

FEDERAL INTERAGENCY BIG DATA AND SOCIAL SCIENCE VIRTUAL CONFERENCE

September 30-October 2, 2020

Conference Program

Agenda | Day 1, Wednesday September 30

Time	Event
12:30-12:45	Welcome and overview <i>Bryan Hubbell, EPA</i> <i>Bill Riley, NIH</i> <i>Arthur Lupia, NSF, conference host</i>
12:45-2:00	Empowering social science to understand and ameliorate major challenges of human society <i>Gary King, Harvard</i>
2:00-2:15	Break
2:15-4:00	Linking social science data to big data platforms <i>Moderator: Arthur Lupia, NSF</i> <ul style="list-style-type: none">· Modernizing data and encouraging social science within the U.S. Census Bureau <i>Misty Heggeness, US Census Bureau</i>· Modernizing crime data reporting: linking law enforcement incident-based crime data to state and local indicators of public health and quality of life <i>Kim Martin, DOJ</i>· Linking disparate data sources to advance environmental health sciences <i>Chris Duncan, NIEHS</i>· Collecting and integrating patient-reported outcomes with electronic health records <i>Chun-Ju (Janey) Hsiao, AHRQ</i>· Educational data <i>Anthony Kelly, George Mason University</i>
4:00-4:15	Break
4:15-4:45	Interfacing between small and big data: Social interactions and educational inequality <i>Rebecca Johnson, Dartmouth College</i>
4:45-5:00	Day 1 wrap up

Agenda | Day 2, Thursday October 1

Time	Event
12:00-12:10	Welcome and day 1 recap
12:10-1:00	Lunch and learn presentation on big data tools for social sciences <i>Frauke Kreuter, University of Maryland</i>
1:00-1:15	Break
1:15-2:00	Keynote presentation: NOAA big data and social science: Ensuring public safety and American prosperity in the 21st century. <i>Tim Gallaudet, Ph.D., Rear Admiral, U.S. Navy (Ret.) Assistant Secretary of Commerce for Oceans and Atmosphere / Deputy NOAA Administrator</i>
2:00-3:45	Ethics of big data, Part 1 Moderator: <i>Anne Weaver, EPA</i> <ul style="list-style-type: none">· When the legitimacy of data comes undone <i>danah boyd, Microsoft Research</i>· Privacy & confidentiality: Perspectives from the intersection of big data & environmental health studies <i>Cavin Ward-Caviness, EPA</i>· New approaches to data access and use <i>Julia Lane, NYU</i>
3:45-4:00	Break
4:00-5:00	Ethics of big data, Part 2 Moderator: <i>Bryan Hubbell, EPA</i> <ul style="list-style-type: none">· Algorithms/AI <i>Aaron Roth, University of Pennsylvania</i>· Environmental epidemiology and big data <i>Ana Rappold, EPA</i>
5:00-5:15	Day 2 wrap up

Agenda | Day 3, Friday October 2

Time	Event
	Welcome and day 2 recap
10:30-10:45	Personally-generated data Moderator: <i>Cavin Ward-Caviness, EPA</i> <ul style="list-style-type: none">· Modeling personally generated health-related data: dynamic, temporally dense, highly contextualized—and messy <i>Donna Spruijt-Metz, USC</i>· Volunteered geographic information for managing recreation on public lands <i>Spencer Wood, University of Washington</i>· Human mobility data, recreation and water quality <i>Nate Merrill, EPA</i>· Participant-generated data in the era of the internet of things <i>Wendy Nilsen, NSF</i>
12:15-1:00	Break
1:00-2:30	Promise and perils of social media data Moderator: <i>Angela Moore, DOJ</i> <ul style="list-style-type: none">· Using crowd-sourced data in public lands recreation monitoring <i>Eric White, USFS</i>· Harnessing social media data in the face of environmental Uncertainty <i>Sonya Sachdeva, USFS</i>· Privacy & ethics in a world of volunteered data <i>Katie Shilton, University of Maryland</i>· Health misinformation on social media: measuring impact and developing responses <i>Wen-Ying Sylvia Chou, NIH/NCI</i>
2:30-3:15	Conference Wrap-up Moderator: <i>Cavin Ward-Caviness, EPA</i> <ul style="list-style-type: none">· Sharing of perspectives – what has changed?· Actions we can take to increase/improve use of big data

Registration & Joining Zoom

- *Panelists and Attendees* must register and will receive an email with the link to join
- *Panelists* should download and use the Zoom Client
 - Refer to the Zoom Getting Started Site providing the download link and instructions for signing in.
- *Attendees* should join using the Zoom Client or via Chrome or Firefox browser
 - Perform a Zoom Test ahead of time: Zoom Automated Test Session - <https://zoom.us/test>
 - Instructions for Joining a Zoom Test Session: https://www.nsf.gov/pdf/about/participant/Zoom_Test_Meeting.pdf

Audience Guidance

Audience participants play a critical role in determining the quality of each session. To ensure the conference is an engaging and fruitful experience for all, please keep the following in mind:

- Be concise in your questions and reflections. As this event will take place virtually, we ask that you remain cognizant of the presenters' and other participants' time, and contact them directly with more detailed questions and comments.
- Help to keep participation lively by identifying key points worthy of further analysis and discussion.

Some guiding questions to consider ahead of the Q&A sessions:

- How can big data be used to address important questions faced by your agency?
- What sources of big data would be most useful in addressing your agency's needs?
- What are some of the challenges that you have faced in finding and using big data to address social science questions relevant to your agency?
- How can social sciences contribute to better understanding how society interacts with big data to inform decisions at scales from individual behavior changes to national policymaking?
- What do you see as the biggest opportunities and challenges to increasing the use of big data to address the needs of your agency?
- What kinds of training and tools do you think are needed to increase the use of big data in federal agencies?

Speakers



danah boyd is a Partner Researcher at Microsoft Research, the founder and president of Data & Society, and a Visiting Professor at New York University. Her research is focused on addressing social and cultural inequities by understanding the relationship between technology and society. She received a bachelor's degree in computer science from Brown University, a master's degree from the MIT Media Lab, and a Ph.D in Information from the University of California, Berkeley.



Wen-Ying Sylvia Chou, PhD, MPH, is a Program Director in the Health Communication and Informatics Research Branch (HCIRB) of the Behavioral Research Program (BRP) at the National Cancer Institute (NCI). Her research interests include health-related misinformation, social media and health, patient-provider communication, and mixed methods health research. Sylvia has a BA in Music from Santa Clara University, a MS and PhD in Linguistics from Georgetown University, and a Master in Public Health from UC Berkeley.



Christopher Duncan, Ph.D., is a Program Administrator in the Genes, Environment, and Health Branch of the National Institute of Environmental Health Sciences (NIEHS) Division of Extramural Research and Training. Chris is responsible for building and managing a portfolio of research grants in areas of data science relevant to environmental health. Duncan received an undergraduate degree in chemical engineering from North Carolina State University and a doctorate in pathology from Duke University.



Rear Admiral Tim Gallaudet, Ph.D., US Navy (ret) is the Assistant Secretary of Commerce for Oceans and Atmosphere and Deputy Administrator of the National Oceanic and Atmospheric Administration (NOAA). In his current position, Rear Admiral Gallaudet leads NOAA's Blue Economy activities that advance marine transportation, sustainable seafood, ocean exploration and mapping, marine tourism and recreation, and coastal resilience. Rear Admiral Gallaudet has a Bachelor's Degree from the U.S. Naval Academy and a Master's and Doctorate Degree from Scripps Institution of Oceanography, all in oceanography.



Misty L. Heggeness is a principal economist and senior advisor at the U.S. Census Bureau. She received her Ph.D. from the University of Minnesota in 2010. Her research focuses on family, gender, poverty & inequality, and the workforce. Within the federal government she has been involved in numerous projects to harness administrative records for research and build forward-thinking integrated data infrastructures that enhance federal statistics and data products throughout the federal government.



Chun-Ju (Janey) Hsiao is a health scientist administrator at the Division of Digital Healthcare Research within the Center for Evidence and Practice Improvement at AHRQ. She leads several initiatives on patient-reported outcomes including funding opportunities and a cross-agency project. Janey received a doctorate in health services research and a master of health science degree in health finance and management from the Johns Hopkins Bloomberg School of Public Health.



Bryan Hubbell is the National Program Director for the Air and Energy Research Program in the US EPA Office of Research and Development. He was previously Senior Advisor on Social Science for the US EPA Office of Research and Development. Bryan earned a Ph.D. in economics from NC State University. He is an author on over 50 peer-reviewed publications on a wide variety of topics and disciplines and has presented extensively in the U.S. and internationally on health and environmental impacts of air pollution and economic benefits and costs of air quality regulations.



Rebecca Johnson is an Assistant Professor in the Program in Quantitative Social Science at Dartmouth College. She is also an academic affiliate with the Office of Evaluation Sciences, a Data Science Fellow with The Lab at DC, and a faculty scholar with the American Bar Foundation. Her research focuses on how K-12 schools prioritize between students and the role of law and ethics in shaping that prioritization, and uses computational methods including text analysis, spatial analysis, and predictive modeling.



Anthony (Eamonn) Kelly is professor of educational psychology at George Mason University. He served as senior advisor to the Directorate of Education and Human Resources at NSF (2014-2018), Among other duties, he contributed to the National Privacy Research Strategy (NITRD] document (2016); and served on the NSF Harnessing the Data Revolution (HDR) committee. He was deeply involved in NSF/industry partnerships.



Gary King is Weatherhead University Professor, Harvard's most distinguished faculty position, and Director of its Institute for Quantitative Social Science. He develops and applies empirical methods in many areas of social science research, focusing on innovations that span the range from statistical theory to practical application. He has been elected to 8 honorary societies and has won more than 55 prizes and awards for his work. His statistical methods are used extensively in academia, government, and industry. He is co-founder of, and an inventor of the technology for, Crimson Hexagon, Learning Catalytics, Perusall, Thresher, OpenScholar and others. See GaryKing.org.



Frauke Kreuter is Director of the Joint Program in Survey Methodology at the University of Maryland, USA; Professor of Statistics and Methodology at the University of Mannheim; and head of the Statistical Methods Research Department at the Institute for Employment Research (IAB) in Nürnberg, Germany. She received her Master in Sociology from the University of Mannheim, Germany, and her PhD in Survey Methodology from the University of Konstanz. Her research focuses on sampling and measurement errors in complex surveys.



Julia Lane is a Professor at the NYU Wagner Graduate School of Public Service, at the NYU Center for Urban Science and Progress, and a NYU Provostial Fellow for Innovation Analytics. She is a senior advisor in the Office of the Federal CIO at the White House, supporting the implementation of the Federal Data Strategy. Julia received her PhD in Economics and Master's in Statistics from the University of Missouri.



Arthur Lupia is Assistant Director of the National Science Foundation, where he leads NSF's Social, Behavioral, and Economic Sciences Directorate. He also serves as co-chair of the National Science and Technology Council's Subcommittee on Open Science. Dr. Lupia is on assignment from the University of Michigan, where he is their Gerald R Ford Distinguished University Professor. Dr. Lupia's research and related public work examines processes, principles, and factors that guide decision-making and learning. He earned a bachelor's degree in economics from the University of Rochester and a PhD at the California Institute of Technology.

Kimberly Martin came to the Bureau of Justice Statistics in December 2013 from the Department of Sociology and Criminal Justice at Old Dominion University, where she was an assistant professor of criminology and taught classes on advanced quantitative research methods and statistics, criminology, and gender and crime. Dr. Martin has a Ph.D. in criminology from the University of Missouri-St. Louis, an M.S. in criminal justice from Georgia State University, and a B.A. in political science and B.S. in criminal justice from Georgia State University.



Nate Merrill is an environmental economist at US EPA's Office of Research and Development, where he studies the costs and benefits of improvements in coastal water quality. Through surveys, modeling, on-site observations and newly available human-mobility datasets, the recreational benefits of water quality can be quantified, informing policy makers on the value of protecting and improving the natural resources in the coastal zone. Nate received his Ph.D. in Environmental and Natural Resource Economics from the University of Rhode Island in Kingston, RI and a B.A. from Bates College in Lewiston, ME.



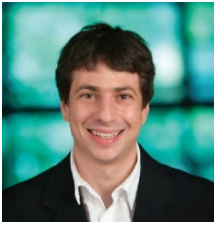
Wendy Nilsen, Ph.D. is in the Program Director in the Computer and Information Science and Engineering Directorate at NSF. She is also the lead Program Director in the Smart and Connected Health program. Her work focuses on the intersection of computing and human functioning. More specifically, her efforts include: serving as cochair of the Health Information Technology Research and Development working group of the Networking and Information Technology Research and Development Program; the lead for the NSF/NIH Smart and Connected Health announcement; convening workshops to address methodology in technology in health research; and, serving on numerous federal technology initiatives.



Dr. Angela Moore is a social and behavioral scientist who serves as a Senior Science Advisor at the National Institute of Justice (NIJ), the research, development, and evaluation agency of the U.S. Department of Justice (DOJ). Dr. Moore has taught criminology, criminal justice, and victimization courses at the University of Maryland, College Park, and the University of Maryland, Global College. Her research interests include interpersonal violence, violence against women, sentencing, corrections, reentry, and wrongful conviction. She received her Ph.D. in Criminology from the University of Maryland, College Park, and her B.S. and M.P.A. degrees from the Pennsylvania State University.



Dr. Ana G. Rappold is a Statistician with EPA's Office of Research and Development. She is a scientific lead of the project aimed to integrate public health messaging with environmental models and understanding their effectiveness to reduce burden in population. She has conducted a number of clinical and epidemiological research studies of health effects from air pollution and has authored a number of studies specific to smoke impacts on health.



Aaron Roth is a professor of Computer and Information Sciences at the University of Pennsylvania, affiliated with the Warren Center for Network and Data Science, and co-director of the Networked and Social Systems Engineering (NETS) program. His research focuses on the algorithmic foundations of data privacy, algorithmic fairness, game theory and mechanism design, learning theory, and the intersections of these topics. Together with Cynthia Dwork, he is the author of the book *The Algorithmic Foundations of Differential Privacy*. Together with Michael Kearns, he is the author of *The Ethical Algorithm*.



Sonya Sachdeva is a computational social scientist with the US Forest Service in the greater Chicago area, serving on the interdisciplinary People and Their Environments research work unit that studies the human component of natural resource management within urban ecosystems. She holds a Bachelors in Economics from the University of Michigan and a doctoral degree in Cognitive Science from Northwestern University. Her current projects involve large-scale automated text analysis of climate change coverage in the media, field assessments of the efficacy of environmental programs and behavioral experiments to study the impact of resource scarcity on conservation behavior.



Katie Shilton is an associate professor in the College of Information Studies at the University of Maryland, College Park. Her research explores ethics and policy for the design of information technologies. She is the PI of the PERVADE project, a multi-campus collaboration focused on big data research ethics. Other projects include developing privacy-sensitive search for email collections; analyzing ethical cultures in computer security research; and building tools to facilitate ethics discussions in mobile application development. Katie received a B.A. from Oberlin College, a Master of Library and Information Science from UCLA, and a Ph.D. in Information Studies from UCLA.



Donna Spruijt-Metz is Research Professor in Psychology and Professor in Preventive Medicine, housed in the USC Dornsife Center for Economic and Social Research. Her work meshes 21st century technologies with transdisciplinary metabolic, behavioral and environmental research to facilitate dynamic, personalized, contextualized behavioral interventions that can be adapted on the fly. She has led two NSF/EU/NIH-funded international workshops on building new computationally-enabled theoretical models to support health behavior change and maintenance in real- or near-time.



Dr. Cavin Ward-Caviness is a Principal Investigator in the Public Health and Integrated Toxicology Division of the US Environmental Protection Agency. With a background in computational biology and environmental epidemiology, Dr. Ward-Caviness seeks to understand the environmental factors which influence health in vulnerable populations and the molecular mechanisms that influence environmental health risks. By integrating molecular and clinical data, Dr. Ward-Caviness seeks to understand environmental health as a way to advance personalized medicine and reduce health disparities.



Dr. Anne Weaver joined the US Environmental Protection Agency in 2017, where her work has focused on studying disparities in air pollution and cardiovascular disease among people with different sociodemographic backgrounds. Her research includes underrepresented groups, such as rural populations, African American populations, and American Indian populations. She uses large datasets, including air pollution databases and medical records, as well as datasets from smaller research studies. Dr. Weaver received her PhD in epidemiology from the University at Buffalo in 2015.



Eric M. White is a Research Social Scientist with the Pacific Northwest Research Station of the USDA Forest Service. His research is focused on understanding recreation use on public lands and the economic effects on local communities, the social and economic outcomes from forest collaboration and restoration, and the ecological and social outcomes from collaborative natural resource management across landscapes of public and private ownerships. He has a Ph.D from Michigan State University and M.S. and B.S. degrees from Southern Illinois University.



Spencer Wood is an interdisciplinary empiricist and modeler who works at the nexus of social and ecological systems. His research on outdoor recreation addresses questions about how conservation and public land management influence the distribution, behaviors, and preferences of park visitors. In partnership with governments, NGOs, and corporations, Spencer leads projects that develop open-source software and reproducible approaches to make geographic data and models more accessible for decision-makers. Spencer is a Senior Research Scientist in the University of Washington's College of the Environment. He earned his PhD from the University of British Columbia.