



Data Infrastructure Issues for Evidence-Based Policy and Practice

#LiveAtUrban

Housekeeping

- Event is being recorded and the recording will be posted online afterward.
- Hide captions or adjust settings with the Live Transcript button.
- Speaker biographies are available online at [Urban.org](https://www.urban.org).
- All participants are muted.
- Type your **questions** and **comments** into the Q&A box at any time.
- Please complete the survey at the end of the event.



#LiveAtUrban



Data Infrastructure Issues for Evidence-Based Policy and Practice

#LiveAtUrban



Cornell University
ILR School

A Vision for a 21st-Century National Data Infrastructure: Implications for Evaluations

Erica L. Groshen, Cornell U. –ILR

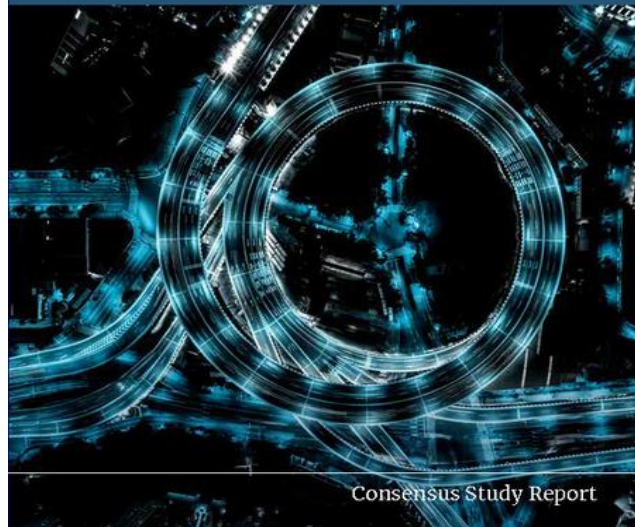
Federal Evaluation Forum

The Urban Institute

December 12, 2022

NATIONAL ACADEMIES
Sciences
Engineering
Medicine

Toward a 21st Century
National Data Infrastructure:
Mobilizing Information for
the Common Good



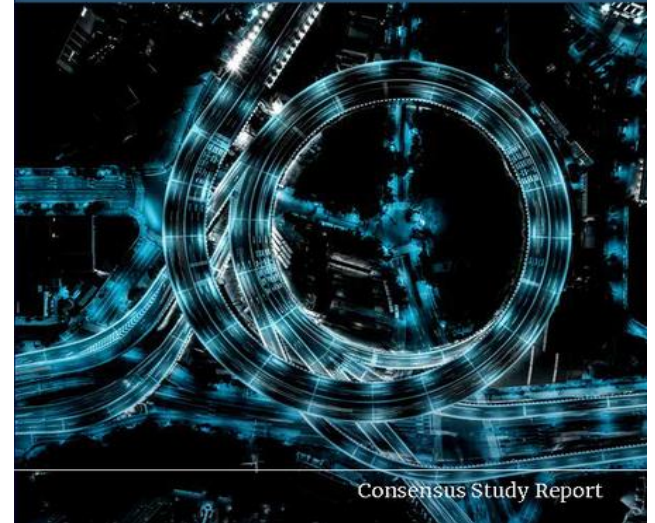
Consensus Study Report

Agenda

- A new national data infrastructure?
- Vision, outcomes, and attributes
- More “blended” data
- Implications for evaluations

NATIONAL ACADEMIES
Sciences
Engineering
Medicine

Toward a 21st Century
National Data Infrastructure:
Mobilizing Information for
the Common Good



Consensus Study Report



“National data infrastructure”???

- Credible statistical info supports democratic society
 - Informs decisions by governments, businesses, and individuals
 - Similar to roads and bridges supporting commerce by facilitating movement of goods, services, and people
 - Examples: Unemployment, inflation, crime, and health statistics inform decisions regarding the economy, society, and peoples’ lives
- Statistical agencies play critical role
 - Collect input data
 - Provide trustworthy statistics equally to all



Building blocks for trustworthy statistics

A red letter 'A' on a wooden block.

ccurate. . . Getting it right

A blue letter 'O' on a wooden block.

bjective. . . Free from bias

A green letter 'R' on a wooden block.

levant. . . Information you can use

A blue letter 'T' on a wooden block.

imely. . . Getting it out quickly

A purple letter 'A' on a wooden block.

ccessible. . . Meeting you where you are



Mid-20th Century data infrastructure

- Official statistics and evaluations rely heavily on surveys
 - Few other sources
 - Sample designed and data collected to fit measurement needs
 - Response rates high
 - Other surveys rare
 - Public service valued
- Social science research and evaluations began with surveys



Overall vision for a new 21st century data infrastructure

- What:
 - Improve quality, timeliness, granularity, and usefulness of national statistics
 - Facilitate more rigorous social and economic research
 - Support evidence-based policymaking and program evaluations
- How: Blending data from multiple sources
 - Overcomes limitations from single sources



What is blended data?

- Collect data from multiple sources
 - Surveys
 - Government agencies
 - Private aggregators and companies
 - Crowdsourcing
- Combine them
 - Match/merge
 - Model
 - Use predictive analytics
- Examples: GDP, productivity, state JOLTS...



Outcomes of a new data infrastructure

1. US information resources are mobilized in a coordinated manner
2. More timely, granular, and useful information
3. Enhanced research insights
4. Enhanced evidence-based policy analysis



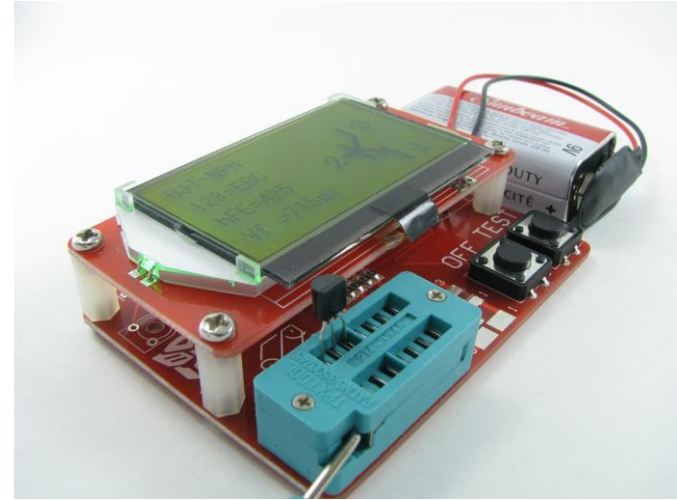
Outcomes of a new data infrastructure, cont'd.

5. Data holders incentivized to share data for statistical purposes
6. Reformed legal framework enhances privacy protections
7. National data infrastructure operates in a transparent, high-trust environment



Key components of new data infrastructure

1. Data assets
2. Data-related technologies
3. People with expertise to manage, use, and understand data
4. Standards governing data
5. Organizations to manage data infrastructure
6. Trust of data subjects and holders



Seven attributes of the vision for the statistical system

1. Safeguards and advanced privacy-enhancing practices to minimize possible individual harm
2. Statistical uses only, for common good information, with statistical aggregates freely shared with all
3. Mobilization of relevant digital data assets, blended in statistical aggregates, providing benefits to data holders, with societal benefits proportionate to possible costs and risks



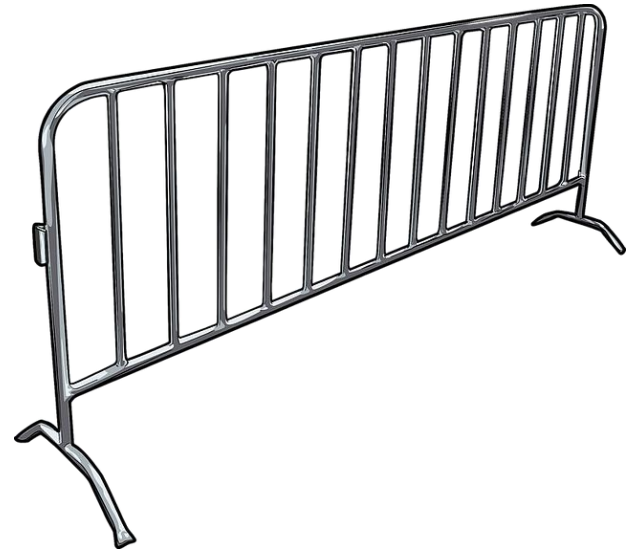
Seven attributes of the vision for the statistical system, cont'd.

4. Reformed legal authorities protecting all parties' interests
5. Governance framework and standards effectively supporting operations
6. Transparency to the public about analytical operations using the infrastructure
7. State-of-the-art practices for access, statistical, coordination, and computational activities, continuously improved to efficiently create more secure, more useful information



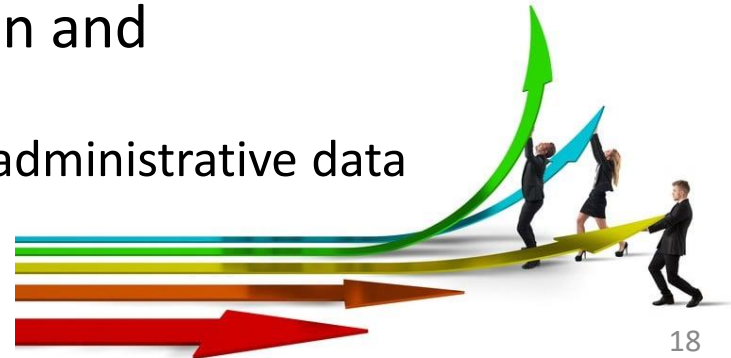
Barriers

- Resources
- Legal impediments
- Learning curve
- Fragmented statistical system
- Privacy concerns



Progress toward the vision

1. Evidence-Based Policymaking Commission, Evidence Act, and ACDEB
 - Increase statistical agency access to federal administrative data
 - Create National Secure Data Service to facilitate blending
2. Statistical agencies, academics, and tech firms pursue initiatives
3. CNSTAT reports develop overall vision and near term steps
 - Include private sector and state & local administrative data



Implications for evaluations

- More data sources available means more need to assess suitability, advantages, and limitations
 - Information (what fields)
 - Coverage (who is missing)
 - Data quality (completeness, accuracy, consistency, history)
 - Matching opportunities
 - Access terms
- **Government administrative data likely to remain key source for evaluations**



Implications for evaluations, contd.

- Data standards for employer-maintained employment and earnings records could be hugely beneficial
 - Enhance state/UI/NDNH wage records with more consistency and new fields (hours, job title, stop & start dates, work & residence location)
 - Enhance access for evaluations
 - Multiple efforts (e.g., BLS wage record pilot, states, JEDx)
- Ensure your voice is at the table to address particular needs
- Reinforce high-trust environment and support resources and response rates for federal surveys



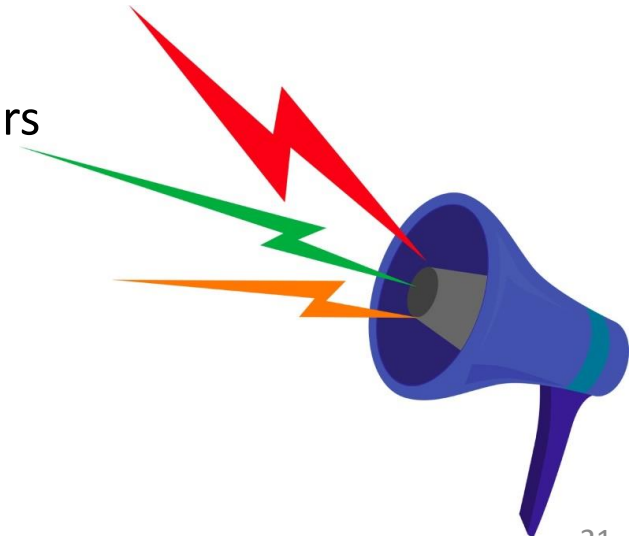
Support independence and modernization for stat agencies

- Decry political interference; independence necessary for trust
- Don't free ride—speak up for evidence and official statistics
 - Attest to trustworthiness; debunk attacks
 - Cite your sources
 - Participate in federal surveys; encourage others
 - Support funding and modernization
 - Join Friends of BLS

www.friendsofbls.org

LinkedIn: “Friends of BLS” group

Twitter: @Friends_of_BLS



Takeaways

- Current national data infrastructure is ill-equipped to meet 21st century data needs
- US needs a data infrastructure that blends data from multiple sources to
 - Improve national statistics
 - Facilitate more rigorous social and economic research
 - Support evidence-based policy and program evaluations
- Progress will require new partnerships, a shared vision, legislative action, consensus regarding activities that move us toward the vision...**and help from evaluation community**



Thank you.

Erica L. Groshen

Cornell-ILR

erica.groshen@gmail.com

<https://nap.nationalacademies.org/catalog/26688>

Full report, Report Highlights and Policy Brief available for viewing/download. Also available are an FAQ section and an interactive website for this report (and, eventually, the two that will follow).

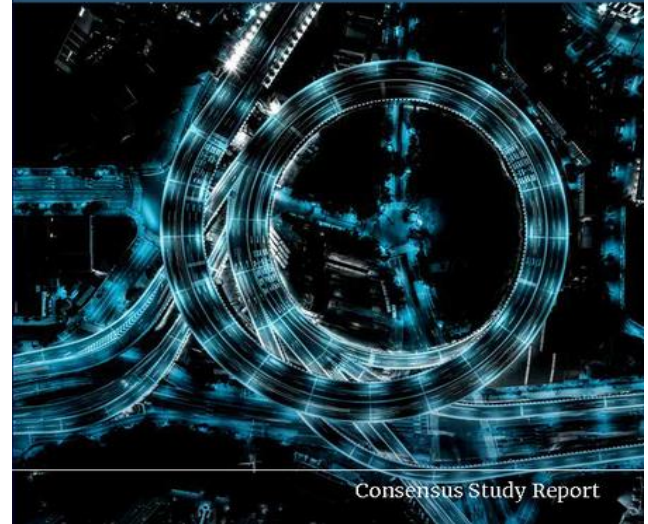


Cornell University
ILR School

NATIONAL
ACADEMIES

Sciences
Engineering
Medicine

Toward a 21st Century
National Data Infrastructure:
Mobilizing Information for
the Common Good



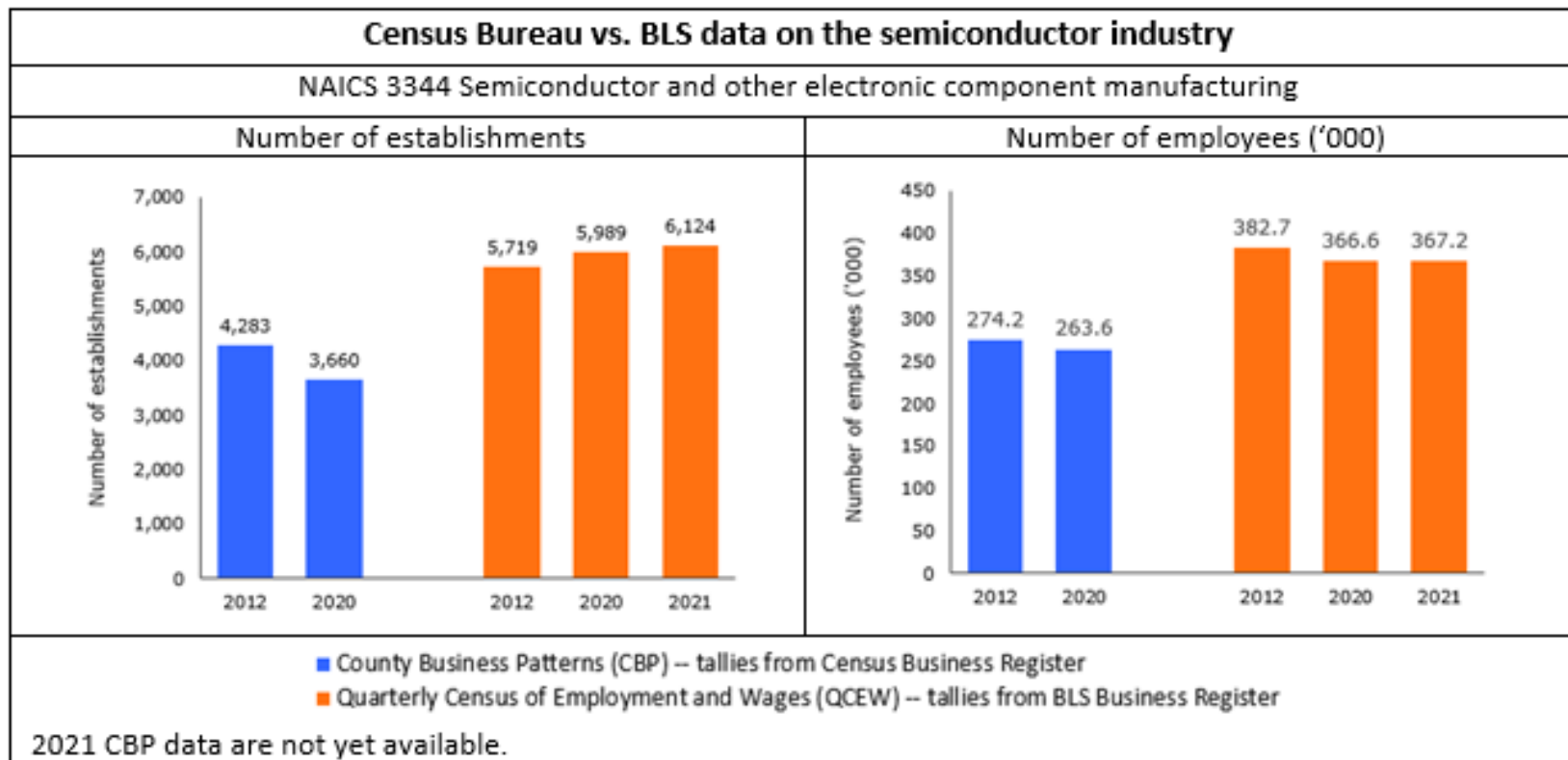
How do private sector and official indicators compare?

- Government and private initiatives are complements, not competitors

Official	Private
<ul style="list-style-type: none">• Transparent, known statistical properties• Access to comprehensive, sensitive government data• High survey response rates• Objective• Long history	<ul style="list-style-type: none">• Proprietary methods• Speedy production• Quick innovation• Access to transactional data• Tailored to special needs



BLS and Census employment discrepancies





Data Infrastructure Issues for Evidence-Based Policy and Practice

#LiveAtUrban