NATIONAL ACADEMIES Sciences Engineering Medicine

AI Day for Federal Statistics 2026
April 30, 2026

National Academy of Sciences Building
2101 Constitution Avenue NW, Washington, DC

PROGRAM

Location: Fred Kavli Auditorium

1:00 pm–1:10 pm **Introductory Remarks**
Katharine Abraham, University of Maryland, College Park and Chair, Committee on National Statistics

1:10 pm-1:20 pm **Statisticians in AI Leadership and Collaboration**
David Matteson, Mayo Clinic, Cornell University, and Director, National Institute of Statistical Sciences

1:20 pm-2:30 pm **Panel – Artificial Intelligence in Federal Statistics: Applications, Impacts, and ROI**

Moderator: Robert Sivinski, Chair, Federal Committee on Statistical Methodology

ROI of GenAI
Benjamin Rogers, Centers for Disease Control

Valid Survey Response Simulations: Considerations for Using AI to Improve Survey Production
Kristina Gligorić, Johns Hopkins University

From Legacy Autocoders to Large Language Models: New Approaches to Coding Industry and Occupation Data
Lynda Laughlin, U.S. Census Bureau

Looking Ahead: Trends for Statistical Agencies in a World With AI
Brian Quistorff, Bureau of Economic Analysis

2:30 pm–2:45 pm **Break** (move to breakout rooms)

Location: Various Rooms

2:45 pm–4:00 pm **Training and Capacity-building Breakout Sessions** *(all sessions concurrent)*

1) Working With AI Governance and Implementation Challenges
(Lecture Room)

GSA's USAi Platform of Ready-to-Use AI Tools, Capabilities, and Services
Zach Whitman, Chief AI and Data Officer, General Services Administration

Considerations for GenAI Adoption
Benjamin Rogers, Centers for Disease Control

Advancing Responsible AI: Governance and Implementation in Federal Applications
Gizem Korkmaz, Westat

Moderator: David Matteson, Mayo Clinic, Cornell University, and Director, National Institute of Statistical Sciences

2) AI-ready Data
(Room 120)

Artificial Intelligence for Enhancing Data Quality, Standardization, and Integration: Data Transformation Toolkit
Ramond Robinson, Bureau of Transportation Statistics

Building with MCP, Agentic Systems, and Open Data: A Practitioner's Perspective
Brock Webb, Census Bureau

Moderator: Amy O'Hara, Georgetown University

3) AI in Statistical Production and Workflow
(Fred Kavli Auditorium)

NCHS Data Linkage Program: Incorporating Artificial Intelligence (AI) and Machine Learning (ML) into the Data Linkage Workflow
Cordell Golden, National Center for Health Statistics

AI Foundation Models for Science at NASA
Rahul Ramachandran, National Aeronautics and Space Administration

2022: A Machine Learning Odyssey
Emily Wiley, Census Bureau

Moderator: Kristen Olson, University of Nebraska-Lincoln and Committee on National Statistics

4) AI and Modernization of Software Systems for Next Generation Analysts
(Members' Room)

BEA Boost: Modern Tools for Modern Analysts
Amanda Lyndaker, Bureau of Economic Analysis

AI is My Copilot: The BEA Experience with an AI Coding Assistant
Brian Quistorff, Bureau of Economic Analysis

Moderator: Brian Habing, University of South Carolina and National Institute of Statistical Sciences

5) Responsible AI: Privacy, Security, and Ethics Applied to CIPSEA and Other Restricted Data
(Room 125)

Advancing Privacy Enhancing Technologies in the National Secure Data Service
Lisa Mirel, National Center for Science and Engineering Statistics

Navigating AI Ethics, Privacy and Security via Current U.S. Regulatory Structures
Baron Rodriguez, WestEd

Moderator: Salil Vadhan, Harvard University and Committee on National Statistics

Location: West Court, East Court, and Great Hall

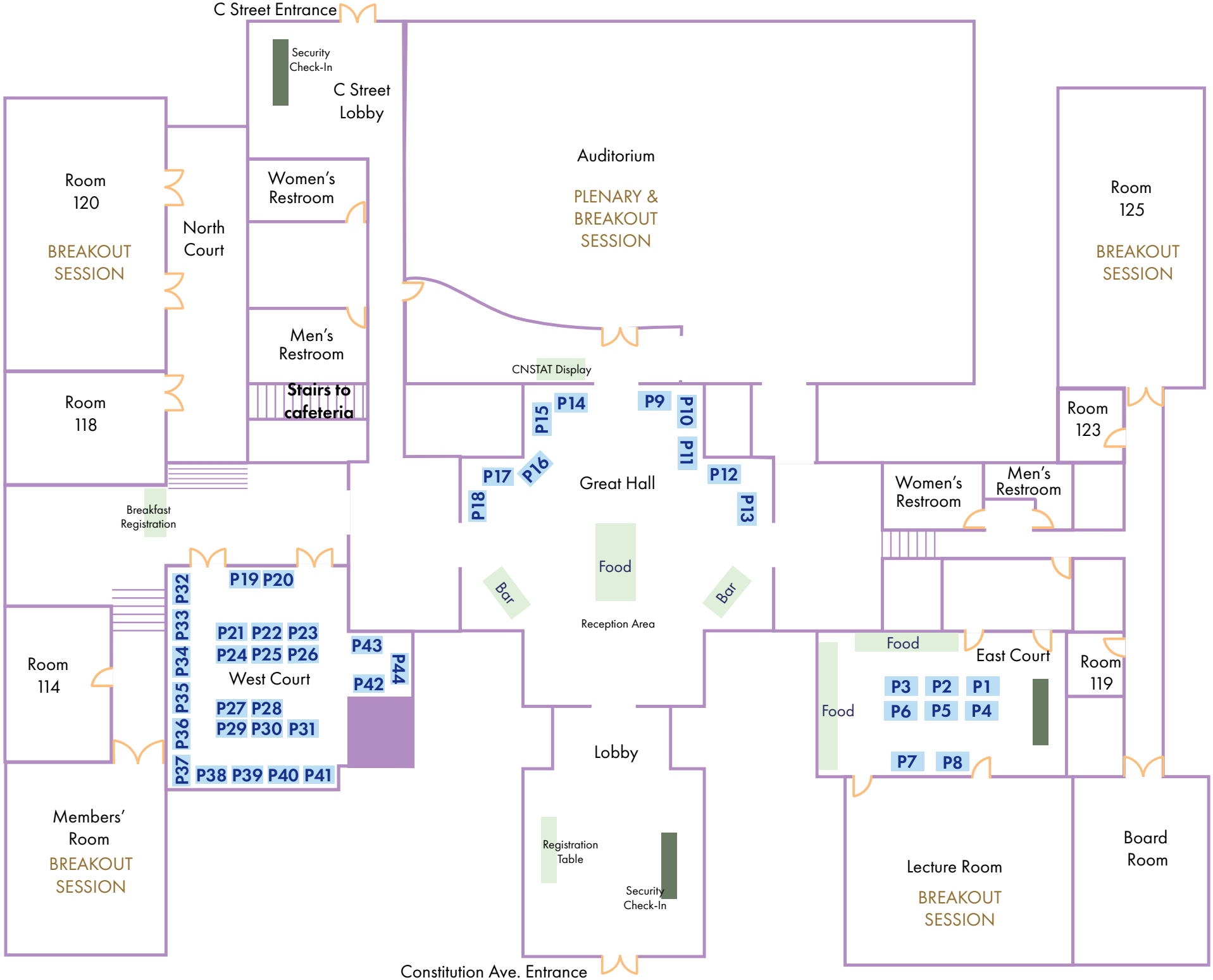
4:00 pm–5:00 pm **Poster Session and Reception**

East Court		
Location	**Poster Title**	**First Author**
P1	AI-assisted survey translation: Exploring the role of LLMs as expert reviewers	**Matt Dearstyne** Census Bureau matthew.j.dearstyne@census.gov
P2	Reimagining the Survey Development Process as an AI/Human Collaboration	**Renee Ellis** Census Bureau renee.ellis@census.gov
P3	Breaking the Bottleneck: Agentic AI for Timely Government Statistics	**Kim KiDeuk** Urban Institute kkim@urban.org
P4	AI-Assisted Travel Surveys: Enhancing Federal Travel Behavior Data with Generative AI	**Nivedya Madankara Kottayi** APDU Assoc. of Public Data Users nivedya.akkin@gmail.com
P5	Classifying Job Tasks Using Sentence-Transformer Model	**David Oh** Bureau of Labor Statistics oh.david@bls.gov
P6	Neural Network for Acreage Estimation Using Precision Agriculture and Survey Data	**Sean Rhodes** National Agricultural Statistics Service sean.rhodes@usda.gov
P7	Projecting AI Impacts on Employment	**Emily Rolen** Bureau of Labor Statistics rolen.emily@bls.gov
P8	Evaluation of an AI model for the Automation of In-Field Enumeration Reinterview Clerical Matching	**Kexin Zhang** Census Bureau Kexin.zhang@census.gov

GREAT HALL		
Location	**Poster Title**	**First Author**
P9	Using Machine Learning and Industry and Occupation Narratives to Identify Alternative Work Arrangements	**Joelle Abramowitz** University of Michigan Survey Research Center jabramow@umich.edu
P10	Combining Probability and Non-Probability Samples through Gaussian Process Mass-Imputation	**Soumadeep Bhowmick** University of Maryland Baltimore County xn84770@umbc.edu
P11	Privacy and Data Security for safer AI use	**Amrit Kohli** Bureau of Labor Statistics kohli.amrit@bls.gov
P12	A Practical Guide to Differentially Private Deep Learning using the Pseudo Posterior Mechanism	**Alexander J. Preiss** RTI International apreiss@rti.org
P13	Generative AI-Based Synthetic Data Generation	**Minsun Riddles** WESTAT MinsunRiddles@westat.com
P14	Privatizing an AI-Enhanced Pipeline for the US Census of Agriculture	**Luca Sartore** National Agricultural Statistics Service luca.sartore@usda.gov
P15	Democratizing Innovation Microdata: How Synthetic Public Use Files Can Address the Credibility Crisis in Applied Economics	**Matthew Williams** RTI International mrwilliams@rti.org
P16	AI in Statistics at BrightQuery	**BrightQuery** BrightQuery
P17	AI in Statistics at the National Institute of Statistical Sciences	**Megan Glenn** National Institute of Statistical Sciences mglenn@niss.org
P18	AI in Statistics at Summit	**Summit** Summit

WEST COURT		
Location	**Poster Title**	**First Author**
P19	Occupational Cluster Analysis	**Sertan Akinci** Bureau of Labor Statistics akinci.sertan@bls.gov
P20	Who Does What Now? Payroll-Based Occupation Imputation to Track AI-Driven Transitions	**Michael Dalton** Bureau of Labor Statistics dalton.michael@bls.gov
P21	The Microstructure of AI Diffusion: Evidence from Firms, Business Functions, and Worker Tasks	**Cory Breaux** U.S. Census Bureau cory.breaux@census.gov
P22	Are we there yet? An AI Roadmap Using Productivity Data	**Lamae Maharaj** Bureau of Labor Statistics Maharaj.Lamae@bls.gov
P23	AI and Industry Productivity Growth	**Peter Meyer** Bureau of Labor Statistics meyer.peter@bls.gov
P24	Enhancing Data Review Efficiency Through AI Assisted Excel Automation	**Maggie Behlen** U.S. Census Bureau margaret.behlen@census.gov
P25	2020 Post-Enumeration Survey Estimation Information Retrieval Using AI Agents	**Guorong Chen** U.S. Census Bureau guorong.chen@census.gov
P26	Unsupervised Learning for Anomaly Detection	**Shannon McDougall** U.S. Census Bureau shannon.mcdougall@census.gov
P27	Discrete Choice Person-Place Models: Address Prediction from Administrative Records Data for Census and Survey Operations	**Nathan G Welch** U.S. Census Bureau/Mitre nwelch@mitre.org
P28	REDI: Automated, Auditable Data Preparation for AI	**Sean R. Wilkinson** Department of Energy Oak Ridge National Labs wilkinsonsr@ornl.gov
P29	Applying Large Language Models to Maritime Near-Miss Safety Data Analysis	**Amanda Lemons** Bureau of Transportation Statistics amanda.lemons@dot.gov
P30	RTR Fusion: Integrating Retrieval, Reasoning, and Tools for Supply Chain Intelligence	**Jesse Ponnock** Department of Commerce/Mitre jponnock@mitre.org
P31	Metabolomic machine learning to predict serum perfluorooctane sulfonate (PFOS) and perfluorooctanoate (PFOA) levels in population studies	**Pei Zhang** National Institutes of Health zhangp9@nih.gov
WEST COURT		

Location	Poster Title	First Author
P32	Improving LLM Access to Federal Data	**Haley Johnson** General Services Administration haley.johnson@gsa.gov
P33	Leveraging Generative AI to Improve the Survey Process: Use Cases and Challenges	**Andreea Erciulescu** Westat AndreeaErciulescu@westat.com
P34	Mapping Smithsonian Research Activity Through Automated Geographic Entity Extraction	**Sydney Leiher** Smithsonian leihersj@si.edu
P35	The Data Navigator: Making Federal and State Data Accessible	**Heather Madray** National Science Foundation, National Center for Sci. and Eng. Statistics hmadray@nsf.gov
P36	Measuring Large Language Model Understanding of Federal Statistical Data	**Zachary Palmer** Department of Commerce zpalmer@doc.gov
P37	AI-Assisted Development for Fair Housing Data Mapping Tool	**Thiyaghessan Poongundranar** Urban Institute tpoongundranar@urban.org
P38	Artificial Intelligence for Enhancing Data Quality, Standardization, and Integration: Tools to Expedite and Streamline Statistical Workflows	**Ramond Robinson** Bureau of Transportation Statistics ramond.robinson@dot.gov
P39	Enhancing Qualitative Data Coding in Federal Statistics: A Human-Machine Learning Framework	**Judy H. Tang** Westat JudyTang@Westat.com
P40	Machine Learning using Tax Information: Improving Entity Classification in IRS Form 1041	**Sudeshna Roy** Statistics of Income Division, IRS Sudeshna.Roy@irs.gov
P41	Automating Text Classification in Corporate Tax Statistics: A Case Study of "Other Assets"	**Abigail Gaudinier** Statistics of Income, IRS abigail.gaudinier@irs.gov
P42	Bridging AI and Social Sciences for Robust Societal Decision-Making and Federal Statistical Modernization	**Norman Gottron** Carnegie Mellon University ngottron@andrew.cmu.edu
P43	AI-Supported Practice and Evaluation for Census Enumerator Training: A Closed-Loop Approach to Responsible AI in Census Workforce Development	**Nelson Er** U.S. Census Bureau nelson.l.er@census.gov
P44	TIDE (Tax Imputation & Data Engine): A Modern, Form-Agnostic Approach to Tax Data Validation featuring AI-Assisted Architecture Design	**Mark Xu** Statistics of Income Division, IRS mark.xu@irs.gov


C Street Entrance
Security Check-In
C Street Lobby
Auditorium
PLENARY & BREAKOUT SESSION
Room 120
BREAKOUT SESSION
North Court
Women's Restroom
Men's Restroom
Stairs to cafeteria
Room 118
Room 125
BREAKOUT SESSION
Room 123
CNSTAT Display
P14
P15
P9
P10
P11
P12
P13
P16
P17
P18
Great Hall
Women's Restroom
Men's Restroom
Breakfast Registration
Food
Bar
Bar
Reception Area
P19 P20
P32
P33
P34
P35
P36
P37
P21 P22 P23
P24 P25 P26
West Court
P27 P28
P29 P30 P31
P38 P39 P40 P41
P43
P44
P42
Room 114
Food
East Court
Room 119
P3 P2 P1
P6 P5 P4
P7 P8
Food
Lobby
Registration Table
Security Check-In
Members' Room
BREAKOUT SESSION
Lecture Room
BREAKOUT SESSION
Board Room
Constitution Ave. Entrance

Statisticians in AI Leadership and Collaboration

Topic: **Statisticians in AI Leadership and Collaboration**

Speaker: **David Matteson**, Mayo Clinic, Cornell University, and Director, National Institute of Statistical Sciences, dm484@cornell.edu

Bio: David S. Matteson is Professor of Biostatistics at Mayo Clinic, Professor of Statistics & Data Science at Cornell University, and Director of the National Institute of Statistical Sciences. His research centers on statistical AI methods for modeling complex human and natural processes and systems, with expertise in dynamic, spatial, functional, and network data science and machine learning. He received a CAREER Award from the National Science Foundation, the Chancellor's Award for Scholarship and Creative Activities from the State of New York, the inaugural Ann S. Bowers Research Excellence Award, Faculty Research Awards from the Xerox/PARC Foundation and LinkedIn, and is a Fellow of the American Statistical Association and the Institute of Mathematical Statistics. He has served as lead PI and Director for several NSF funded collaborative institutes, and he founded and serves as Editor-in-Chief for the new open access ASA affiliated journal Data Science in Science.

Panel Session: Artificial Intelligence in Federal Statistics: Applications, Impacts, and ROI

Title: **ROI of GenAI**

Speaker: **Benjamin Rogers**, Centers for Disease Control, qtw4@cdc.gov

Bio: Ben Rogers, CDC's Acting Deputy Chief Artificial Intelligence Officer (CAIO), has been with CDC for more than five years, serving as the Acting Deputy CAIO since early 2025. Previously, he served as a data scientist and AI Technology Advisor within the Office of Public Health Data, Surveillance, and Technology. He served on multiple details to CDC's COVID-19 response, completed a 2024 detail to the HHS AI Task Force, and has served as a Data Scientist in the National Center for Health Statistics. He holds a Master of Science in Data Science from the University of Virginia.

Title: **Valid Survey Response Simulations: Considerations for Using AI to Improve Survey Production**

Speaker: **Kristina Gligorić,** Assistant Professor of Computer Science, Whiting School of Engineering, Johns Hopkins University, gligoric@jhu.edu

Abstract: Large language models can augment the survey lifecycle, from pretesting questionnaires and simulating respondents to improving sampling. Studies based on surveys such as NHANES illustrate accuracy, effective sample size gains, and return on investment, while highlighting key considerations for using LLMs in a valid way. The talk concludes by examining the practical value that LLMs offer as a tool that can reduce costs, strengthen the explanatory power of survey items, and augment traditional survey operations at various stages.

Bio: Kristina Gligorić is an assistant professor of computer science at Johns Hopkins University. She received her PhD in computer science from EPFL in 2022 and was a postdoctoral scholar in the Computer Science Department at Stanford University before joining Johns Hopkins in 2025. Her

work sits at the intersection of AI and the social sciences, focusing on developing AI tools that expand the empirical toolkit of the social sciences, enabling measurement and understanding of human communication and behavior.

Title: **From Legacy Autocoders to Large Language Models: New Approaches to Coding Industry and Occupation Data**

Speaker: **Lynda Laughlin,** U.S. Census Bureau, lynda.l.laughlin@census.gov

Abstract: This presentation traces the evolution of industry and occupation autocoding in the American Community Survey, from traditional rules-based methods to emerging AI approaches, including multi-task transfer learning and large language models. It reviews the approaches explored, lessons learned, and the LLM-based method adopted for production, highlighting new opportunities for coding text data in survey operations.

Bio: Lynda Laughlin is a Senior Advisor in the Social, Economic, and Housing Statistics Division at the U.S. Census Bureau. She previously served as Chief of the Industry and Occupation Statistics Branch and has led work on occupational coding and classification systems. Her research interests include women and work outcomes, occupational measurement, and applying large language models to improve text coding in federal surveys.

Title: **Looking Ahead: Trends for Statistical Agencies in a World With AI**

Speaker: **Brian Quistorff,** Bureau of Economic Analysis, Brian.Quistorff@bea.gov

Abstract: Looking farther out in the future, there are fundamental ways that AI may impact statistical agencies, cross agency collaboration, and the broader statistical system. I highlight a few basic capabilities of AI, trace how they may affect the federal statistics, and pose some questions that agencies and leaders may want to consider going forward.

Bio: Brian Quistorff is the chief data scientist at BEA, where he works across DS and AI. He worked as a research economist at both Microsoft AI+Research division and the Bureau of Economic Analysis. He received a bachelor's degree in computer science from Stanford and a Ph.D. in economics from the University of Maryland.

Moderator: **Robert Sivinski**, Chair, Federal Committee on Statistical Methodology, Robert.Sivinski@hq.dhs.gov

Training and Capacity-building Breakout Sessions

1. Working With AI Governance and Implementation Challenges

Topic: **GSA's USAi Platform of Ready-to-Use AI Tools, Capabilities, and Services**

Speaker: **Zach Whitman**, Chief AI and Data Officer, General Services Administration, zach.whitman@gsa.gov

Bio: Zach Whitman is the Chief Data Scientist and Chief Artificial Intelligence Officer (CAIO) at the General Services Administration (GSA), where he supports the Data Analytics Center of Excellence (CoE) team, and leads cloud migration, data architecture modernization, change

management, data and analytics, data management, technology procurement, and strategy formation modernization efforts at external federal agencies.

Title: **Considerations for GenAI Adoption**

Speaker: **Benjamin Rogers**, Centers for Disease Control, qtw4@cdc.gov

Bio: Ben Rogers, CDC's Acting Deputy Chief Artificial Intelligence Officer (CAIO), has been with CDC for more than five years, serving as the Acting Deputy CAIO since early 2025. Previously, he served as a data scientist and AI Technology Advisor within the Office of Public Health Data, Surveillance, and Technology. He served on multiple details to CDC's COVID-19 response, completed a 2024 detail to the HHS AI Task Force, and has served as a Data Scientist in the National Center for Health Statistics. He holds a Master of Science in Data Science from the University of Virginia.

Title: **Advancing Responsible AI: Governance and Implementation in Federal Applications**

Speaker: **Gizem Korkmaz,** Vice President, Data Science and AI, Statistics and Data Science Center, Westat, gizemkorkmaz@westat.com

Abstract: AI-driven technologies are rapidly reshaping federal statistical work, offering opportunities to improve efficiency and scale. Federal agencies—and the researchers and contractors who support them—have a responsibility to maintain for accuracy, transparency, and the protection of sensitive information. While AI can accelerate processes, its adoption must be grounded in principles such as validity, reliability, fairness, transparency, and objectivity. AI governance plays a critical role in ensuring that systems are developed and deployed responsibly, in alignment with legal and ethical standards, as reflected in frameworks such as those from the National Institute of Standards and Technology. Despite these frameworks, many governance approaches may be too high-level to guide implementation in practice. This talk discusses practical strategies for operationalizing responsible AI, including human-in-the-loop review, documentation and reproducibility standards, model validation and monitoring, and structured governance approaches that translate principles into actionable steps across the project lifecycle.

Bio: Gizem Korkmaz, Ph.D., the vice president of data science and AI at Westat, leads and coordinates data science and AI activities across the company. Dr. Korkmaz is an economist with expertise in conducting social, behavioral, and policy research. Dr. Korkmaz applies innovative data science and AI methods and leverages both traditional and novel data sources to address science policy and measurement questions raised by policymakers and communities. Her work at Westat involves working with government agencies to harness the power of data to support evidence-based policy, as well as spearheading efforts to develop strategies for responsible use of traditional and generative AI in policy and survey research. Prior to joining Westat, Dr. Korkmaz was a research associate professor at the University of Virginia (UVA) and worked as a senior research scientist at Coleridge Initiative.

Moderator: **David Matteson**, Mayo Clinic, Cornell University, and Director, National Institute of Statistical Sciences, dm484@cornell.edu

2. AI-Ready Data

Title: **Artificial Intelligence for Enhancing Data Quality, Standardization, and Integration: Data Transformation Toolkit**

Speaker: **Ramond Robinson,** Director, Office of Transportation Analysis, U.S. Bureau of Transportation Statistics, Ramond.Robinson@dot.gov

Abstract: The AI-DQSI Toolkit was developed to demonstrate how targeted, modular AI capabilities can support DQSI activities routinely performed by federal statistical agencies. The design reflects the needs identified in the project's Framework Plan, particularly the need to: 1) improve metadata completeness; 2) address inconsistent schemas across data providers; 3) structure free text; 4)dynamically address missing data; and 5) extract structured information from documents that are not machine-readable.

Bio: Ramond Robinson serves as the Director of the Office of Transportation Analysis within the United States Department of Transportation (USDOT), Bureau of Transportation Statistics (BTS). With over 25 years of experience, he leads initiatives focused on compiling, analyzing, and managing multimodal transportation statistics.

Title: **Building with MCP, Agentic Systems, and Open Data: A Practitioner's Perspective**

Speaker: **Brock Webb,** Census AI Lead, Office of the Chief Information Officer/Computer Services Division (CSvD), U.S. Census Bureau, Brock.Webb@census.gov

Abstract: MCP servers are trivial to build. Making them "official" is where agencies get stuck. But the harder question is making government data AI-ready in the first place. This talk shares lessons from building agentic systems with open data, covering MCP's real friction, web-layer protocols like llms.txt and WebMCP, and what breaks when AI meets specialized data without context.

Bio: Brock Webb is the AI Lead at the U.S. Census Bureau, with over 20 years of experience across Census, DISA, and OMB. His work focuses on AI workflow design, evaluation methods, and pragmatics for official statistics, including building open-source tools for making Census data accessible to AI systems. He is currently exploring the practical challenges of integrating large language models into federal statistical workflows.

Moderator: **Amy O'Hara**, Georgetown University, Amy.OHara@georgetown.edu

3. AI in Statistical Production and Workflow

Title: **NCHS Data Linkage Program: Incorporating Artificial Intelligence (AI) and Machine Learning (ML) into the Data Linkage Workflow**

Speaker: **Cordell Golden,** Chief of the Data Linkage Methodology and Analysis Branch in the Division of Analysis and Epidemiology, National Center for Health Statistics, cdg4@cdc.gov

Abstract: The NCHS Data Linkage Program links NCHS survey data with administrative data to enhance the analytic utility of the survey data. The linkage creates new data resources that can support a wide range of studies on health care utilization and outcomes, public health, and policy evaluation. The presentation will include an overview the Data Linkage Program's efforts to

incorporate machine learning and artificial intelligence into the production workflow for developing linked data files to improve data quality and linkage efficiency.

Bio: Cordell Golden is the chief of the Data Linkage Methodology and Analysis Branch in the Division of Analysis and Epidemiology at the National Center for Health Statistics (NCHS). He directs the NCHS Data Linkage Program, leading agency efforts to expand the analytic potential of NCHS survey data by integrating it with vital records and other administrative data. He also oversees the application of innovative techniques to improve linkage efficiency, expand linkage opportunities, and increase access to linked files while safeguarding survey participant privacy.

Title: **AI Foundation Models for Science at NASA**

Speaker: **Rahul Ramachandran,** Senior Data Science Strategist, AI for Science Lead, Office of Data Science and Informatics (ST50), NASA/Marshall Space Flight Center, rahul.ramachandran@nasa.gov

Abstract: NASA manages petabytes of observational data, spanning scales from galactic structures to microscopic biological systems. While these archives offer unparalleled opportunities for scientific discovery, their sheer volume challenges traditional extraction of meaningful insights. Conventional deep learning addresses these issues but remains constrained by the resource-intensive requirement for large, labeled datasets. Foundation Models (FMs) offer an alternative by employing self-supervised learning to distinguish latent patterns within unlabeled data, thereby streamlining diverse downstream applications.

To operationalize this potential, NASA's Office of the Chief Science Data Officer implemented a strategy to develop open-source FMs using flagship datasets from all science divisions, alongside a science-specific language model. Key achievements include two iterations of the Prithvi Geospatial model for environmental monitoring and the Prithvi Weather and Climate model, designed to reconstruct atmospheric states and forecast future conditions. Additionally, the initiative recently deployed a heliophysics foundation model to advance space weather analysis. This presentation will provide an overview of the foundation models released to date, the workflows used in their design and development, and an open challenge for the community to build unified models.

Bio: Dr. Rahul Ramachandran is a Senior Research Scientist at NASA's Marshall Space Flight Center, where he directs the center's data science and artificial intelligence initiatives. He leads the Pathfinder AI for Science Portfolio for the Office of the Chief Science Data Officer, overseeing the development of foundation models including Prithvi-EO, Prithvi-WxC, and Surya Helio FM. Dr. Ramachandran is working on the concept of Accelerated Knowledge Discovery, a framework shifting AI from an analytic tool to a scientific collaborator through agentic software. His operational leadership includes modernizing the Global Hydrology Resource Center DAAC into NASA's first cloud-native archive and establishing the Satellite Needs Working Group Management Office to align federal priorities with NASA's Earth observations. A recipient of the Presidential Early Career Award for Scientists and Engineers (PECASE), NASA Exceptional Achievement Medal and AGU Leptoukh Lecture, Dr. Ramachandran fosters cross-sector partnerships to support open science and collaborative research.

Title: **2022: A Machine Learning Odyssey**

Speaker: **Emily Wiley,** Chief, Statistics Modernization Branch, Economy-Wide Statistics Division, U.S. Census Bureau, Emily.L.Wiley@census.gov

Abstract: In 2022, the U.S. Census Bureau implemented machine learning classification models in two major economic programs: the Economic Census and the Commodity Flow Survey. These models were introduced in real-time, respondent-facing instruments. This talk will cover the motivation behind these efforts, the methods used to build the models, and the key challenges we faced along the way.

Bio: Emily Wiley is the chief of the Statistics Modernization Branch at the U.S. Census Bureau. Her team of data scientists build analytical tools and machine learning models to innovate and modernize the Census Bureau's economic programs. Emily holds a master's degree in mathematics from the University of Central Florida and has earned two Department of Commerce Bronze Medals for innovation.

Moderator: **Kristen Olson** is the Leland J. and Dorothy H. Olson Professor in Sociology and Director of the Bureau of Sociological Research at the University of Nebraska-Lincoln. Dr. Olson's research focuses on mixed-mode surveys, survey costs, and questionnaire design, among other topics. She is a fellow of the American Statistical Association, the American Association for the Advancement of Science, and the Midwest Association for Public Opinion Research. She has an M.S. in survey methodology from the University of Maryland and a Ph.D. in survey methodology from the University of Michigan. kolson5@nebraska.edu; kolson5@unl.edu

4. AI and Modernization of Software Systems for Next Generation Analysts

Title: **BEA Boost: Modern Tools for Modern Analysts**

Speaker: **Amanda Lyndaker,** Bureau of Economic Analysis, Amanda.Lyndaker@bea.gov

Abstract: BEA is undertaking a sweeping modernization effort to replace legacy, opaque desktop applications with modern, Python-based workflows that strengthen statistical production and analytical capability. This presentation highlights the bureau's multi-year journey—its training strategy, community-building efforts, and the challenges faced as economists and statisticians adopt new tools at scale. Attendees will gain insight into how BEA is redesigning processes, empowering staff, and laying the technical foundation for future data-driven innovation.

Bio: Amanda Lyndaker is the Deputy CIO at the Bureau of Economic Analysis, leading modernization efforts that equip analysts with modern tools and data-driven capabilities. With deep experience bridging technology and federal statistical production, Amanda drives initiatives that strengthen collaboration, streamline workflows, and expand access to Python and data science across the bureau. She holds a Master of Arts in Applied Economics from Johns Hopkins University and a Bachelor of Science in Mathematics, Economics, and Statistics, from St. Lawrence University.

Title: **AI is My Copilot: The BEA Experience with an AI Coding Assistant**

Speaker: **Brian Quistorff,** Bureau of Economic Analysis, Brian.Quistorff@bea.gov

Abstract: BEA has been working for years to move away from rare/proprietary programming languages to modern/featureful ones such as Python. We share our experience using AI coding assistants to help us with this conversion process, both from an initial pilot project in 2024 and in more targeted work with a rare language (FAME). We show what can be done with off-the-shelf products and without significant developer resources.

Bio: Brian Quistorff is the chief data scientist at BEA, where he works across DS and AI. He worked as a research economist at both Microsoft AI+Research division and the Bureau of Economic Analysis. He received a bachelor's degree in computer science from Stanford and a Ph.D. in economics from the University of Maryland.

Moderator: **Brian Habing**, University of South Carolina and National Institute of Statistical Sciences, bhabing@niss.org

5. Responsible AI: Privacy, Security, and Ethics Applied to CIPSEA and Other Restricted Data

Title: **Advancing Privacy Enhancing Technologies in the National Secure Data Service**

Speaker: **Lisa Mirel**, National Center for Science and Engineering Statistics**,** lbmirel@nsf.gov

Abstract: In support of data modernization efforts, federal agencies have been linking data and making data more accessible. However, challenges exist when there are legal protections around data that protect privacy. Privacy enhancing technologies can support work in this space. For federal agencies and others to adopt such tools and technologies, a path to implementation with a strict focus on privacy and confidentiality needs to be established.

In 2022 the CHIPS and Science Act authorized the U.S. National Science Foundation (NSF) to establish a National Secure Data Service (NSDS) Demonstration project operated directly or via a contract by the National Center for Science and Engineering Statistics (NCSES). The goal of the NSDS Demonstration project is to inform efforts for developing a shared services model that would streamline and innovate data access and sharing to enable evidence-based decision making.

To date, nine NSDS projects have assessed or are assessing privacy enhancing technologies. The projects aim to support high quality, data-informed decision making while protecting privacy. This talk will give an overview of privacy enhancing technology projects in the NSDS. It will describe the lessons learned from these projects, the successful transition from pilot to implementation, and the tools and services available to support the use of privacy enhancing technologies.

Bio: Lisa Mirel is the Program Director of the Statistics and Methods group at the National Center for Science and Engineering Statistics within the U.S. National Science Foundation where she leads a team of statisticians and methodologists to support the Center's priorities related to survey design, statistical standards, privacy and confidentiality, data quality, and demonstration projects for a future National Secure Data Service. She previously served as chief of the Data Linkage Methodology and Analysis Branch at the National Center for Health Statistics (NCHS) where she directed NCHS's data linkage program. She is a member of the Federal Committee on Statistical Methodology and is also the current Washington Statistical Society President Elect.

Title: **Navigating AI Ethics, Privacy and Security via Current U.S. Regulatory Structures**

Speaker: **Baron Rodriguez**, WestEd, Baron.Rodriguez@wested.org, brodrig@wested.org

Abstract: Federal statistical agencies face a growing tension between executive directives accelerating AI adoption and durable statutory obligations under CIPSEA, FERPA, and the Public Trust Rule —

obligations backed by criminal penalties. This presentation examines three unresolved compliance questions: designating AI systems as "agents" under CIPSEA, maintaining the statistical purpose boundary when AI blurs statistical and administrative use, and whether synthetic data inherits CIPSEA protections. It addresses generative AI immaturity — credential persistence, retention uncertainties, and supply chain risks illustrated by recent federal vendor disruptions. The session concludes with responsible paths forward including trusted execution environments, privacy-preserving computation, and the ethical obligations accompanying AI use against pledge-protected data.

Bio: Baron Rodriguez is Vice President of AI & Client Facing Technology at WestEd and founder of the philanthropically funded Data Integration Support Center (DISC), a national center providing privacy, security, architecture, and responsible AI technical assistance to public agencies. A nationally recognized expert in data integration, privacy, security, and AI governance, Baron advises public organizations on designing and safeguarding data infrastructures while preserving the confidentiality of sensitive information. He plays a key role supporting policymakers, researchers, and agency leaders in leveraging data and technology to advance state-level policy goals. Prior to WestEd, Baron founded and led the U.S. Department of Education's Privacy Technical Assistance Center (PTAC) and served as Chief Information Officer for the Oregon Department of Education.

Moderator: **Salil Vadhan**, Harvard University and Committee on National Statistics, salil_vadhan@harvard.edu

THANK YOU TO OUR SPONSORS

CNSTAT Core Funders:

National Science Foundation: Methodology, Measurement, and Statistics Program; National Center for Science and Engineering Statistics | Social Security Administration | U.S. Department of Agriculture: Economic Research Service, National Agricultural Statistics Service | U.S. Department of Commerce: Bureau of Economic Analysis | U.S. Department of Health and Human Services: National Center for Health Statistics, National Institute on Aging | U.S. Department of Justice: Bureau of Justice Statistics | U.S. Department of Labor: Bureau of Labor Statistics | U.S. Department of Transportation: Bureau of Transportation Statistics | U.S. Department of the Treasury: Statistics of Income Division, Internal Revenue Service | Russell Sage Foundation

The Reception is Supported By:

National Institute of Statistical Sciences | American Statistical Association | BrightQuery, Inc. | Summit Consulting | Westat

This Event was Organized By:

Brandon Kopp, Bureau of Labor Statistics
Ed Strocko, Bureau of Transportation Statistics
David Matteson, National Institute of Statistical Sciences
Brian Habing, National Institute of Statistical Sciences
Chris Chapman, National Institute of Statistical Sciences
Linda Young, National Institute of Statistical Sciences
Maddi Nicol, Committee on National Statistics
Rebecca Krone, Committee on National Statistics
Melissa Chiu, Committee on National Statistics