

NATIONAL
ACADEMIES

Sciences
Engineering
Medicine

Looking to the Future: Toward a 21st Century National Data Infrastructure

Melissa Chiu, Director, Committee on National Statistics



MARCH 3, 2023
COPAFS QUARTERLY MEETING

Agenda

1

This Moment – the Need for a National Data Infrastructure

2

Vision: Scope, Components, and Key Characteristics of a 21st-Century Data Infrastructure

3

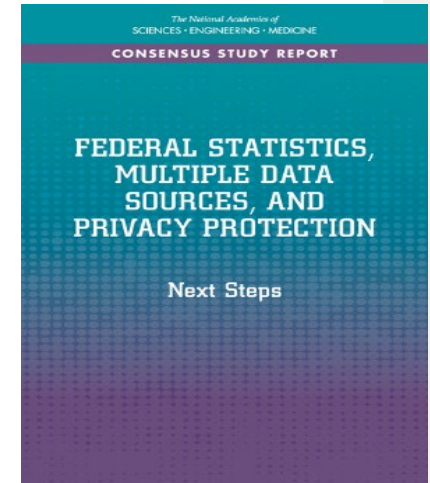
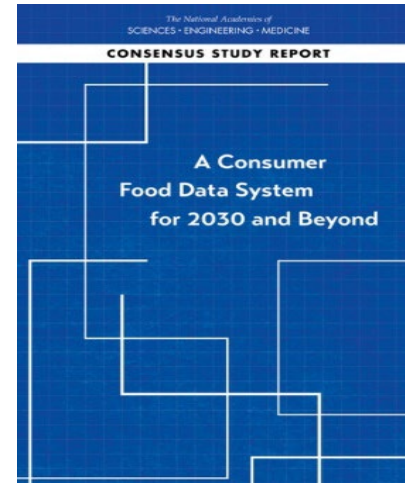
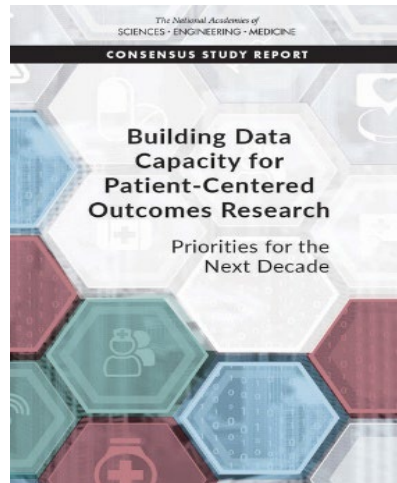
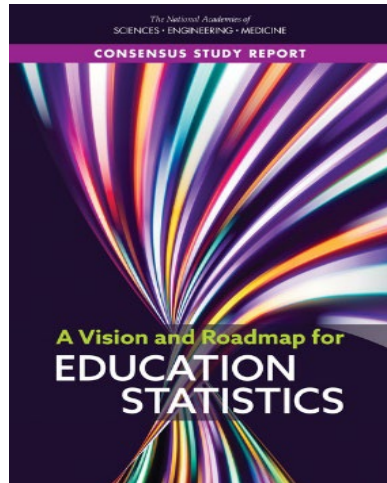
National Data Infrastructure for Specific Areas

Serving the Statistical Community

- **Mission of CNSTAT** is to provide advice to the federal government and the nation grounded in the current best scientific knowledge and practice that will lead to improved statistical methods and information upon which to base public policy.
- Assisting agencies with **specific issues**
- Addressing system-wide and **cross-cutting issues**
- **Expanding our reach** and making our products more widely used and more useful to the statistical **community**

Building a Vision for a New Data Infrastructure

- An Integrated System of U.S. Household Income, Wealth, and Consumption Data and Statistics to Inform Policy and Research



Expanding into the Future

- Statistical system and researchers are seeking more **alternative data sources** for new insights
- Expanding into area that is **vital for the 21st Century**
- **Strategic priority** for CNSTAT

This Moment

Why the U.S. Needs a New National Data Infrastructure

- Survey-centric paradigm no longer sustainable
- Surveys are still key data sources but cannot be relied upon solely anymore; need to be buttressed
- **Explosion of digital data** in many forms
 - Administrative including state and local governments, private sector, social media, sensors, scraped websites, unstructured data, etc.
- Statistical system and researchers seeking more **alternative data sources** for new insights

The Changing Data Access and Use Landscape

- Increasing **blending** of administrative data with survey data
- **Legislative** actions
- Greater interest in **evidence-based decision making**
- Ad hoc uses of **private sector** data
- Challenge is how to create a new data infrastructure that broadly serves the **public good**

We envision a **national** data infrastructure
that creates **national** statistics using
all relevant data resources of the country

CNSTAT Focus on the New National Data Infrastructure

- Strengthen **capacity** of the federal statistical system and research communities
- Launch **convenings** of private sector, state, local, other data holders
- Assist envisioning **technical transformation** from survey-centric to blended data sets
- Focus on **incentives** to business-to-government data sharing for statistical purposes

Vision: Scope, Components, and Key Characteristics of a 21st-Century Data Infrastructure



2

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

- Study 1: Scope, Components and Key Characteristics of a 21st Century Data Infrastructure
 - December 9 & 16, 2021; [Final report published March 3, 2023](#)
- Study 2: The Implications of Using Multiple Data Sources for Major Survey Programs
 - May 16 & 18, 2022; [Report expected March 2023](#)
- Study 3: Approaches for Data Governance and Protecting Privacy
 - Spring 2023; [Report expected fall 2023](#)

NATIONAL ACADEMIES Sciences
Engineering
Medicine

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



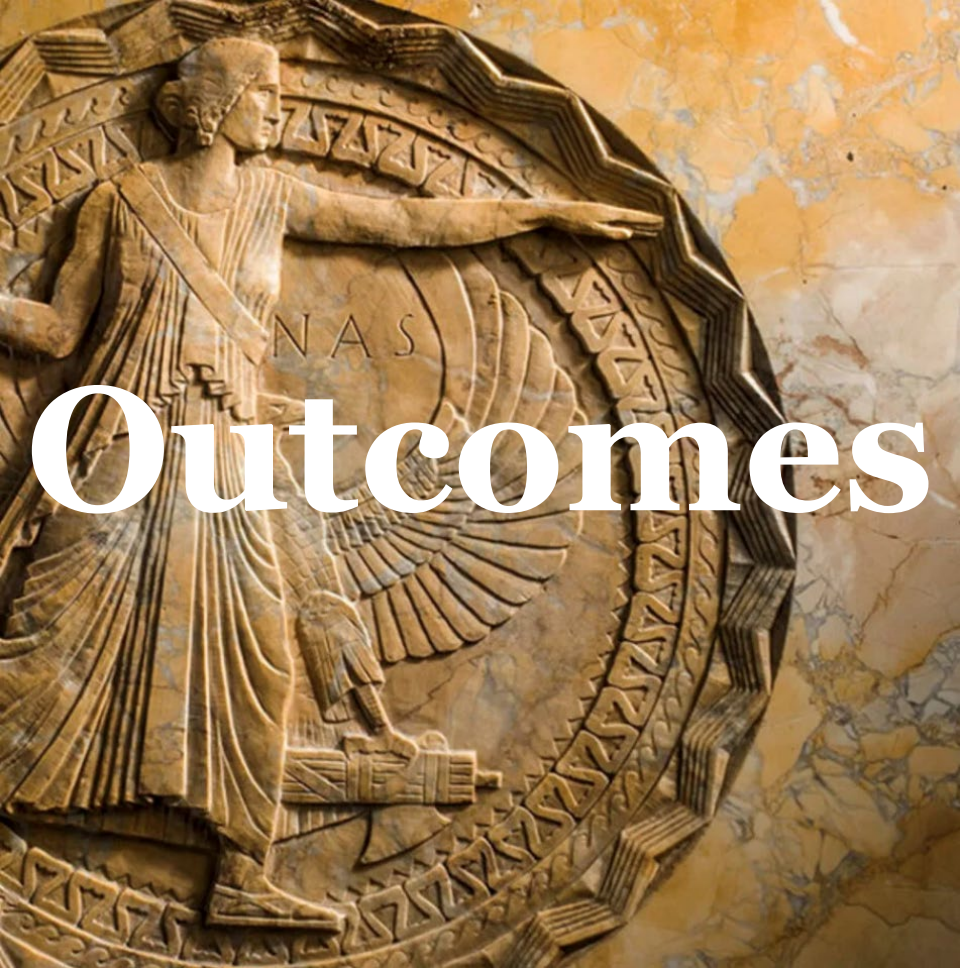
Goals – Interpretation of the Charge

- **Improve national statistics**, social and economic research, and evidence building **for the common good**
- Hold December 2021 **workshop** focused on private sector data use
- **Use additional evidence and resources**, such as
 - Commission on Evidence-Based Policymaking (CEP) report
 - Foundations for Evidence-Based Policymaking Act of 2018 (Evidence Act)
 - Advisory Committee on Data for Evidence Building (ACDEB)
 - previous expert groups
 - international data initiatives
- Synthesize the evidence into **consensus conclusions** that provide a **vision** for a **new national data infrastructure**



Vision

- Access to data assets
- Research needs
- Statistical uses only
- For the common good
- Consider
 - users
 - operations
 - components
 - capabilities



Outcomes

- Information strengthened by blending data, new methods
- Timely, granular, useful
- Expanded access
- Illuminate issues of national importance
- Evidence-based policy analysis to inform governments of all levels
- Incentives to share data
- Legal framework, operation undergirds protection, trust



Attributes

1. **Safeguards** and advanced privacy-enhancing practices
2. **Statistical uses only**, for common good information
3. Mobilization of relevant digital data assets, **blended** in aggregates
4. Reformed **legal authorities**
5. **Governance** framework and standards
6. **Transparency** to the public
7. State-of-the-art practices

Takeaways



The United States needs a 21st century national data infrastructure that

- **blends data from multiple sources** (*including surveys*)
- 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.

Data from

- **federal, state, tribal, territory, and local governments;**
- the **private sector;**
- **nonprofits** and **academic** institutions; and
- **crowdsourced** and **citizen-science** data holders

are crucial components of a 21st-century national data infrastructure.

(in addition to surveys)

Data sharing is **incentivized** when

- **all data holders** enjoy **tangible benefits** valuable to their missions, and when
- **societal benefits** are **proportionate** to possible **costs and risks**.

It is **ethically necessary** and technically possible to

- **preserve privacy** and **fulfill confidentiality pledges** regarding data while simultaneously
- **expanding the statistical uses** of diverse data sources.

Trust in a new data infrastructure **requires**

- **transparency of operations** and
- **accountability** of the operators,
with
- ongoing **engagement** of stakeholders.

Effective data governance is critical and should be **inclusive** and **accountable**;

governance policies and standards facilitating interoperability **include input** from **key stakeholders and oversight bodies.**

The **operations** of a new data infrastructure would **benefit** from the **inclusion of continually evolving**

- **practices,**
- **methods,**
- **technologies,** and
- **skills,**

to **ethically leverage** new technologies and advanced methods.

Multiple organizational structures can support a new data infrastructure but organizational **options** and **questions warrant further study.**

Legal and regulatory changes are **necessary** to achieve the full promise of a 21st century national data infrastructure. Example actions:

- Establish legislation underlying **secure data sharing**
- Implement **Evidence Act regulations**
- Identify **state-related legislation/regulatory priorities** derived from the Commission on Evidence-Based Policymaking
- Develop legislative strategy for **data synchronization bill**
- Identify legal options that would **incentivize data holders** to share data for statistical purposes.

Takeaway

1

0

Building a data infrastructure requires **identifying short- and medium-term activities.**

*Report identifies **40** short-term and **25** medium-term activities that moves the U.S. toward the panel's vision.*

Report Resources

Publication and resources

Report

Report highlights

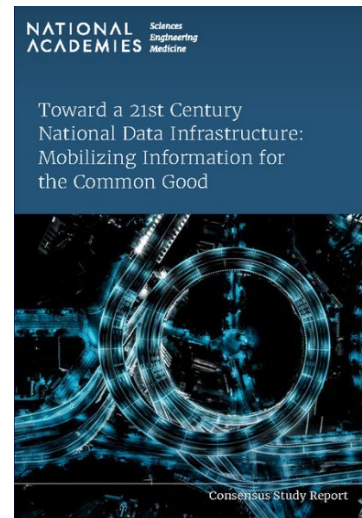
Policy brief

Private sector data brief (*coming soon*)

Interactive site

Accessible, digestible format

FAQ section



Toward a 21st Century



Future CNSSTAT Data Infrastructure Reports

- Report 2: The Implications of Using Multiple Data Sources for Major Survey Programs (*expected March 2023*)
- Report 3: Approaches for Data Governance and Protecting Privacy (*workshop spring 2023, report expected fall 2023*)

National Data Infrastructure for Specific Areas

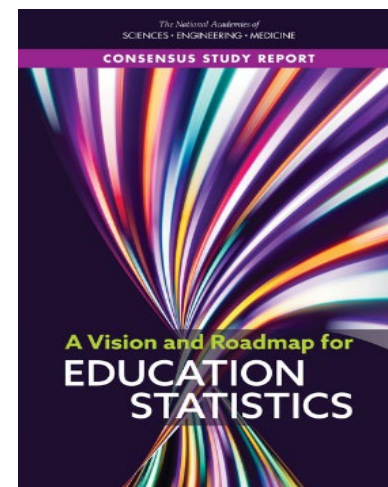


3

A Vision and Roadmap for Education Statistics

The National Center for Education Statistics (NCES) as a leader of the education data ecosystem

- Charge from Department of Education's Institute of Education Sciences (IES) to study the National Center for Education Statistics (NCES)
 - **Vision and roadmap for next 7 years**
- **Education data landscape changing rapidly**
 - Student populations more diverse
 - Explosion of data sources
 - Strong interest in evidence-based decision making
 - Foundations for Evidence-Based Policymaking Act of 2018 (Evidence Act)
- Published April 2022 ([report and resources](#))



Maximize NCES's Unique Value for Evidence Building

Increasing value of data products and services, Rec. 2-3

- Mission expanded by Evidence Act – new authority and mandate to **acquire data** from federal agencies and to **facilitate access to data**
- Collaborate **within ED – Statistical Official connects** Chief Data Officer, Evaluation Officer by turning data into high-quality information fit to inform policy, decision making
- Build stronger partnerships **with NCEE, NCER, NCSER – collaboratively decide**
 - the priority datasets NCES should acquire and link, and
 - providing access to external researchers as a **force multiplier** for answering Learning Agenda questions (also Rec. 4-5)
- **Collaborate** with data-holding agencies and organizations (Rec. 4-2)
- **Support state and local** education agencies in data access and linkage (Rec. 4-4)

Expand Data Sources to Gain New Insights

Acquiring data strategically to improve efficiency, timeliness, and relevance,
Recs. 2-5, 2-6

- Use and develop **new data sources**, data linkage, especially administrative data
- Study quality, fitness for use, bridging, and series continuity
- Expand responsible, secure access and linkage tools
- **Partner within IES, ED, and with other federal statistical agencies**
 - Data-science methods to harness and link data
 - Strengthen impact of techniques
 - Develop useful products and processes (Rec. 4-2)
- Modernize consent language, permit secondary uses, privacy-protected linkage

Support State and Local Education Agencies

Scaling NCES's impact by strengthening state and local capacity for evidence building, Rec. 4-4, Con. 4-1

- Set priority activities, goals, outcomes for Statewide Longitudinal Data Systems (SLDS) awards
 - Award proposals that are high value and meet these criteria, e.g., for –
 - **Infrastructure** investments for **high-value data linkages** (e.g., across social domains) to build new insights
 - **States** to share data, facilitate data access, provide technical assistance, analytic assistance **to local education agencies** (LEAs)
 - **States** to facilitate collaborations among regions or groups of LEAs
 - Generating products and tools useful to LEAs and SEAs
 - Creating **shareable data** to benefit all states

Ongoing project: An Integrated System of U.S. Household Income, Wealth, and Consumption Data and Statistics to Inform Policy and Research

Modernizing measurement and improving data infrastructure

- The nation's **disparate federal statistics** make it **difficult to accurately measure** income and wealth inequality, economic well-being for U.S. households and families
- Provide guidance on producing **high-quality estimates, information**, considering:
 - potential for using **multiple data sources**, including surveys, state and federal administrative records, and commercial data, and modeling
 - needed **quality improvements** for relevant data collection programs
 - the **cost and respondent burden** for collecting the data
 - **legal, administrative, and other barriers** to creating an integrated system of statistics
- Publication expected fall 2023 (watch [project page](#))

Concluding Thoughts

Principle 5 Continual improvement and innovation

from [*Principles and Practices for a Federal Statistical Agency*](#)

Federal statistical agencies must
continually seek to improve and innovate their
processes, methods, and statistical products
to **better measure an ever changing world**.

- There are many challenges and opportunities in the federal statistical system
- CNSTAT is here to help the federal statistical system harness the opportunities and address the many challenges ahead

Melissa C. Chiu

CNSTAT Director

mchiu@nas.edu

Brian Harris-Kojetin

Senior Scholar

bkojetin@nas.edu

CNSTAT website, events, and projects <http://www.nationalacademies.org/cnstat>

All CNSTAT reports available from National Academies Press (free)

<https://nap.nationalacademies.org/author/CNSTAT/>

[P&P main page](#) and [interactive site](#)

[Sign up for email updates](#)

Select DBASSE – “National surveys and statistics”

Thank you!



Celebrating 50 years of Service
to the Statistical Community

1 9 7 2 – 2 0 2 2