

Toward a Vision for a New Data Infrastructure for Federal Statistics and Social and Economic Research in the 21st Century

Workshop 3: Approaches to Sharing Blended Data in a 21st Century Data Infrastructure

AGENDA

MAY 23, 2023 (DAY 2) 10 a.m. - 12:30 p.m. (EDT)

-
- 10:00 – 10:05 a.m. **Welcome and Workshop Objectives (Day 2)**
 • **Jerome P. Reiter**, Consensus Panel Chair, *Duke University*
- 10:05 – 11:15 **Session 4: In-depth Case Study--Unique Privacy Considerations When Blending Education Data**
 10:05-10:10
 • **Robert Goerge**, Panel Member, *University of Chicago*

- 10:10-10:30
 • **Andrew M. Penner**, *University of California-Irvine*

This talk will discuss the legal, technical, and perceptual challenges in blending education data. Implications of FERPA and public perceptions of confidentiality protections will be explored. The talk will conclude with a discussion of data ownership as residing with the public, rather than with individuals.

- 10:30-10:50
 • **Kelly Dauberman**, *Statistics of Income Office*
 • **Paul Arnsberger**, *Statistics of Income Office*

This presentation will discuss the partnership between the Department of Education and the IRS Statistics of Income office to produce earnings data for the College Scorecard website. The dataset is a blend of federal student aid recipient data derived from the National Student Loan Data System (NSLDS) and IRS administrative tax records data. The presentation will provide an overview of how the data are obtained, and the implementation and evaluation of differential privacy.

10:50-11:15

• **Peggy Carr**, *National Center for Education Statistics*

Reflections on the prior presentations will be shared.

Panel/Audience Discussion

11:15 – 12:25 p.m.

Session 5: In-depth Case Study--Unique Privacy Considerations in Blending Financial Services Data

11:15-11:20

• **Mark Watson**, Panel Member, *Federal Reserve Bank of Kansas City (formerly)*

11:20-11:40

• **Larry Cordell**, *Federal Reserve Bank of Philadelphia*

The Federal Reserve System's Board and 12 Regional Banks obtain access to a broad array of confidential and public consumer data used in research, analysis, and public presentations. This talk will describe some research studies at the Federal Reserve, which blend consumer data from confidential and public sources. The Risk Assessment, Data Analysis, and Research (RADAR) Group at the FRS, which anonymizes these data to ensure individual privacy, will be emphasized.

11:40-12:00

• **John W. Towns**, *National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign*

Ensuring appropriate access to useful blended data will require not only physical and software infrastructure, but also human infrastructure—the collaboration of technical experts with users of blended data. This talk will describe some of the challenges and opportunities that exist in the development of the community of technical professionals needed to address this critical element of the ecosystem.

12:00-12:25

Panel/Audience Discussion

12:25 – 12:30

Wrap-up Day 2

Jerome P. Reiter, Consensus Panel Chair, *Duke University*

12:30 p.m.

Adjournment of Open Meeting

CLOSED MEETING OF PANEL

PANEL ON APPROACHES TO SHARING BLENDED DATA IN A 21ST CENTURY DATA INFRASTRUCTURE

Jerry Reiter (Chair), Duke University
Claire Bowen, Urban Institute
Aloni Cohen, University of Chicago
Diana Farrell, JPMorgan Chase Institute
Robert Goerge, University of Chicago
Nick Hart, Data Foundation
HV Jagadish, University of Michigan
Dan Kifer, The Pennsylvania State University
Karen Levy, Cornell University
Salomé Viljoen, Michigan Law School
Mark Watson, Federal Reserve Bank of Kansas City (formerly)