

Toward a 21st-Century National Data Infrastructure

International Total Survey Error Workshop

September 18, 2024



Three Eras of Sample Surveys, circa 2011

- A coalescence of theory and practice (1930-1960)
- Vast expansion of practice (1960-1990)
- Decay of participation; rising costs (1990-)

A fourth era is emerging.

This Moment

Data as national
resources used for
common good
statistical purposes

Paper Forms



Digital Records

Management by
Directive



Real-time Management
by Data/Algorithms



Structured Numeric Data



Unstructured, Text Data



Expensive Compute
and Storage



Cheap Compute and
Storage

Data designed for
official statistics



Data created to run processes





Majority of data on
persons and
enterprises owned by
the private sector



Data sets shared by
acquiring copies



Remote access within
firewalls

Blending relevant
harvested organic
data with survey data

Data Sharing for Statistical Purposes

Requirements

Estimation focus on
coverage, measurement,
modeling error

Requirements

Organizational infrastructure for remote access/blending

Needed Developments

- Legal changes
- Transparency
- Shared governance

Needed Developments

- Model-based estimation as default
- Attention to multiple errors
- Redesign of surveys

Needed Developments

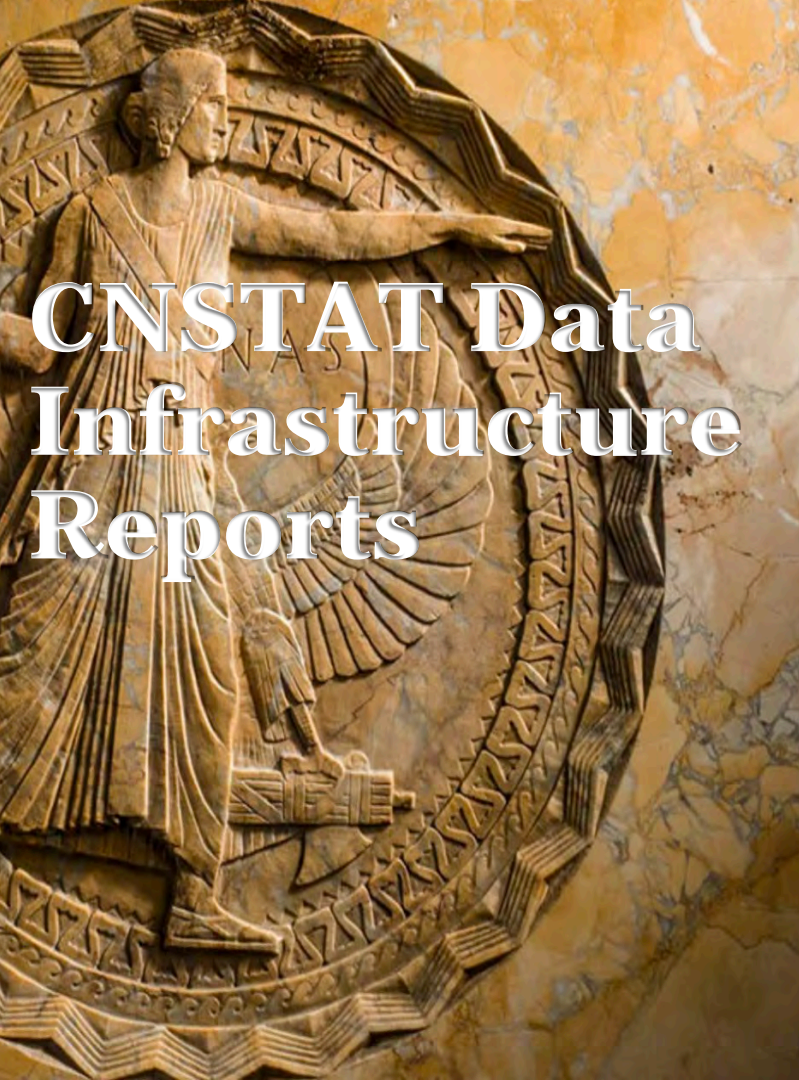
- Centralized ↔ Networked design
- Privacy controls/credibility
- New, more timely information products

Report Conclusion

The United States needs a 21st-century national data infrastructure that blends data from multiple sources to improve the quality, timeliness, granularity, and usefulness of national statistics, facilitates more rigorous social and economic research, and supports evidence-based policymaking and program evaluations.

Speculations regarding TSE

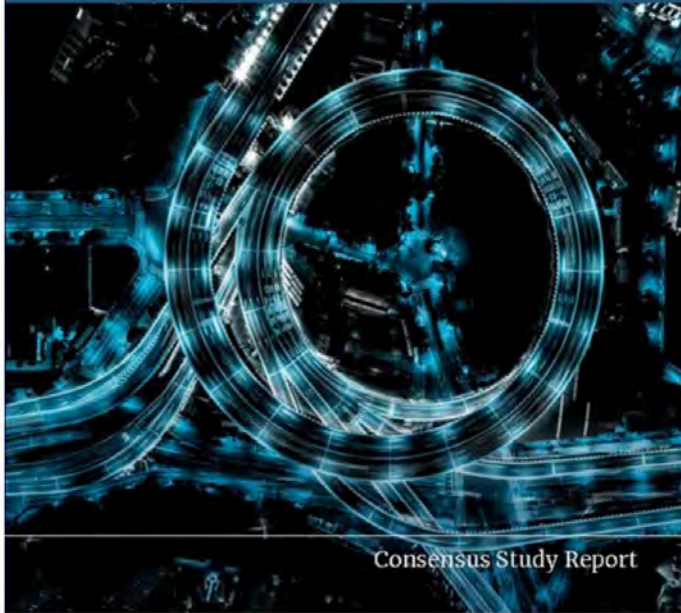
- A renewal of attention on errors of nonobservation
- A measurement error framework that deals with less structured data
- A unified treatment of modeling errors and observational errors



CNSTAT Data Infrastructure Reports

- Report 1: Toward a 21st Century Data Infrastructure: Mobilizing Information for the Common Good
- Report 2: Toward a 21st Century Data Infrastructure: Enhancing Survey Programs by Using Multiple Data Sources
- Report 3: Toward a 21st Century Data Infrastructure: Managing Privacy and Confidentiality Risks with Blended Data

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



Consensus Study Report

Reports available at

<https://nap.edu>

Search for “data infrastructure”

An interactive website for this report can be accessed there. Short overview documents are there.