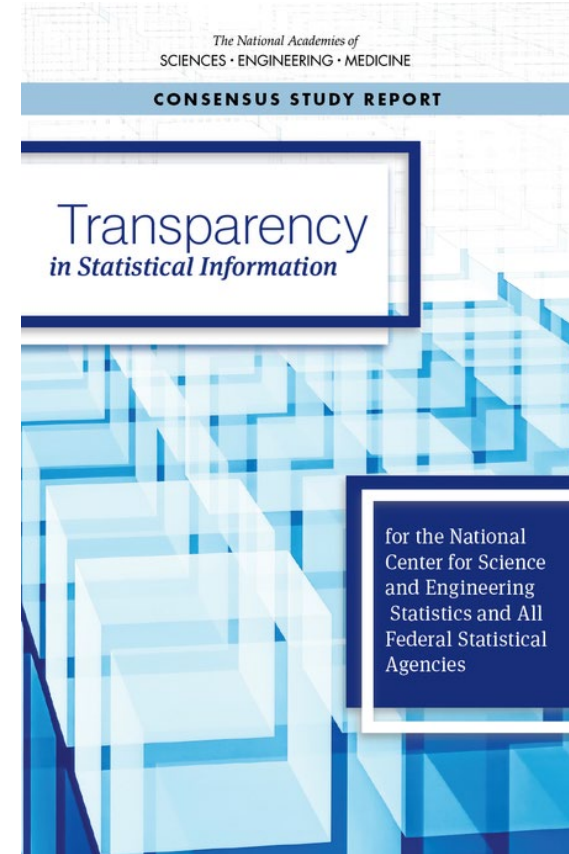


Transparency of Official Statistics: *Part of the recent emphasis on reproducibility of scientific activities*

FCSM Session G-2
October 29, 2022



Panel Members

Dan Kasprzyk (NORC at the University of Chicago)

Philip Ashlock (GSA)

David Barraclough (OECD)

Christopher Chapman (NCES)

Dan Gillman (BLS)

Linda Jacobsen (Population Reference Bureau)

H.V. Jagadish (University of Michigan)

Frauke Kreuter (JPSM and the University of Mannheim)

Margaret Levenstein (ICPSR)

Peter Miller (Northwestern)

Audris Mockus (University of Tennessee)

Sarah Nusser (Iowa State)

Eric Rancourt (Statistics Canada)

Lars Vilhuber (Cornell University)

With special thanks to sponsor NSF NCSES and CNSTAT staff **Michael Cohen & Michael Siri**





PANEL ON REPRODUCIBILITY AND TRANSPARENCY

Transparency: Current Contexts & Future Directions

Sarah M Nusser

Iowa State University

Transparency in Official Statistics

Transparency

... the provision of sufficiently detailed documentation of all the processes of producing official estimates. (p. 2)

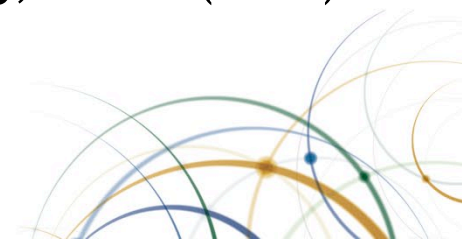


Transparency in Official Statistics

Transparency:

the provision of sufficiently detailed documentation of all the processes of producing official estimates

Archived “documentation” includes survey program description, data (inputs, analysis, public use, ...), data treatments (editing, imputing, processing, linking, privacy, ...), official estimates, assessments and indicators of quality, *et al.* (Ch 7)



Transparency in Official Statistics

Goal

Enable consumers of federal statistics to accurately understand and evaluate how estimates are generated

(reuse)



Transparency in Official Statistics

Goal: enable consumers of federal statistics to accurately understand and evaluate how estimates are generated

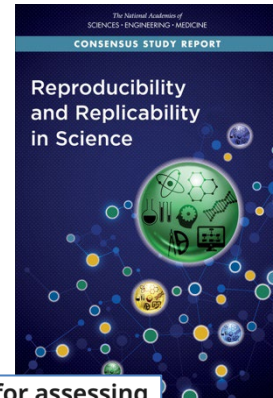
Federal statistical system and its policies have promoted transparency for decades.

So why now?



Paradigm Shifts in Scholarship

- **Focus on Openness and Rigor for scientific activities has evolved and deepened**
 - Long-standing practices that support scrutiny, debate, self-correction, new discoveries being modernized
 - Scholarly articles now expected to include access to data, metadata, methods, software, and other outputs
 - International principles for data sharing widely accepted and maturing through implementation
 - Findable, Accessible, Interoperable, Reusable (FAIR)
 - Greater emphasis on research integrity, reproducibility and replicability in scholarship



Paradigm Shifts in Scholarship

- **Transparency is the primary mechanism to fostering & demonstrating rigor, advancing knowledge, ensuring equitable access**
- **Changing how academic researchers conduct research**
 - Higher expectations of openness
- **2022 White House OSTP public access policy expands prior mandates by removing embargos and broadening data & agencies subject to mandate**
 - With protections for confidential, proprietary, national security concerns



EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

August 25, 2022

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Dr. Alondra Nelson *Alondra Nelson*
Deputy Assistant to the President and Deputy Director for Science and Society
Performing the Duties of Director
Office of Science and Technology Policy (OSTP)

SUBJECT: Ensuring Free, Immediate, and Equitable Access to Federally Funded Research

This memorandum provides policy guidance to federal agencies with research and development expenditures on updating their public access policies. In accordance with this memorandum, OSTP recommends that federal agencies, to the extent consistent with applicable law:

1. Update their public access policies as soon as possible, and no later than December 31st, 2025, to make publications and their supporting data resulting from federally funded research publicly accessible without an embargo on their free and public release;
2. Establish transparent procedures that ensure scientific and research integrity is maintained in public access policies; and,
3. Coordinate with OSTP to ensure equitable delivery of federally funded research results and data.

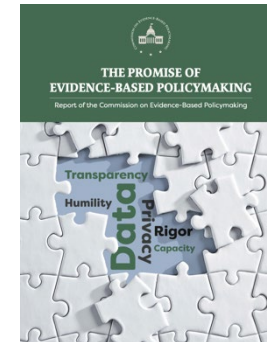
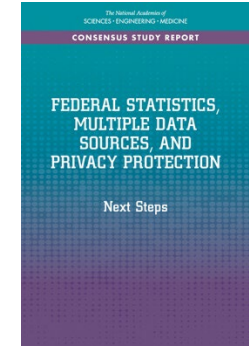
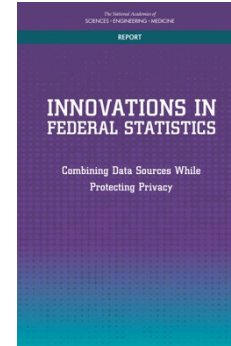
1. Background and Policy Principles

Since February 2013, federal public access policy has been guided by the *Memorandum on Increasing Access to the Results of Federally Funded Research* (2013 Memorandum).¹ Issued by the White House Office of Science and Technology Policy (OSTP), the 2013 Memorandum directed all federal departments and agencies (agencies) with more than \$100 million in annual research and development expenditures to develop a plan to support increased public access to the results of federally funded research, with specific focus on access to scholarly publications and digital data resulting from such research.

Nearly ten years later, every federal agency subject to the 2013 Memorandum has developed and implemented a public access policy in accordance with its guidance.² As a result, the American

Other Transformations

- **Continued innovation in approaches, standards & tools for official statistics**
 - Expanding use of non-survey data sources in creating statistical products
 - International official statistics community practice and standards development
- **Government policy for data sharing and reuse**
 - Evidence-based Policy Making Act
 - Federal Data Strategy
 - National Secure Data Service development



What is our Current State?

- **Transparency practices vary widely**
 - Across agencies and throughout the life cycle
- **Documentation not always prioritized**
 - Internal documentation in better shape than external documentation
- **User and their needs not always considered**



Federal Statistics & Transparency

Report calls upon the federal statistical system to update its underst

anding and implementation of transparency

Documenting
Quality &
Processes
for Users

User needs
Methodologies
Quality assessments

Metadata
&
Standards

Describing data
Standards
Collaborative systems

Archiving
Practices

Internal, external use
Where, when
Paradata



What is to be Gained?

(beyond alignment with shifts in current practices)

- Enables understanding, quality and resilience in operations (**efficiency and effectiveness**)
- **Promotes trust and credibility** with users, public
- Undergirds the **ability to innovate and improve methods**
- Supports **reuse and potential integration of data for new applications**



How? (system)

- **Dedicated leadership from stewarding bodies**
(e.g., OMB, ICSP, agencies ...)
 - **Prioritize & continually reinforce** updating of transparency practices
 - Provide **resources & training** to adopt practices and tools
 - Organize **pilots as collaborations** across agencies
 - Encourage **staff involvement** in emerging standards and tool development communities



How? (agency/program practice)

- **This is a long game - report offers several smaller steps**
- **Proactive planning** for data management and archival (Ch 3)
- **Understanding user documentation needs and usability** (Ch 4)
- **Learning to use systems for collaborative software development & document workflow and methodologies** (Ch 4) Talk 2:
Lars Vilhuber
- **Adopting incremental approaches to metadata standards development** (Ch 5) Talk 3:
Dan Gilman
- **Establishing contract terms to support transparency and openness in deliverables** (Ch 6)



Best practices (Ch 7)

- Tables outlining information needed to support transparency in production of official statistics
 - Internal vs external users
 - Varying data types (survey, admin, digital trace)
 - Data integration
 - Paradata
 - Data archival

Table 7-1 Documenting Basic Elements of a Statistical Program

Information to retain or archive ^a	To be available internally, to program staff	To be available externally, to the public
<p>The Estimation Problem: The concept or concepts this official statistical program is measuring (e.g., the percentage of U.S. households in poverty at the county level in April 2020, where poverty is defined using ... and households are defined as ...). This should include:</p> <ul style="list-style-type: none">✓ the precise definition of the key concepts✓ the relevant population✓ the levels of aggregation at which the estimates are provided✓ the relevant time period covered by the estimates✓ the nature of the products, e.g., tabulations, confidential, microdata, or public-use files.	<p>Description should be updated regularly, versioned, and curated^b for easy access. When concepts or definitions change, the data documentation for specific data products should be able to precisely connect to the version of the description that applies.</p>	<p>Description should be on the appropriate Web page for access by the public, updated regularly, versioned, and curated. When concepts or definitions change, the data documentation for specific data products should be able to precisely connect to the version of the description that applies. In addition, the relationship between the old and new versions should be explained for the benefit of the public.</p>



PANEL ON REPRODUCIBILITY AND TRANSPARENCY

Transparency: Current Contexts & Future Directions

Sarah M Nusser

Iowa State University

nusser@iastate.edu