Lars Vilhuber

lars.vilhuber@cornell.edu (607) 330-5743 Ithaca, NY larsvilhuber 0000-0001-5733-8932
lars.vilhuber.com Google Scholar Citations: 3,544 | H-index: 23 RePEc Citations: 562 | H-index: 12

Experience

American Economic Association, Data Editor — USA
Oversee data and code availability for economics journals published by the AEA — 2018 – present — 8 years
- Developed reproducibility verification processes
- Created training materials for reproducible research

Labor Dynamics Institute, Cornell University, Executive Director — Ithaca, NY
Lead research on — 2011 – present — 15 years
- labor dynamics
- confidentiality protection
- data access mechanisms
- reproducibility in social science research

Cornell University, ILR School, Senior Research Associate — Ithaca, NY
Research on labor economics and statistical disclosure limitation — 2007 – present — 19 years

ACES-Research, LLC, Vice-President — USA
2007 – present — 19 years

U.S. Census Bureau, Center for Economic Studies, Senior Research Associate — Washington, D.C.
Research on confidentiality protection and administrative data, creation of LEHD data infrastructure — 2004 – 2020 — 16 years

Cornell Institute for Social and Economic Research (CISER), Cornell University, Economist, Senior Research Associate or contractor — Ithaca, NY, USA
2001 – 2007 — 6 years

U.S. Census Bureau, Longitudinal Employer-Household Dynamics (LEHD) Program, Economist — Washington, D.C.
1999 – 2001 — 2 years

Dept. of Economics, York University, Adjunct Professor — Toronto, Canada
1999 – 2002 — 3 years

Dept. of Economics, York University, Research Fellow — Toronto, Canada
Co-responsible for COEP 1996 data administration and dissemination — 1998 – 1999 — 1 year

Publications

Full publication list — Feb 2026
See the complete publication list on my website.
Lars Vilhuber
lars.vilhuber.com/publications

Impact

Google Scholar: 3,527 citations | H-index 23

CitEc (RePEc): 562 citations | H-index 12

Publications: 244 publications and documents

Research Funding: 19 grants (2007–present)

As AEA Data Editor: 2791 manuscripts | 4973 authors reached

LDI Replication Lab: 201 undergradudates trained

Professional Service

Data Editor, American Economic Association journals 2018 – present

Oversee data and code availability for economics journals published by the AEA

- Impact: https://aeadataeditor.github.io/aea-cumulative-summary/impacts_of_aea_data_editing.html

Oversight Committee for Registry of Randomized Controlled Trials 2025 – present

Supervise the operations of the AEA's Registry of Randomized Controlled Trials, which promotes transparency and reproducibility in economics research by providing a platform for researchers to pre-register their randomized controlled trials.

- Chair, 2025-present
- Member: Nicholas Bloom
- Member: Esther Duflo
- Member: Edward Miguel

Harvard Data Science Review 2019 – present

- Associate Editor since 2019
- Column Editor, "Reinforcing Reproducibility and Replicability", since 2023

Conseil d'orientation du Centre interuniversitaire québécois de statistiques sociales (CIQSS)/Advisory board of the Quebec Inter-University Centre for Social Statistics (QICSS) [Canada], member Canada 2014 – present

Comité scientifique du Centre d'accès sécurisé aux données (CASD) [Scientific advisory board of the Center for secure data access, France] France 2012 – present

- Chair 2014-present
- Member 2012-2014

Working Group Open Science of the Verein für Socialpolitik [German Economic Association] Germany 2025 – present

Works on promoting open science practices in economics, within the German economics community. Convened by VfS President in 2025, Prof. Klaus Schmidt (LMU).

- Member

Urban-IRS Validation Server Advisory Board Meeting Washington, DC

- Member 2022 – present

Board of Directors of the Canadian Research Data Centre Network/ Réseau canadien des Centres de données de recherche Canada 2019 – 2026

- Member

Journal of Privacy and Confidentiality 2018 – 2024

Relaunched the journal with Cynthia Dwork, implemented open access infrastructure, ran the day-to-day process of the journal.

- Relaunch, 2018
- Executive Editor, 2018-2024

World Bank Review: Improving Reproducibility in the World Bank Policy Research Working Paper Series 2023

Peer reviewer of concept for improving reproducibility in the World Bank Policy Research Working Paper Series.

- World Bank launched https://reproducibility.worldbank.org/ and associated processes the same year.

Journal of Econometrics 2019 – 2021

- Guest Editor for special issue on Linked Employer-Employee Data

Committee on the Future of Scholarly Communications (Cornell University) Ithaca, NY

- Member 2019 – 2020

Social Sciences and Humanities Research Council of Canada (SSHRC) Canada

- Adjudication committee for Insight Grants, 2018-2019 2018 – 2019

American Statistical Association, Committee on Privacy and Confidentiality USA

- Chair 2020-2021 2016 – 2021
- Member 2016-2020

Contributions to activities of the National Academies of Science, Engineering, and Medicine USA 2016 – 2019

- Forecasting Costs for Preserving, Archiving, and Promoting Access to Biomedical Data (DEPS-BMSA-18-02)
- Transparency and Reproducibility of Federal Statistics for the National Center for Science and Engineering Statistics (DBASSE-CNSTAT-17-04)
- Workshop on Transparency and Reproducibility in Federal Statistics (DBASSE-CNSTAT-16-02), presentation in June 2017, workshop report see Vilhuber, 2019a
- Committee on Reproducibility and Replicability in Science (DBASSE-BBCSS-17-03), commissioned paper prepared July 2018 (Vilhuber, 2018c), final consensus report see National Academies of Sciences, Engineering, and Medicine, 2019

National Science Foundation, Methodology, Measurement, and Statistics (MMS) Advisory Panel USA 2016 – 2018

International Data Service Centre, IZA Institute of Labor Economics, Advisory Board Germany 2016 – present

LEHD Technical Papers 2000 – 2007

Professional Associations

- Society of Labor Economists (SOLE)
- American Economic Association (AEA)
- American Statistical Association (ASA)
- Elected member, The International Statistical Institute, 2018-

Grants & Funding

To train the next generation of scholars to conduct replication studies and produce transparent, reproducible, and replicable research 2023 – 2026

To train the next generation of scholars to conduct replication studies and produce transparent, reproducible, and replicable research.
Lars Vilhuber, Abel Brodeur
sloan.org/grant-detail/g-2023-22326

Collaborative Research: Elements: TRAnsparency CErtified (TRACE): Trusting Computational Research Without Repeating It July 2022 – Dec 2025

The central goal of the TRACE project is the development, validation, and implementation of a technical model of certified transparency. This includes a set of infrastructure elements that can be employed by system owners to (1) declare the dimensions of computational transparency supported by their platforms; (2) certify that a specific computational workflow was executed on the platform; and (3) bundle artifacts, records of their execution, technical metadata about their contents, and certify them for dissemination. The first phase of the project focuses on the development of a conceptual model and technical specification that can be used to certify the description of a system, termed a Transparency-Certified System (TRACE system), and the aggregation of artifacts along with records of their execution, termed Transparency-Certified Research Objects (TROs). The second phase focuses on the development of reusable software components implementing the TRACE model and approach. To demonstrate certified transparency, the toolkit is used to TRACE-enable existing platforms including Whole Tale, SKOPE, and the SLURM workload manager. These TRACE-enabled systems produce certified TROs that can be trusted and do not need to be repeated or re-executed to verify that results were obtained as claimed.
Lars Vilhuber (this grant: $182,000), Bertram Ludaescher (Grant 2209628), Timothy M McPhillips (Grant 2209628), Kacper Kowalik (Grant 2209628), Craig Willis (Grant 2209628), Thu-Mai Christian (Grant 2209630)
www.nsf.gov/awardsearch/showAward?AWD_ID=2209629&HistoricalAwards=false

Mid-Scale RI-2: the Research Data Ecosystem (Rde), a National Resource for Reproducible, Robust, and Transparent Social Science Research in the 21st Century June 2024 – June 2026

This project will implement a new platform for social and behavioral science data. Diverse types of data enable path-breaking analyses into human behavior but also present challenges of scale, sensitivity, and structure. Current barriers to conducting research include multiple incompatible standards for data, lack of interoperability, and the inherent difficulty of managing big data. There is an urgent need for new modes of access, confidentiality protection, methodological approaches, and tools so that research using a variety of data types meets accepted scientific standards. The Research Data Ecosystem (RDE) will modernize the management of data to enable a new era of interconnected research for the social and behavioral sciences. The platform will improve the quality of data-driven social and behavioral science research over the entire data life cycle. My role is to help identify issues that emerge from actual usage of the current infrastructure, and forecasting the needs of economists and possibly other social sciences, based on my experience as a researcher and as AEA Data Editor.
Margaret C Levenstein, Lars Vilhuber (sub-award)
www.nsf.gov/awardsearch/showAward?AWD_ID=1946932

Conferences on Reproducibility and Replicability in Economics and the Social Sciences (CRRESS) Aug 2022 – July 2025

This award provides partial support for a series of virtual and in-person conferences on the topics of reproducibility, replicability, and transparency in the social sciences. The conferences address educational and procedural barriers slowing down adoption of best practices, whether journals should be the verifiers of reproducibility, whether (and how) scientists' work can be made to be reproducible at every stage of the research process, and implications thereof for funding, technical infrastructure, and the training of undergraduate and graduate students.
Lars Vilhuber, Aleksandr Michuda
www.nsf.gov/awardsearch/showAward?AWD_ID=2217493

Subaward from: CC*DNI DIBBS: Merging Science and Cyberinfrastructure Pathways: The Whole Tale Mar 2016 – Feb 2023

The goal of this project is to strengthen the second layer of research output, and to build a robust third layer that integrates all parts of the story, conveying the holistic experience of reproducible scientific inquiry by (1) exposing existing cyberinfrastructure through popular frontends, e.g., digital notebooks (IPython, Jupyter), traditional scripting environments, and workflow systems; (2) developing the necessary 'software glue' for seamless access to different backend capabilities, including from DataNet federations and Data Infrastructure Building Blocks (DIBBs) projects; and (3) enhancing the complete data-to-publication lifecycle by empowering scientists to create computational narratives in their usual programming environments, enhanced with new capabilities from the underlying cyberinfrastructure (e.g., identity management, advanced data access and provenance APIs, and Digital Object Identifier-based data publications). The technologies and interfaces will be developed and stress-tested using a diverse set of data types, technical frameworks, and early adopters across a range of science domains. (I was not involved in the original proposal)
Bertram Ludaescher, Victoria Stodden, Niall I Gaffney, Matthew J Turk, Kyle Chard, Subaward: Lars Vilhuber, $35,000
www.nsf.gov/awardsearch/showAward?AWD_ID=1541450&HistoricalAwards=false

To identify and test effective interventions for enhancing the quality and completeness of randomized-controlled-trial registries 2021 – 2025

To identify and test effective interventions for enhancing the quality and completeness of randomized-controlled-trial registries.
Lars Vilhuber, Sarah Kopper, Jack Cavanagh
sloan.org/grant-detail/g-2021-17105

To improve policy-relevant research by facilitating the use of administrative data among economists who design, run, and analyze randomized controlled trials 2019 – 2022

To improve policy-relevant research by facilitating the use of administrative data among economists who design, run, and analyze randomized controlled trials. Grant supported the Innovations in Data and Experiments for Action Initiative (IDEA) at J-PAL Global, including core technical staff, two pilot projects in the United States, and creation of the Handbook on Using Administrative Data for Research and Evidence-Based Policy.
Anja Sautmann, Shawn Cole, Iqbal Dhaliwal, Lars Vilhuber (subaward to Cornell University)
sloan.org/grant-detail/g-2019-11391

To support a conference on research using linked employer-employee data to study labor markets and disseminate these insights to the wider economics community 2019 – 2021

To support a conference on research using linked employer-employee data to study labor markets and disseminate these insights to the wider economics community.
Lars Vilhuber, Ian Schmutte, David Card
sloan.org/grant-detail/g-2019-12486

To enhance the transparency, reproducibility, and replicability of empirical research in the social sciences by simplifying how authors can, when submitting a paper to a journal, also provide structured metadata about the provenance and archiving of code, data, and other supplementary materials 2018 – 2020

To enhance the transparency, reproducibility, and replicability of empirical research in the social sciences by simplifying how authors can, when submitting a paper to a journal, also provide structured metadata about the provenance and archiving of code, data, and other supplementary materials.
Lars Vilhuber, Carl Lagoze
sloan.org/grant-detail/g-2018-11377

To study the economics of socially efficient protocols for managing research databases containing private information 2015 – 2017

This project characterized the 'efficient frontier' of privacy-accuracy tradeoffs in research data protocols, assembling a library of protocols (including homomorphic encryption, secure computation, and differential privacy) and measuring public attitudes toward these trade-offs.
John M Abowd, Lars Vilhuber
sloan.org/grant-detail/g-2015-13903

RCN: Coordination of the NSF-Census Research Network Oct 2014 – June 2017

Coordinates a national research network to advance theory, methodology, and tools for the Census Bureau and the federal statistical system.
Lars Vilhuber, Margaret C Levenstein, Jerome P Reiter
www.nsf.gov/awardsearch/showAward?AWD_ID=1507241

RCN: Coordination of the NSF-Census Research Network July 2012 – Mar 2015

Coordinates a national research network to advance theory, methodology, and tools for the Census Bureau and the federal statistical system.
Alan F Karr, John M Abowd, Jerome P Reiter, Lars Vilhuber
www.nsf.gov/awardsearch/showAward?AWD_ID=1237602

NCRN-MN: Cornell Census-NSF Research Node: Integrated Research Support, Training and Data Documentation Oct 2011 – Sept 2017

Creates tools and methods for documenting, preserving, and sharing restricted-access social science data, supporting reproducible research with confidential datasets.
Lars Vilhuber, John M Abowd, William C Block, Ping Li
www.nsf.gov/awardsearch/showAward?AWD_ID=1131848

Synthetic Data User Testing and Dissemination Sept 2010 – Aug 2015

Develops and tests methods for making detailed government data available to researchers while protecting individual privacy, focusing on synthetic data.
John M Abowd, Lars Vilhuber
www.nsf.gov/awardsearch/showAward?AWD_ID=1042181

TC: Large: Collaborative Research: Practical Privacy: Metrics and Methods for Protecting Record-level and Relational Data July 2010 – June 2017

Bridges statistics and computer science to develop privacy-preserving methods for releasing confidential data, focusing on synthetic data and formal privacy guarantees.
John M Abowd, Johannes E Gehrke, Lars Vilhuber
www.nsf.gov/awardsearch/showAward?AWD_ID=1012593

Joint NSF-Census-IRS Workshop on synthetic data and confidentiality protection, July 2009 Washington, DC July 2009 – June 2011

Workshop to discuss and improve synthetic and partially synthetic public-use micro-data, fostering collaboration between researchers and statistical agencies.
John M Abowd, Ronald Jarmin, Jerome P Reiter, Lars Vilhuber
www.nsf.gov/awardsearch/showAward?AWD_ID=0922494

Social Science Gateway to TeraGrid July 2009 – June 2013

Expands access to large-scale confidential Census Bureau data for social science research, building infrastructure for tera-scale data analysis.
John M Abowd, Lars Vilhuber
www.nsf.gov/awardsearch/showAward?AWD_ID=0922005

The economics of mass layoffs: displaced workers, displacing firms, and causes and consequences Oct 2008 – Sept 2013

Measures the cost of job loss for workers using mass layoffs as a natural experiment, addressing exogeneity and representativeness in labor market research.
John M Abowd, Lars Vilhuber
www.nsf.gov/awardsearch/showAward?AWD_ID=0820349

Synthesis of HRS-SSA linked data Sept 2007 – Aug 2010

This project is a Phase I feasibility study of applying these new methods, called synthetic data, to a subset of the variables in the SSA records that link to the general-use RAND-HRS data. The project has three main components: (1) port a general data synthesizer that was developed at the U.S. Census Bureau for use with SSA data linked to the Survey of Income and Program Participation for adaptation to the HRS/SSA link; (2) synthesize a few variables from the HRS/SSA link and test their usefulness in statistical modeling; (3) perform studies of the statistical disclosure risk associated with linking synthetic SSA data to the RAND-HRS general-release ?le. If the confidentiality-protected data prove scientifically useful and if the statistical disclosure risk can be controlled, then Phase II of the research would synthesize the entire HRS/SSA data link.
John M Abowd, Lars Vilhuber
reporter.nih.gov/search/tfrpgQe35EqhqkYzSl6Rmg/project-details/7223120

Education

Université de Montréal, Economics — Montreal, Canada, 1993 – 1999

Université de Montréal, Economics — Montreal, Canada, 1992 – 1993

Universität Bonn, Economics — Bonn, Germany, 1990 – 1992

Languages

English
Fluent

German
Native speaker

French
Fluent

Portuguese
Advanced (used to be fluent)

Spanish
Conversational (but opinions might differ)