ph.D. program, the Department of Statistics and Data Science might scale back course requirements as well as consider an alternate name that would better embrace the broad spectrum of data science. Brock highlighted the important roles that administrators and funding agencies play in making these programs successful. A Ph.D. program in statistics and data science may motivate faculty to collaborate beyond their disciplinary siloes, which is crucial for the future of science. And NSF is creating conduits for graduate students to work in a domain area and data science, as well as promoting discussions across university boundaries. He added that establishing industry–university partnerships is essential as the data science landscape continues to evolve.

Alfred Hero III, University of Michigan, asked about industry’s perspective of a Ph.D. in statistics and data science. Spielman said that industry has high demand for students with undergraduate degrees in statistics, computer science, and applied mathematics, so he expects the same to be true for Ph.D.’s in statistics and data science because they further develop these skillsets. Alok Choudhary, Northwestern University, asked if the Ph.D. program teaches students how to build scalable software, and Spielman explained that

individual graduates will emerge with varied skills and strengths. This will best prepare them to be productive members of data science teams in the workplace, he continued. Philip Bourne, University of Virginia, emphasized the importance of breaking down traditional disciplinary siloes and transferring best practices across departments and institutions, both in the United States and abroad. Spielman agreed that it is important to engage faculty from other departments and universities to create intellectual diversity and introduce new methods.

INTRODUCTION TO STATISTICS AND DATA SCIENCE AT THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Devavrat Shah, Massachusetts Institute of Technology

Shah described the [Statistics and Data Science Center \(SDSC\)](#), which is part of the Massachusetts Institute of Technology’s (MIT’s) [Institute for Data, Systems, and Society](#), as an interdisciplinary academic center with the mission to advance statistics and data science programs and research activities across campus. The SDSC encourages connections with the social sciences, life sciences, and computational sciences.

The SDSC began offering an undergraduate minor in statistics and data science in 2016, professional education in data science in 2016, and an [interdisciplinary Ph.D. in statistics](#) in 2018, and will launch an online micro-master in statistics and data science for professionals in fall 2018. Shah said that MIT hosts the interdisciplinary Ph.D. through its five schools—Architecture and Planning; Engineering; Humanities, Arts, and Social Sciences; Management; and Science—because students need to be trained in statistics, computation, and data science in order to be successful, and no single unit at MIT could achieve this. The Ph.D. program is managed by an institute-wide standing committee, with representatives from and within each academic unit. Shah emphasized community-building as an essential part of the program, with weekly activities and annual events (e.g., [SDSCon](#)) sponsored by the SDSC as well as a required semester-long advanced research seminar.

Shah explained that students must be admitted to a home unit first before becoming eligible to apply to the interdisciplinary statistics Ph.D. program in a subsequent semester; that admission decision will be made by the home unit first and then by the institute-wide standing committee. In addition to course requirements from the students’ home units, Shah continued, courses across four foundational areas

¹ The studies in practical data analysis include a case studies course, a practical exam with a data problem that has to be solved within 1 week, and practical work through a semester-long project with a faculty member in another department.

(i.e., probability, statistics, computation and statistics, and data analysis) are required. While the probability and statistics courses share a common curriculum, the computation and statistics and the data analysis courses may vary across domains. Shah added that a student's thesis must be relevant to both statistics and data science in order to earn the interdisciplinary Ph.D.

In response to a concern from Mark Tygert, Facebook Artificial Intelligence Research, Shah said that prospective students are aware that acceptance into the interdisciplinary Ph.D. program is not guaranteed. Replying to questions from Isbell and Spielman, Shah noted that a graduate of this program would receive a degree that reads, "Ph.D. in 'X' and 'statistics and data science.'" Because of the community that is developed and the work that is required to complete the program, this degree signifies more than a "badge." Frew wondered about the administrative management of such a program, and Shah explained that the interdisciplinary program is relatively straightforward to manage as all units are provided a clear set of checkpoint guidelines. In response to a question from Choudhary, Shah commented that students are able to take courses in the interdisciplinary program without obtaining the interdisciplinary Ph.D. and added that qualification is determined by the individual units, not a centrally administered exam. Dhar asked what volume of students is expected for the program, and Shah replied that because the burden on students is substantial with six additional courses, only one or two students at a time are expected to apply to the interdisciplinary Ph.D. from each participating unit.

OPEN DISCUSSION

Diversity and Interdisciplinarity

An online participant asked how the diversity of students' backgrounds impacts the curricula of graduate programs. Dhar noted that although NYU's Ph.D. applicants come from many disciplines, no formal pathways have been created. This decision will be evaluated as the program evolves. Spielman noted that the statistics Ph.D. at Yale historically accepted and trained students with diverse academic backgrounds, and he is hopeful that the same can be done for participants in the Ph.D. program in statistics and data science. Shah explained that because MIT's program is interdisciplinary by design, students are expected to be heterogeneous. He believes this level of diverse experience attracts students to the program and ensures the best contributions from

each. Choudhary asked if the thesis in each of these Ph.D. programs is driven by domain data. Spielman replied that acceptable theses come in many forms: some are driven by data, some develop methods, and others prove a theorem without data. Dhar and Shah emphasized that a Ph.D. in data science allows for broad inquiry. Horton suggested that Roundtable members read the National Academies' report [Graduate STEM Education for the 21st Century](#) and underscored the importance of implementing evaluation and assessment, fostering a community, offering faculty development, and promoting diversity and inclusion in emerging data science Ph.D. programs.

Ethics and Curriculum Development

Lise Getoor, University of California, Santa Cruz, asked how these Ph.D. programs integrate responsible data science practice and data science ethics. Dhar said that CDS is collaborating with [AI Now](#) on this issue; although courses that discuss such topics are already available, it would be beneficial to create a formal course requirement in ethical data science for the Ph.D. program. Spielman noted that Yale has begun a search to hire faculty with the expertise to integrate ethics into the program, but the university does not yet offer a formal course beyond what is covered in a graduate case studies class. Shah asserted the need to involve social scientists in this conversation. He added that although such topics have been introduced in some courses, there is no single course in the ethics of data science at MIT. Mark Krzysko, Department of Defense, emphasized that a framework is needed around social and domain norms for data science practice. Tygert encouraged people to engage Facebook in these conversations about ethics, as the company continues to explore similar questions. Brock described a master's program at Brown University that includes a course on data and society, whose students noted that they do not believe privacy is important. Because such students do not experience data and the world in the same way as faculty, faculty have an added challenge in understanding how this dichotomy of viewpoints should affect course content and delivery.

DATA SCIENCE AT THE UNIVERSITY OF CALIFORNIA, DAVIS

Duncan Temple Lang, University of California, Davis

Temple Lang shared that the University of California (UC), Davis, perceives data science as a distinct academic discipline that focuses on the process of data-enabled research; explores the breadth of the entire data pipeline; and integrates mathematics,

computer science, and statistics. The data science curriculum focuses on applying data science across the domains as well as solving problems within the domains. UC Davis's goal is to engage and impact all disciplines from engineering to religious studies.

[The Data Science Initiative](#) began at UC Davis in 2014 when the provost provided funding to explore the best structure for data science education, including collaborative research, community-building across disciplines, training and consulting opportunities, and dedicated space in the campus library to connect people across diverse areas. Temple Lang explained that a new academic unit for data science, led by a multidisciplinary coalition of faculty, will be in place in the 2018-2019 academic year. This unit will provide an opportunity for a new perspective and culture in research and education and serve as a complement to the mathematics, statistics, and computer science departments.

UC Davis will ultimately offer three types of doctoral study in data science: a Ph.D. in data science; a Ph.D. in computer science, mathematics, or statistics; and a Ph.D. in a domain discipline. The latter two are of greatest focus for UC Davis currently, Temple Lang noted, because they attract the largest number of students. To provide educational opportunities for these currently enrolled Ph.D. students, UC Davis plans to offer two types of add-on data science credentials: a "designated emphasis" and a "graduate academic certificate." Both credentials require an additional four courses: Survey of Statistical Machine Learning; Data Technologies and Computational Reasoning; an elective; and a capstone project. The designated emphasis also requires a data science-related thesis and qualifying exams. Both credential programs are especially attractive to students in computer science, statistics, mathematics, and domain sciences, Temple Lang continued, because they give students practice with real data science problems. Both programs prepare graduates who seek employment outside of academia as well as graduates who may want to teach data science in a discipline.

Temple Lang is often asked, "Why offer a data science Ph.D. if all Ph.D.'s use data to do science?" He reiterated that data science has a unique culture and concept; therefore, an academic home that emphasizes the data science process, the entire data science pipeline, and multidisciplinary is essential. Such a home encourages students to engage in systematic research in workflows, data science problem-framing, computational environments for data analysis, data

visualization, data sources and fusion, reproducibility, and ethics. He added that UC Davis is committed to its acceptance of interdisciplinarity; for example, faculty can advise students in many different Ph.D. programs beyond those in their home departments. In addition to developing the new academic unit in data science, the Ph.D. in data science, and the add-on Ph.D. data science credentials, UC Davis also plans to continue its complementary data science initiative as well as develop a data science undergraduate major, data science minors with varied foci, and a data science master's degree.

Ullman expressed his skepticism of data science as a unique intellectual domain. Temple Lang responded that the core of data science is the composition of the process: framing data science problems, enabling qualitatively different research in existing fields, and communicating about data. Although there is overlap in the content of data science and other disciplines, he continued, data science has a unique focus. In response to a question from Hero about the role of information scientists in building the Ph.D. program in data science at UC Davis, Temple Lang noted that although UC Davis does not have a school of information, faculty with such expertise could find a home in the new academic unit for data science. Brock commented that Temple Lang's systematic research topics frame data as primary and domains as essential; these are the types of topics with which a data science Ph.D. student would engage. Magdalena Balazinska, University of Washington, mentioned that as data science departments emerge, computer science and statistics departments are evolving—data science departments often play an important role in uniting all of these efforts.

SOCIAL SCIENTIFIC DATA SCIENCE: BUILDING THE PENN STATE PH.D. IN SOCIAL DATA ANALYTICS

Burt Monroe, Pennsylvania State University

Monroe discussed [Social Data Analytics](#) (SoDA) at Penn State, which aims to integrate social science and data science approaches to better understand human interactions. SoDA resulted from an NSF-funded [Big Data Social Science Integrative Graduate Education and Research Traineeship Program](#) (BDSS-IGERT), and it hosts a dual-title Ph.D. (in cooperation with six departments), a doctoral minor, and a bachelor's of science degree. According to Monroe, with its focus on socially relevant problems, SoDA excels in attracting and recruiting women and underrepresented groups.

Monroe provided a historical overview of the data science education-related efforts at Penn State that led to the development of SoDA, starting with the Social Science Statistics Partnership (SSSP) in 2004. This initiative began in an effort to raise the level of methodology within the political science and sociology departments. With funding from the College of Liberal Arts, SSSP expanded into the campus-wide Quantitative Social Science Initiative in 2006. The BDSS-IGERT grant of \$3 million in 2012 allowed for 2-year academic research rotations in interdisciplinary projects and summer externships for students, initial plans to create the SoDA curriculum, and community building through the establishment of the “Data-basement”—a central campus location where SoDA students meet.

Monroe explained that the [dual-title Ph.D. in SoDA](#) is structured to offer interdisciplinary programs without creating new departments, similar to the model used at MIT. Penn State’s program is different from MIT’s, however, in that it is possible for a student to be accepted simultaneously into the home department and the SoDA program. Students complete a series of requirements in their home disciplines and an additional four courses to satisfy SoDA requirements (e.g., two data approaches and issues seminars and two courses from approved options in analytical, social, quantitative, and computational sciences). Monroe discussed some program design challenges, including agreeing upon the number and type of requirements for the Ph.D. program, navigating boundaries between social science and non-social science disciplines that think about data in different ways, achieving true interdisciplinarity, and balancing the levels of faculty ownership for the program.

Horton wondered whether this model is similar to a data science + X degree program. Monroe responded that it is different because it extends beyond substantive engagement with domain theories to exploration of methodological approaches unique to social science. He noted that no one model of data science education will meet everyone’s needs. Benjamin Ryan, Gallup, Inc., asked Monroe about his ideal relationship with industry, and Monroe said that SoDA has an industry advisory board and often invites industry speakers to campus. He emphasized that not all Ph.D. graduates should become university professors; if a graduate secures employment in any position that requires Ph.D. training, Monroe considers that a success.

DATA SCIENCE SPECIALIZATIONS IN PH.D. PROGRAMS AT THE UNIVERSITY OF WASHINGTON

Magdalena Balazinska, University of Washington

Balazinska explained that the University of Washington’s (UW’s) [eScience Institute](#) was founded in 2008 and has become a permanent fixture on campus, owing to funding from UW and the Washington state legislature. As UW’s neutral hub of data science activity, the eScience Institute strives to empower students and faculty to accelerate discovery and leverage data, no matter how complex. It aims to build community, further research, and educate. The eScience Institute includes more than 100 affiliated faculty from across the university as well as a number of postdoctoral and Ph.D. students from a 2013 NSF-IGERT award, and it extends open office hours to anyone on campus with a data problem.

The motto of the eScience Institute is “data science for all,” Balazinska continued. The eScience Education Working Group makes data science education available to any interested student through formal programs, short courses, domain-themed hack weeks, workshops, and seminars and encourages an interdisciplinary community for students. Because the students generally fall into two categories—those who want to use data science tools and those who want to build data science tools—varied educational approaches are needed.

Balazinska commented that UW’s formal data science education programs include a Ph.D. in a discipline with either an [“advanced data science option”](#) or a [“data science option”](#); an undergraduate degree with a data science option; a professional data science master’s degree; and a variety of professional certificates. To enroll in either of the Ph.D. options, students are first admitted to their participating home departments³ and then can simply elect an option. The options are managed by the individual departments, although a single framework under the eScience umbrella is used and a central steering committee oversees the process. The advanced data science option, which is intended for data science tool developers, requires students to complete three of four courses in basic statistics, machine learning, data management, and data visualization, in addition to any home department requirements. Students also take four quarters of an eScience seminar, Balazinska said. In the data science option, which is designed for data science tool users, students and departments have a bit more flexibility in the course requirements. More than 60 students currently par-

ticipate in the Ph.D. options. These Ph.D. options evolved out of an NSF-IGERT program that had additional requirements: IGERT students are co-advised by faculty in data science methods and in domain sciences, encouraged to participate in internships, and regularly attending seminars. Several networking activities are also available for students interested in data science, Balazinska explained, such as an annual retreat, student-led seminars, lunches, summits, program evaluation, and a career fair, all facilitated by the eScience Institute infrastructure and resources as the IGERT grant draws to a close.

In response to a question from McKeown, Balazinska confirmed that UW would like to expand its data science options in the humanities and social sciences. Replying to Atma Sahu, Coppin State University, Balazinska said that the core domain framework for both options was initially developed by the eScience Education Working Group and continues to evolve. Balazinska added that departments play a central role in developing and maintaining the options, with special consideration for issues of accreditation. An audience participant asked about prerequisites for the data science options, and Balazinska reiterated that the two levels of data science options target different audiences, depending on their needs and interests (i.e., some courses in the data science option have minimal or no prerequisites). Hero inquired about the interdepartmental partnerships that are required to develop successful Ph.D. programs in data science. Balazinska explained that UW tries to increase its capacity within individual departments by hiring additional faculty. She also said that UW has various departments teaching data science courses; as a result, departments are starting to specialize in certain areas and offer more courses.

SMALL GROUP DISCUSSIONS AND CONCLUDING CONVERSATIONS

Roundtable participants divided into two groups to discuss key questions that emerged earlier in the day. On behalf of his group, Bourne summarized conver-

sations surrounding the following questions: *From an employer's point of view, what are the anticipated advantages of a Ph.D. in data science in contrast to a Ph.D. in a domain? More broadly, as asked by Temple Lang, "Why [offer] a data science Ph.D. if all Ph.D.'s use data to do science?"* Bourne noted that because data science skills are in such high demand in industry, graduates of either type of program are likely to gain employment. A number of factors are important: if employers are seeking knowledge of the complete data life cycle (which he defined as acquisition, engineering, analytics, visualization, dissemination, ethics), a Ph.D. in data science would be more useful than a Ph.D. in a domain. Bourne observed that the unique cultures of different fields also play a role in educational preparation and hiring decisions—industry focuses on a product, academia focuses on knowledge creation, and government focuses on policy. The scope, scale, and topic of a particular project would also influence the type of knowledge and training best suited for success.

On behalf of his group, Frew summarized discussions in response to the following question: *Data science education at the Ph.D. level is multi-faceted, and institutions are coming up with many different approaches. Is it possible to identify emerging best practices to common process challenges?* Frew identified three models for emerging Ph.D. programs: (1) A start-up (i.e., a new entity created with existing faculty), (2) an expansion of an existing entity, or (3) an overlay (i.e., data science superimposed on top of preexisting departments). Best practices may vary by model. No matter which model is chosen, Frew continued, it is vital to recognize that contributing disciplines have diverse perspectives and to react to those appropriately. Institutions themselves also have varying levels of ease in piloting new programs. Frew added that all three models would benefit from the inclusion of a physical space, independent from any specific department, which allows for cross-disciplinary interactions and collaborations at an appropriate level. When implementing new models, Frew explained, it is important for institutions to incentivize cross-disciplinary collaboration. For example, Stanford University allows faculty members to serve as advisors of record for Ph.D. students in any department. Frew emphasized that cross-disciplinarity should not be seen as a barrier to tenure.

² Departments of astronomy, chemical engineering, genome sciences, and psychology currently offer the data science option; Departments of applied mathematics, astronomy, biology, chemical engineering, computer science and engineering, genome sciences, mathematics, oceanography, psychology, and statistics currently offer the advanced data science option.

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, 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 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 seventh meeting on June 13, 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; Jeffrey Brock, Brown University; Alok Choudhary, Northwestern University; James Frew, University of California, Santa Barbara; Lise Getoor, University of California, Santa Cruz; Alfred Hero III, University of Michigan; Nicholas Horton, Amherst College; Charles Isbell, Georgia Institute of Technology; Mark Krzysko, U.S. Department of Defense; Duncan Temple Lang (via webcast), University of California, Davis; Uri Treisman, University of Texas, Austin; Mark Tygert, Facebook Artificial Intelligence Research; and Jeffrey Ullman, Stanford University.

GUESTS PRESENT: Bilikis Akindel, Duke Health Technology Solutions Analytics Center of Excellence; Magdalena Balazinska, University of Washington; Rocco Blais, National Intelligence University; Karim Boughida, University of Rhode Island; Philip Bourne, University of Virginia; Arlyn Burgess, University of Virginia; Patricia Cifuentes, Pan American Health Organization; Vasant Dhar, New York University; Jason Dunavant, Selbst; Lisa Federer, National Institutes of Health; Narryn Fisher, Technology Consulting; Michael Garriss, National Institute of Standards and Technology; Robert Hershey, Robert L. Hershey, P.E.; Aditya Johri, George Mason University; Erick Jones, National Science Foundation; Yasir Khalid, Georgetown University; Elizabeth Linton, Mount Sinai Hospital; Bert Little, University of Louisville; Raghu Machiraj, The Ohio State University; Burt Monroe, Pennsylvania State University; Ademola Okerinde, Kansas State University; George Onyullo, Department of Energy and Environment; Chuba Oraka, University of the Potomac; Abani Patra, University at Buffalo; Bryan Pijanowski, Purdue University; Hridayesh Rajan, Iowa State University; Benjamin Ryan, Gallup, Inc.; Atma Sahu, Coppin State University; Steve Sawyer, Syracuse University; Eugenia Schenecker, George Washington University; Devavrat Shah, Massachusetts Institute of Technology; Surja Sharma, University of Maryland; Sharad Sharma, Bowie State University; Susan Singer, Rollins College; Martin Skarzynski, Foundation for Advanced Education in the Sciences; Andrew Sostek, Katanya; Daniel Spielman, Yale University; Ethan Steininger, Virtue Theory Inc; Michael Turner, University of North Carolina, Charlotte; Dila Udum, Bahcesehir University; Lucianne Walkowicz, Library of Congress; Bob Weinberg, Massachusetts Institute of Technology; Olivia Wong; Brian Wright, George Washington University; Bin Wu, Liberty Language Service Inc; and Bing Xue, Consultant.

STAFF PRESENT: Linda Casola, Janki Patel, Catherine Pollack, 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