

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Committee on National Statistics
Division of Behavioral and Social Sciences and Education

**Panel on Improving Federal Statistics for Policy and Social Science Research
Using Multiple Data Sources and State-of-the-Art Estimation Methods**

**February 25, 2016
Paul Brest East, Munger Building 4
Stanford University
Stanford, CA 94305**

**WORKSHOP ON THE CURRENT PRACTICES OF PRIVATE COMPANIES AND
THEIR USE OF BIG DATA AND KEY ISSUES AND CHALLENGES WITH PRIVACY
AND CONFIDENTIALITY**

AGENDA

OPEN SESSION

8:00 *Continental Breakfast Available*

8:30 **Welcome and Introductions**
Robert Groves, Chair: Georgetown University

8:45 – 10:00 **Data Collection and Assembly**
Moderator: Ophir Frieder, Georgetown University

Modernizing Statistical Data Collection
Joe Reisinger, Premise

10:00 – 10:15 BREAK

10:15 - 11:30 **Data Sharing and Access**
Moderator: Robert Groves, Georgetown University

Climate Change Data: Management and Distributed Resources
Dean Williams, Lawrence Livermore National Laboratory

Corporate Data Access and Sharing
Stephen J. Eglash, Stanford Data Science Initiative

11:30 – 12:30 Preserving Privacy

Moderator: Marc Rotenberg, Electronic Privacy Information Center

The Practice of (Privacy-preserving?) Data Sharing.
Aleksandra Korolova, University of Southern California

12:30 – 1:30 Lunch

1:30 – 2:45 Transaction Data: Analysis and Access

Moderator: Roberto Rigobon, MIT Sloan School of Management

An Overview of MasterCard's SpendingPulse."
Kamalesh Rao, MasterCard Advisors

The JPMorgan Chase Institute: Challenges, Opportunities and Next Steps in Using Proprietary Transaction Level Data for Economic Research
Rachel Pacheco, JPMorgan Chase Institute

2:45 – 4:00 Social Media Data

Moderator: Frauke Kreuter, University of Maryland

Data @ LinkedIn
Ya Xu and Kevin Morsony, LinkedIn

Curtiss Cobb, Facebook

4:00 – 5:00 Combining Survey Data with Organic Data

Moderator: Michael Chernew, Harvard Medical School

Google Tools for Data
Hal Varian, Google

5:00 Adjourn

6:30 Working Dinner

Joe Reisinger is the CTO and co-founder of Premise, a mobile technology platform orchestrating and crowdsourcing the collection of micro-data for precise and robust economic measurement. He holds a PhD in Computer Science from the University of Texas and spent his academic career building natural language understanding systems at Google Research and IBM T.J. Watson. Prior to co-founding Premise, he was Chief Scientist at Metamarkets.

Dean N. Williams has been a leading researcher in computational science at Lawrence Livermore National Laboratory (LLNL) since 1987. He has unique experience in distributed computing and networking technologies and practical application in the areas of climate change, biology, and other large-scale scientific data projects. For close to three decades, Mr. Williams has been the Principal Investigator (PI) or Chair for several large DOE projects related to “Big Data” initiatives, including the Earth System Grid Federation, the Ultrascale Visualization Climate Data Analysis Tools, and the International Climate Network Working Group. These combined software investigations are essential to the national and international climate communities and share in the 2007 Nobel Prize-winning Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4). After receiving his BA in Applied Mathematics and Statistics, in 1985, and his MS in Computer Science, in 1987, from the California State University at Chico, Mr. Williams has been engaged in research activities focusing on data modeling, data analytics and visualization, high-performance and distributed computing, and network technologies. As the Project Leader for LLNL’s Analytics and Informatics Management Systems, Mr. Williams has garnered many national and international awards, including the three more recent 2013, 2014, 2015 Federal Laboratory Consortium (FLC) awards for Advancing Federal Research and Technology” and the 2010 American Meteorological Society (AMS) special award for “Changing the Way Climate Science Does Business”.

Steve Eglash is an industry executive and technologist focusing on data science, energy, cleantech, optoelectronics, and materials with a strong technical background and experience in corporate management, engineering, marketing and sales, venture capital, and the management of innovation. Steve works at Stanford University where he manages programs in data science, energy, environment, and other fields. Previously, Steve was President and CEO of the solar energy company Cyrium Technologies, a full time consultant to the National Renewable Energy Laboratory and the US Department of Energy, and a venture capitalist at Worldview Technology Partners. Steve was also Vice President at SDL (JDSU) and a research scientist at MIT Lincoln Laboratory. Steve received a PhD and MS from Stanford University, and a BS from the University of California at Berkeley, all in Electrical Engineering. He has more than 40 papers published in peer-reviewed journals and four patents. Steve is a member of the Board of Directors of the Materials Research Society, a Fellow of the SPIE, and a member of the IEEE and OSA. Steve is a member of the Santa Clara University College of Arts & Sciences Leadership Board. He is on the City of Palo Alto Utilities Advisory Commission. Steve is involved in humanitarian projects such as the Tech Museum Awards for Technology Benefiting Humanity and the Global Social Benefit Incubator at Santa Clara University.

Aleksandra Korolova is a WiSE Gabilan Assistant Professor of Computer Science at USC, where she does research on privacy in the world of big data. Prior to joining USC, she was a Research Scientist at Google. Aleksandra received her Ph.D. in Computer Science from Stanford University. Her Ph.D. thesis, focused on protecting privacy when mining and sharing user data, has been recognized by 2011-2012 Arthur L. Samuel Thesis Award for the best Ph.D. thesis in the Computer Science department at Stanford. She is also a co-winner of the 2011 PET Award and a runner-up for the 2015 PET Award.

Kamalesh Rao is a statistician, data scientist and economist working with the SpendingPulse group of MasterCard Advisors. He specializes in applying combining statistics and machine learning to extract economic insight from large, messy datasets. Kamalesh and SpendingPulse has been quoted in the media by outlets including: Bloomberg, Reuters, Dow Jones, The Financial Times and The New York Times. His main focus at SpendingPulse is the macroeconomy and using the group's resources and insights to further explore the relationship of economic trends and consumer behavior. To this end, Kamalesh and his team track the retail sales of the US, Brazil, Japan, Hong Kong, Australia and the UK, and continues to add new fields of inquiry to SpendingPulse's research portfolio. Prior to joining MasterCard Advisors, Kamalesh was a Statistical Economist and Assistant Vice-President at Moody's, where he worked for more than seven years. At Moody's, Kamalesh worked under the company's Chief Economist, John Lonski, tracking and analyzing economic statistics and financial indicators concerning a wide range of topics including: the fixed-income markets, corporate credit, consumer spending, the emerging markets, the macroeconomy, unemployment, inflation and credit risk. Kamalesh holds a Masters in Statistics and a Bachelor's degree in Economic/Statistics, both from Columbia University in New York.

Rachel Pacheco is the Chief of Staff and Director of Operations for the JPMorgan Chase Institute. Rachel has worked at the intersection of the public and private sector for the past ten years, and specializes in corporate involvement in economic and social development challenges specifically in emerging markets. Her past consulting and research includes government reforms in Kazakhstan with Tony Blair Associates, climate financing in Indonesia, and political and social risk research in global infrastructure projects as a doctoral candidate at the University of Pennsylvania. Until 2009, Rachel served as a senior project manager at the Monitor Group, where she was a founding member of the company's Middle East practice based in Dubai and consulted regional governments on issues of economic development in the education and small business sectors. Rachel holds a B.S. in mathematics from Georgetown University and an M.B.A. from the Wharton School at the University of Pennsylvania.

Ya Xu is a Senior Staff Statistician & Engineer at LinkedIn Data Team. She leads a team of Engineers and Data Scientists to create their world-class experimentation platform and to advance methodologies for causal inference. In addition, she spearheads the effort of improving survey methodology at LinkedIn. Before LinkedIn, she worked at Microsoft Bing. She holds a Ph.D. in Statistics from Stanford University, and is a frequent speaker at top-tier conferences worldwide, including: WWW, WSDM, KDD and RecSys.

Kevin Morsony is a Senior Product & Privacy Counsel at LinkedIn. Kevin focuses on creating privacy protections for LinkedIn projects that provide data to external parties including for academic research collaborations and LinkedIn's Economic Graph programs. Kevin attended the University of Texas School of Law and UC Berkeley.

Hal R. Varian is the Chief Economist at Google. He started in May 2002 as a consultant and has been involved in many aspects of the company, including auction design, econometric analysis, finance, corporate strategy and public policy. He is also an emeritus professor at the University of California, Berkeley in three departments: business, economics, and information management. He received his SB degree from MIT in 1969 and his MA in mathematics and Ph.D. in economics from UC Berkeley in 1973. He has also taught at MIT, Stanford, Oxford, Michigan and other universities around the world. Dr. Varian is a fellow of the Guggenheim Foundation, the Econometric Society, and the American Academy of Arts and Sciences. He was Co-Editor of the American Economic Review from 1987-1990 and holds honorary doctorates from the University of Oulu, Finland and the University of Karlsruhe, Germany. Professor Varian has published numerous papers in economic theory, industrial organization, financial economics, econometrics and information economics. He is the author of two major economics textbooks which have been translated into 22 languages. He is the co-author of a bestselling book on business strategy, *Information Rules: A Strategic Guide to the Network Economy* and wrote a monthly column for the New York Times from 2000 to 2007.