



Committee Charge from the National Center for Science and Engineering Statistics (NCSES)

Emilda B. Rivers
Director, NCSES

1ST Meeting of the
Committee on Transparency and Reproducibility of Federal Statistics
for NCSES

May 21, 2019

National Center for Science and Engineering Statistics
Directorate for Social, Behavioral and Economic Sciences
National Science Foundation
<http://ncses.nsf.gov>

What to expect

- About NCSES
- Impetus for this CNSTAT Panel Study
- Charge to the Panel



About NCSES...

One of 13 Principal Federal Statistical Agencies





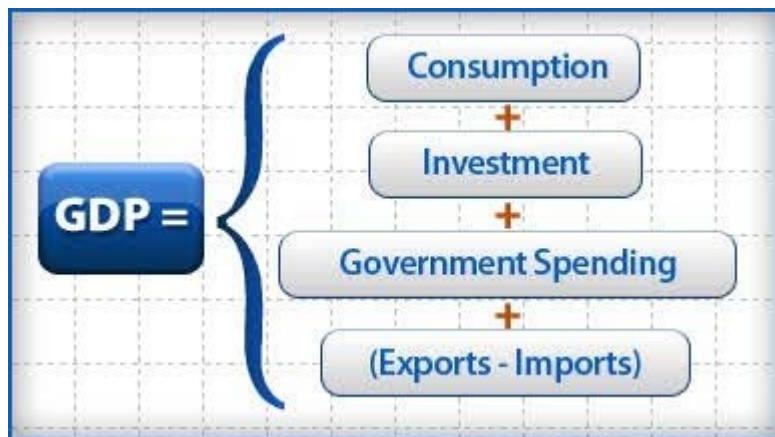
Organizational Placement of NCSES

National Science Foundation

Directorate for Social, Behavioral and Economic Sciences

National Center for Science and Engineering Statistics

National Use of NCSES Data



In 2013 introduced methodological change, R&D part of Investment



Carnegie Classification uses NCSES
R&D expenditures and research staff data



Statistical and Methodological Research Agenda

OPEN FOR BUSINESS

Researchers estimate that Python, Javascript, and R contribute billions to GDP

By Dan Kopf • November 27, 2018



The image shows a person's hand resting on a dark laptop keyboard, with a finger pointing towards the screen. The screen displays a news article from Quartz about the economic impact of programming languages like Python, JavaScript, and R. The article title is "Researchers estimate that Python, Javascript, and R contribute billions to GDP". The Quartz logo is visible at the bottom of the screen.

Latest Featured Obsessions @ Editions

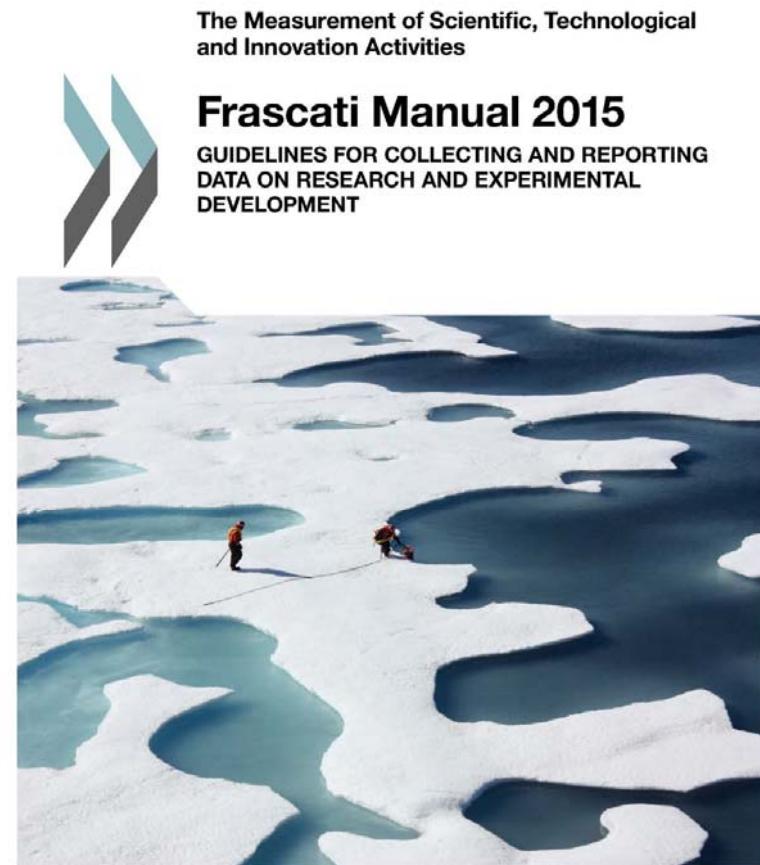
qz.com/1474770/researchers-estimate-that-python-javascript-and-r-contribute-billions-to-gdp/ 1/7

2018 Python, Javascript, R and Julia contribute billions to GDP — Quartz

Open source software is not well counted in GDP.

Gross domestic product, perhaps the most commonly used statistic in the world for evaluating economic progress, [has some issues](#).

International Collaborations





National Center for Science and Engineering Statistics

Administrative and Program Operations Group (APOG)

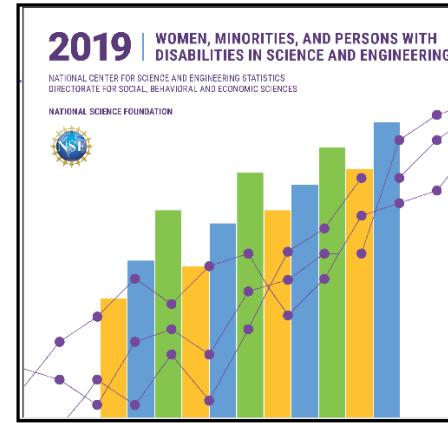
Statistics and Methods Program (SMP)

Information and Technology Services Program (ITSP)

Research and Development Statistics Program (RDS)

Human Resources Statistics Program (HRS)

Science and Engineering Indicators Program (SEI)

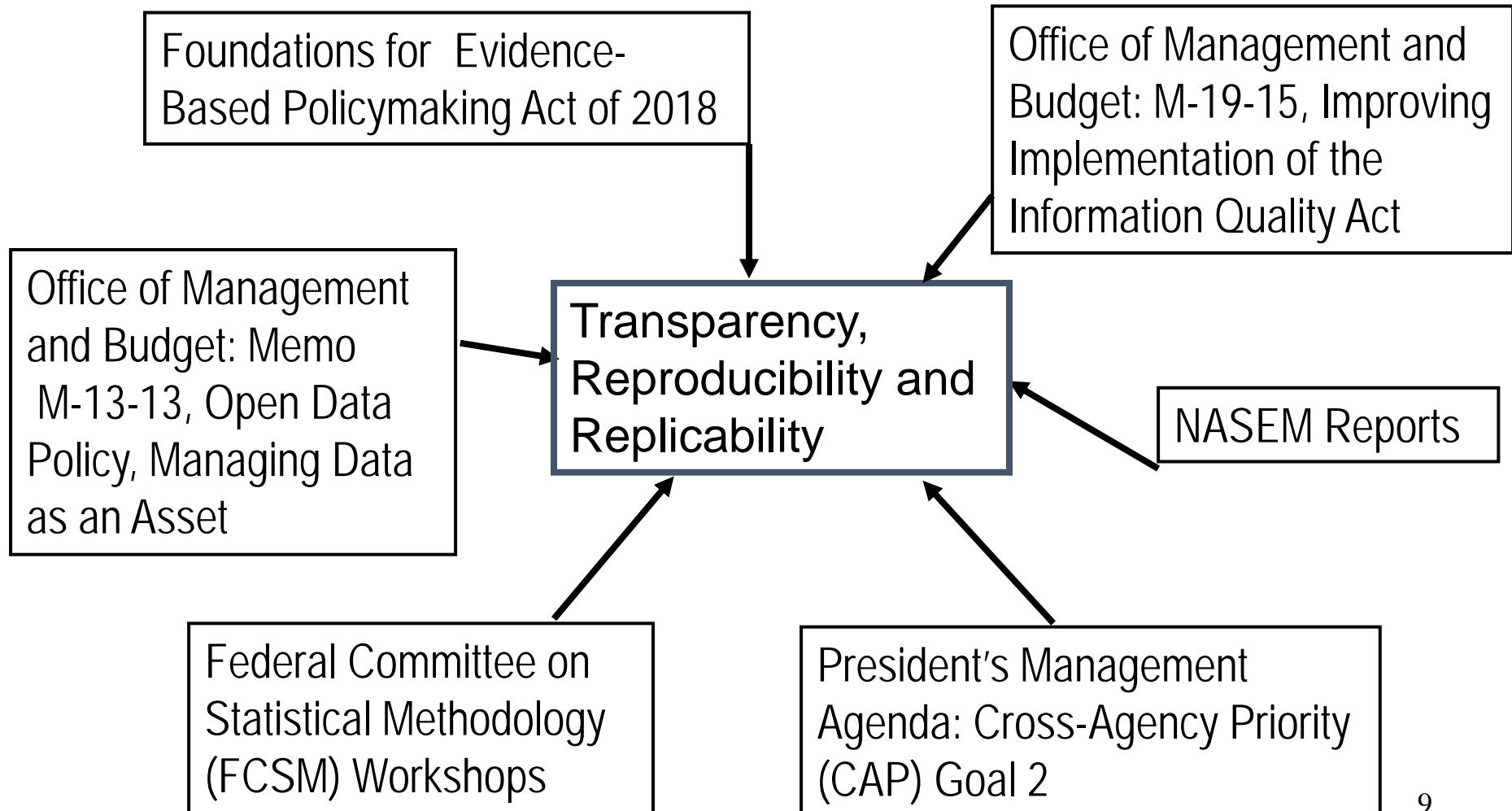




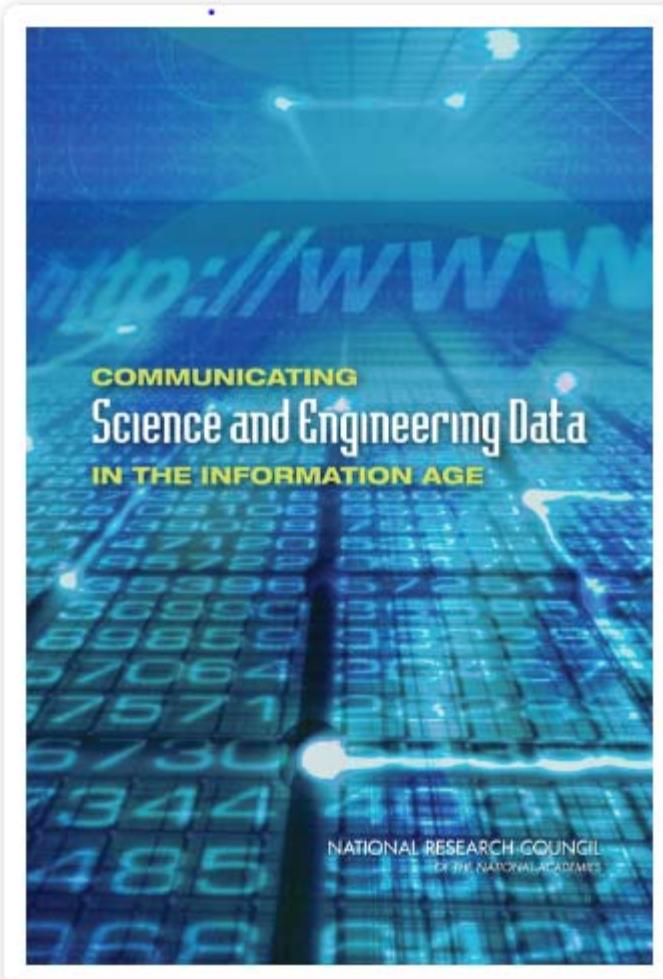
Impetus for this CNSTAT Panel Study

2017 Workshop on Methods to Foster Transparency and
Reproducibility of Federal Statistics (2019 report)

Recent initiatives, workshops and reports informing current panel



NCSES Efforts



- CNSTAT report with several recommendations to improve our dissemination program
- NCSES is developing new Integrated Data System



Why is NCSES Taking on this Complicated Task?



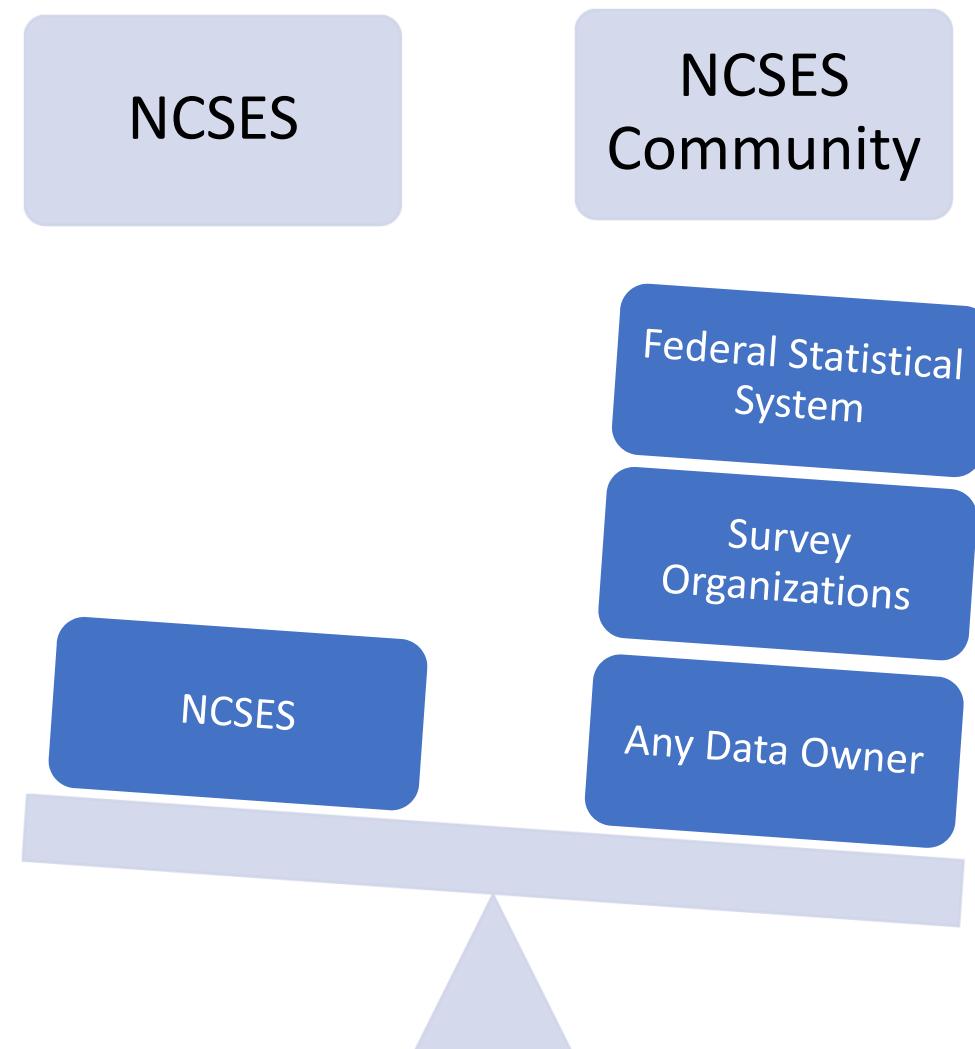
	13 Principal Federal Statistical Agencies	Established	Full-time permanent staff	FY2018 Budget (million \$)
1	Bureau of the Census	1903	6,543	\$1,681
2	Bureau of Labor Statistics	1884	2,035	\$608
3	National Center for Education Statistics	1867	95	\$330
4	National Agricultural Statistics Service	1961	1,022	\$186
5	National Center for Health Statistics	1960	443	\$155
6	Energy Information Administration	1977	368	\$118
7	Bureau of Economic Analysis	1972	480	\$97
8	Economic Research Service	1961	330	\$77
9	National Center for Science and Engineering Statistics	1950	56	\$57
10	Bureau of Justice Statistics	1979	56	\$50
11	Statistics of Income Division	1862	136	\$34
12	Office of Research, Evaluation, and Statistics, SSA	1935	64	\$33
13	Bureau of Transportation Statistics	1992	75	\$26

Why is NCSES Taking on this Complicated Task?

- Staff size and composition
- Survey types
- Data collectors current practices

Beyond NCSES

Charge to the Panel



Charge to the Panel

- Best practices to foster transparency and reproducibility
- Guidance, standards, and tools for documenting and archiving
- Approaches to minimize cost
- Feasible implementation steps – low hanging fruit



Thank You



More information on initiatives, reports and other efforts

Questions to be considered by the panel

- 1) What **documentation and archiving guidance, standards, and tools** currently exist to assist NCSES to facilitate transparency and reproducibility? ...For sample surveys? For administrative records?...
- 2) In what ways can the **costs** of transparency for NCSES and its data users be minimized and the **benefits** maximized?
- 3) How can NCSES obtain value from saving and using the history of statistical program **workflow processes** to facilitate validation ...editing and imputation?
- 4) What are best practices to foster **transparency internal to NCSES** through ...documentation and archiving of methods and data?

Questions to be considered by the panel - continued

- 5) What are best practices to foster **transparency external to NCSES** while [protection] for confidential microdata?
- 6) What requirements for documentation and archiving standards and tools should NCSES include in **contracts with data collectors, such as the U.S. Census Bureau and private survey firms?**
- 7) What are feasible implementation steps toward better documentation and archiving for NCSES in the next 2-3 years? What should be the goals for a longer term R&D effort in this area?
- 8) How can NCSES work **with other federal statistical agencies** to facilitate **adoption of documentation and archiving standards and tools in common?**

FCSM Metadata Workshop: September 14th

- Focuses on metadata and is the 4th in the series of workshops on transparent reporting for blended information products
- Intended to inform OMB and the Interagency Committee on Statistical Policy (ICSP) agency heads on practical metadata requirements to include in a prospective documentation standard
- Broadly covers coverage of current agency practices and their priorities to balance the benefits and burdens of collecting and using metadata
- Concludes with a panel discussion to get a consensus on recommendations for a documentation standard that will be useful yet practical for all federal statistical agencies to implement

OMB memo M-19-15, Improving Implementation of the Information Quality Act (April 2019)

- “This memorandum updates implementation of the Guidelines (2002 *Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility and Integrity of Information Disseminated by Federal Agencies*) to reflect recent innovations in information generation, access, management, and use, and to help agencies address common problems with maintaining information quality”

Major components:

1. **Fitness for Purpose** and Pre-Dissemination Review
2. Enhancing Downstream **uses** of data
3. **Reproducibility** of influential information
4. Requests for correction

OMB memo M-13-13, Open Data Policy—Managing Information as an Asset

- Pursuant to the Executive Order of May 9, 2013, *Making Open and Machine Readable the New Default for Government Information*
- Memorandum establishes a framework to help institutionalize the principles of effective information management at each stage of the information's life cycle to promote interoperability and openness
- Whether or not particular information can be made public, agencies can apply this framework to all information resources to promote efficiency and produce value

President's Management Agenda: Cross-Agency Priority (CAP) Goal #2: Leveraging Data as Strategic Asset

Federal Data Fellows to develop a comprehensive government-wide data strategy

- Develop practices and action steps for four data strategy working groups
- Develop a government-wide perspective on leveraging program, statistical, and mission-support data to deliver on mission, serve customers, and steward resources
- Conduct research, engage Federal and non-Federal stakeholders, analyze inputs, and produce solutions

Evidenced-Based Policymaking

Transparency

- Those engaged in generating and using data and evidence should operate transparently, providing meaningful channels for public input and comment and ensuring that evidence produced is made publicly available
- Increased transparency will enable the public to be informed about how data are being used to improve their government, even as data are being stringently protected

FCSM Working Group on Transparent Reporting

Pilot projects that include different types of combined data

- Blending multiple-survey sources from statistical agencies
- Blended data for national accounts statistics
- Blending administrative and survey data
- Use of private sector data, possibly in combination with other data

Transparency, Reproducibility, and Replicability

Promoting transparency and confidentiality by

- Preserving and accessing shared data and data-related content
- Safe data linkage while protecting confidential data
- Researcher Passport

“ Safe data, safe places, safe people, safe outputs...
Safe data and safe places are complements, not substitutes”

Enhancement of Transparency, Reproducibility, and Replicability (TRR) in the Integration of Multiple Data Sources

- Integration scenarios
 - Survey data as basis; supplement with variables from admin/3rd-party data
 - Organic data sources as basis; supplement with special surveys
- Elements of a conceptual framework and related operational definitions
- Public confidence

*“Media and Public – Transparency emphasis,”
Are we “trustworthy”*