

The National Academies of
SCIENCES • ENGINEERING • MEDICINE
Committee on National Statistics

**Challenges and New Approaches for Protecting Privacy
in Federal Statistical Programs**

June 6-7 2019

National Academies of Sciences, Engineering and Medicine, Lecture Room
2101 Constitution Ave. NW, Washington DC

For decades, federal statistical agencies have striven to balance the legal and ethical obligations to protect the confidentiality of data subjects with the need to provide informative statistics and access to data for secondary analysis. In recent years, balancing these objectives has become increasingly difficult. The digital revolution has seen an explosion in the growth of available data, both from public and private sources, which ill-intentioned actors could use to compromise confidentiality protections.

This workshop, organized by the Committee on National Statistics with the support of the federal statistical agencies will discuss new approaches to protecting data confidentiality, with a particular focus on methods that offer formal guarantees of privacy protection, like differential privacy. Discussions will cover the policy and implementation issues from both provider and user perspectives, including the promises and limitations of using formal privacy methods.

Workshop Agenda

June 6

Welcome

8:45 – 8:50 am Jerry Reiter, Committee Chair, Duke University
8:50 – 9:00 Brian Harris-Kojetin, Director, Committee on National Statistics,
National Academies of Sciences, Engineering and Medicine

Session 1 **9:00 – 10:15** Risks of existing statistical disclosure limitation methods
Dan Kifer, Pennsylvania State University, organizer and moderator

9:00 – 9:25 Reconstruction of person level data from data presented in multiple tables
Philip Leclerc, U.S. Bureau of the Census
9:25 – 9:50 De-Anonymizing the Australian Medicare Data Release
Ben Rubinstein, University of Melbourne, Australia
9:50 – 10:15 Discussion

Break **10:15 – 10:30**

Session 2 10:30 – 11:30 Insights from federal agencies about privacy protection needs
Michael Hawes, U.S. Bureau of the Census, organizer

- 10:30 – 10:40 Mary Theofanos, National Institute of Standards and Technology, moderator
- 10:40 – 10:55 Overview of the needs for federal statistics
Michael Hawes, U.S. Bureau of the Census
- 10:55 – 11:05 Darius Singpurwalla, National Center for Science and Engineering Statistics
- 11:05 – 11:15 Donna Miller, National Center for Health Statistics
- 11:15 – 11:30 Discussion

Session 3 11:30 – 12:30 PM Tutorial on differential privacy: a basic understanding of formal privacy methods; privacy budgets; methods to implement formal privacy. Alexandra Wood, Harvard University, organizer. Aleksandra Korolova, Moderator.

- 11:30 – 12:10 Kobbi Nissim, Georgetown University; Alexandra Wood, Harvard University
- 12:10 – 12:30 Discussion

Lunch 12:30 – 1:30

Session 4 1:30 – 2:45 Panel discussion: Current technical capabilities in formal privacy. What can we release via existing methods to satisfy formal privacy and still have high data usefulness? What can we not do yet? Applications from statistical agencies and elsewhere. Aleksandra Korolova, University of Southern California, organizer and moderator

- 1:30 – 2:15 Panelists: Ashwin Machanavajjhala, Duke University; Kamalika Chaudhuri, University of California San Diego; Ilya Mironov, Google, Salil Vadhan, Harvard University.
- 2:15 – 2:45 Discussion

Break 2:45 – 3:00

Session 5 3:00 – 4:00 Panel discussion: Policy issues and practical experiences with formal privacy. What do agencies need to consider when implementing and evaluating formal privacy solutions? Jerry Reiter, Duke University, organizer and moderator

3:00 – 3:40 Panelists: Nancy Potok, U.S. Office of Management and Budget; John Abowd, U.S. Bureau of the Census; Barry Johnson, Statistics of Income Division of the Internal Revenue Service

3:40 – 4:00 Discussion

Adjourn 4:00

June 7

9:00 – 9:15 AM Welcome and summary of Day 1, Jerry Reiter, Duke University

Session 6 9:15 – 10:30 Panel discussion: Implications of formal privacy for data users. What are users of data products concerned about if agencies move to formal privacy? Jerry Reiter, Duke University, organizer and moderator

9:15 – 10:00 Panelists: Julia Lane, New York University; Joe Salvo, New York City Department of City Planning; Mike Davern, NORC; John Friedman, Brown University, NBER

10:00 – 10:30 Discussion

Break 10:30 – 10:45

Session 7 10:45 – 12:00 Panel discussion: Looking forward to the near future: What can agencies do right now that achieve improved privacy protection? What do they need to investigate in the short and long term? What are intermediate solutions? Michael Hawes, organizer and moderator

10:45 – 11:25 Panelists: Tom Krenzke, Westat; Lars Vilhuber, Cornell University

11:25 – 11:45 Workshop summary – Jerry Reiter, Duke University

11:45 – 12:00 Discussion

Lunch 12:00 – 1:00

Session 8 1:00 – 2:00 Speed session with flash presentations: 5 minutes and 2-3 slides each for people to summarize their privacy research, problems, experiences, and solutions. Moderator: Alexandra Wood, Harvard University

1:01 – 1:07 Data Privacy and the USDA Forest Service's Forest Inventory and Analysis Program Surveys

Brett Butler, USDA Forest Service

1:07 – 1:13 Full-Rank Sufficient Statistics for Limiting Disclosure of Cross-Classifications

- Yves Thibaudeau, U.S. Bureau of the Census
1:13 – 1:19 Changing Farm Structure and its Potential Impact on Ag-Census Disclosure Analysis
Bayazid Sarkar, USDA NASS
1:19 – 1:25 Achieving Differential Privacy Using a Two-Tailed Geometric "Bottom-Up" Mechanism
Stephen Clark, U.S. Bureau of the Census
1:25 – 1:31 Implications of Differential Privacy for America's Children
Bill O'Hare, O'Hare Data and Demographics
1:31 – 1:37 Research on Using Synthetic Microdata to Protect Economic Data: Utility and Privacy Protection
Katherine Thompson, U.S. Bureau of the Census
1:37 – 1:43 Synthetic Data Quality Metrics: Relative vs. Absolute
Christine Task, Knexus Research
1:43 – 1:49 Respondents' Understanding of Disclosure Avoidance
Aleia Yvonne Clark Fobia, U.S. Bureau of the Census
1:49 – 1:55 Bayesian Pseudo Posterior Inference for Data Privacy Protection
Terrance Savitsky, Bureau of Labor Statistics
1:55 – 2:01 CBAMS: Public Microdata with Randomized Response
Caleb Floyd, U.S. Bureau of the Census
2:01 – 2:20 Open Discussion

Adjourn 2:20

Planning Committee Members:

Jerome P. Reiter (Chair), Duke University
Daniel Kifer, Penn State University
Aleksandra Korolova, University of Southern California
Alexandra Wood, Harvard University
Michael Hawes, U.S. Bureau of the Census

This workshop was made possible by the contributions of CNSTAT Core Sponsors:

Agency for Healthcare Research and Quality (DHHS), Bureau of Transportation Statistics (DOT), Citizenship and Immigration Services, Research Evaluation Division (DHS), Economic Research Service (USDA), Energy Information Administration (DOE), Food and Nutrition Service (USDA), Forest Inventory and Analysis Program (Forest Service, USDA), Methodology, Measurement, and Statistics Program (NSF); National Agricultural Statistics Service (USDA), National Center for Education Statistics (Department of Education), National Center for Health Statistics (DHHS), National Center for Science and Engineering Statistics (NSF), National Center for Veterans Analysis and Statistics (Veteran's Administration), National Institute on Aging (DHHS), Office of Policy Development and Research (HUD), Office of Research, Evaluation, and Statistics (SSA), Office of the Assistant Secretary for Planning and Evaluation (DHHS), Statistics of Income Division, (IRS), Bureau of Economic Analysis (Commerce), Bureau of Justice Statistics (DOJ), Bureau of Labor Statistics (DOL), and U.S. Census Bureau (Commerce)