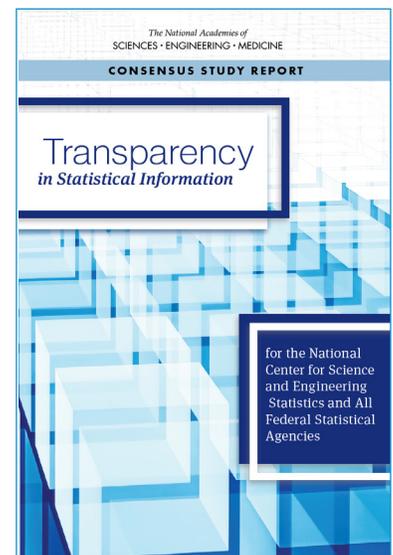


Transparency in Statistical Information for the National Center for Science and Engineering Statistics and all Federal Statistical Agencies

Official statistics, those produced by the federal statistical system, are critical to informing policy decisions on issues ranging from the state of the economy, energy access and use, environmental health, the state of social services, the state of public education, and the state of employment, among many others. These official statistics, which are utilized by government agencies, the private sector, the media, researchers, and the public—given their importance—must be trustworthy and accessible as they are used to guide various public and private actions, such as associated decisions on the allocation of federal funds. In the United States, thirteen principal statistical agencies, as well as units and programs in many other governmental departments and offices, produce official statistics. Their work is coordinated by the chief statistician in the U.S. Office of Management and Budget (OMB), which convenes the Interagency Council on Statistical Policy (ICSP). This council can issue guidances on federal statistical policy.

Complex data collection, processing, and estimation methods are used in the production of official statistics. These methods, which are based on the application of current statistical science, can be challenging for agencies to document and for users to understand, especially given the breadth of user perspectives and needs. Many of the steps underlying the production of official statistics are not currently transparent, even though efforts are made to make their construction clear to their user communities. Transparency, knowing how a set of official statistics are produced from the design of data collection through data processing and estimation, is essential to enable consumers of these statistics to know how these official estimates are generated and their fitness for use for various applications. Along the same lines, assessment of the reproducibility of a set of official statistics, that is the ability to repeat the entire process of planning and design, data collection, data treatment, and statistical estimation, is also needed to ensure that the steps underlying the production of a set of official statistics was as described, and not selected to favor any set of stakeholders, and is consistent with state-of-the-art statistical methods.

Recognizing the need to enhance the transparency and reproducibility of its statistics, the National Center for Science and Engineering Statistics (NCSES) asked the National Academies of Sciences, Engineering, and Medicine's (NAEM) Committee on National Statistics (CNSTAT) to establish a consensus Panel on Transparency and Reproducibility of Federal Statistics to study issues of documentation and archiving of NCSES statistical data products. The panel's report, *Transparency in Statistical Information for the National Center for Science and Engineering Statistics and All Federal Statistical Agencies*, offers recommendations to federal



leadership about the need for agencies who produce federal statistics to establish transparent processes and methods. The report also discusses how federal statistical agencies can strengthen and improve their ability to preserve and archive data; identify and incorporate tools to facilitate the documentation of methodological processes; and lead in framing systemwide standards for the documentation and archiving of official statistics. Also described are priority areas for NCSES to ensure data can be more transparent and helpful to users as well as best practices that federal agencies can use to determine the extent to which their policies are transparent.

TRANSPARENCY: A NECESSARY COMPONENT OF FEDERAL STATISTICS

The panel found that agencies of the federal statistical system are reasonably complete in what they retain internally for their agency; however, these agencies are not as transparent with respect to what is available to the public regarding the treatment of the data and the methodologies used. Also, agencies often do not provide information to the public about what access is permitted and available via secure environments regarding data sets used. To address this, the panel recommends that leadership at the OMB, the ICSP, the NCSES, and all agencies that produce federal statistics should establish transparency of processes and methods as a high priority and continuously reinforce this priority to their staffs.

IMPROVING ARCHIVING PRACTICES

Archiving is an important aspect of government data collection. Federal law requires agencies to maintain record schedules, and to preserve records considered to have long term value in the National Archives of the United States, and to destroy records lacking value after a specified period. To improve current archiving practices, the panel recommends that the agencies that produce federal statistics should fully comply with federal record schedules, ensuring that the input data sets that can legally be retained, as well as official estimates that are produced, are archived in the National Archives and Records Administration. Also, Federal statistical programs who utilize survey data should make information available about the process used to collect that survey data to help users assess the quality of the survey inputs.

ADOPTING NEW TOOLS TO DOCUMENT METHODOLOGICAL PROCESSES

There are several information technology tools widely used in academia and industry that can facilitate the development and documentation of government data collection and treatment. These tools should be examined for their utility in supporting greater transparency of official government statistics. Thus, the Panel recommends that agencies that produce federal statistics, including NCSES, review and make a priority of adopting modern information technology tools that assist in collaborative software development and documentation of workflow and methodology. These tools can support federal agency efforts to make their transparency efforts more efficient.

ADDRESSING ISSUES CONCERNING TRANSPARENCY AND METADATA STANDARDS

The panel identified standards for federal statistical agencies to guide the use and development of metadata, or data that describes other data. It noted that the federal statistical system as a whole, including NCSES, needs to lead in framing systemwide standards for the documentation and archiving of official statistics. This work should review and incorporate international standards and processes that are widely in use, and it should provide training for relevant employees.

To this end, the panel recommended that the ICSP develop and implement a multi-agency pilot project to explore and evaluate employing existing metadata standards and tools to accomplish data sharing, data access, and data reuse. Additionally, the ICSP should: 1) prioritize and emphasize the importance and benefits of federal statistical agency staff engaging in international metadata standards and tool development, and 2) organize a discussion among statistical agencies that leads to an effective, coordinated, and accountable approach for staff in the group of agencies that produce federal statistics to contribute to international metadata standards and tool development.

RECOMMENDATIONS FOR NCSES

The panel reviewed NCSES's general publication standards as well as information available on NCSES's web pages for four programs to develop recommendations for the agency concerning transparency and reproducibility. These programs included the Business Enterprise Research and Development Survey (BERD), Survey of Earned Doctorates (SED), Higher Education Research and Development Survey (HERD), and Early Career Doctorates Survey (ECDS).

In all four cases, the survey design was well described; however, more than one program requested users to go to a technical report to learn more about how data were treated. It was not always clear where the input data or resulting official estimates were archived. Based on this information and suggestions from expert users of NCSES data, the panel identified several priority areas for the agency to improve transparency of its data collection and use. For example, agencies that produce federal statistics should word their contracts and inter-agency agreements that involve planning, collection, processing, and analysis of data products created by the agency so that any information adhering to the agency's transparency expectations obtained by the contractor or federal agency is obligated to be provided to the sponsoring agency unless constrained by legal or proprietary considerations.

In addition, the panel recommended several actions on the part of NCSES senior management, including:

- Renewing their emphasis on the timely production, distribution, and accessibility of methodology reports, data quality assessments, and quality profiles for each program.
- Monitoring and reinforcing agency policy for archiving input data as well as official statistics themselves.
- Investigating how programs use paradata or process data, such as the number of interview contact attempts, including identifying programs that would benefit from the use of paradata.
- Identifying what paradata are valuable to maintain and determining the length of time such data can be made available to researchers.

NCSES should also take the following steps to help users interact with data used to develop government statistics, including:

- establishing ongoing data user groups with contact mechanisms;
- establishing a repeated survey of users as to their current experiences in accessing and using agency data and how estimates could be presented to facilitate time series and cross-sectional analyses;
- ensuring consultations with data users prior to making changes in dissemination systems and statistical programs;
- creating a mechanism that enables members of a statistical program's user group to communicate directly with one another;
- organizing regular meetings with a broad user community representation; and
- identifying ways to improve the transparency, accessibility, and usability of NCSES estimates, data products, documentation, and dissemination systems, including the structure and navigation of the agency's Website.

DEVELOPING SYSTEM-WIDE BEST PRACTICES

The panel also identified best practices that NCSES and other federal statistical agencies can implement to determine the extent to which their policies are transparent. These best practices provide direction for agencies on what information they should retain internally and what should be made available to the public. These include establishing documentation requirements for the basic elements of a statistical program, as well as determining what information to retain and make public for programs that utilize survey, administrative, and digital trace data, as well as issues that arise in data integration and archiving of data used in the production of official statistics.

The panel also recommends that NCSES and other agencies that produce federal statistics should document their data collection methods, their data treatments, their estimation methodologies, and assessments of the quality of their official estimates, and archive their input data sets and their official estimates to support reproducibility and later reuse.

Additionally, senior management at the agencies that produce federal statistics should provide resources and staff support to help transform their current processes, including incorporating the use of data sharing and reuse through use of metadata tools and standards.

Finally, the panel recommends that agencies that produce federal statistics receive additional funds to acquire the necessary training and information technology assistance, as well as cover any increased operational costs, to modify current processes to improve documentation and archiving in support of the greater transparency of official statistics.

PANEL ON TRANSPARENCY AND REPRODUCIBILITY OF FEDERAL STATISTICS FOR THE NATIONAL CENTER FOR SCIENCE AND ENGINEERING STATISTICS

DANIEL KASPRZYK (*Chair*), NORC at the University of Chicago; **PHILIP ASHLOCK**, GSA Technology Transformation Services, General Services Administration; **DAVID BARRACLOUGH**, Practices and Solutions Division, Organisation for Economic Cooperation and Development; **CHRISTOPHER CHAPMAN**, Statistics Sample Surveys Division, National Center for Education Statistics; **DANIEL W. GILLMAN**, Office of Survey Methods Research, U.S. Bureau of Labor Statistics; **LINDA A. JACOBSEN**, Population Reference Bureau, Inc.; **H. V. JAGADISH**, Department of Computer Science and Engineering, University of Michigan; **FRAUKE KREUTER**, Joint Program in Survey Methodology, University of Maryland; **MARGARET LEVENSTEIN**, Inter-university Consortium for Political and Social Research, University of Michigan; **PETER V. MILLER**, U.S. Census Bureau (retired); **AUDRIS MOCKUS**, Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville; **SARAH M. NUSSER**, Center for Survey Statistics and Methodology, Iowa State University; **ERIC RANCOURT**, Modern Statistical Methods and Data Science Branch, Statistics Canada; **LARS VILHUBER**, Department of Economics, Cornell University; **MICHAEL L. COHEN**, Study Director.

For More Information . . . This Consensus Study Report Highlights was prepared by the Committee on National Statistics based on the Consensus Study Report, *Transparency in Statistical Information for the National Center for Science and Engineering Statistics and all Federal Statistical Agencies* (2021). The study was sponsored by the National Science Foundation. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any organization or agency that provided support for the project. Copies of the Consensus Study Report are available from the National Academies Press, (800) 624-6242 or <https://www.nap.edu/26360>.

Division of Behavioral and Social Sciences and Education

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The nation turns to the National Academies of Sciences, Engineering, and Medicine for independent, objective advice on issues that affect people's lives worldwide.

www.nationalacademies.org