



# Government Decisions and Issues about Collecting and Using Data

## Basics of Evidence Brief #3

*Shelley Metzenbaum*  
THE BETTER PROJECT

*Batia Katz*  
URBAN INSTITUTE

*Demetra Nightingale*  
URBAN INSTITUTE

*December 2022*

**Effective public management and high-quality evidence require good data. Data are turned into evidence through data analyses and well-designed research. Quality data turned into evidence can help government agencies, officials, and others decide where to focus their efforts, find ways to improve, increase adoption of good practices, and build public understanding of and trust in government (Metzenbaum, Nightingale, and Katz 2021). This brief provides a general discussion of governments' decisions about what data to collect and how to collect, store, analyze, share, and use (or require others to use) data.**

The main focus of this brief is government decisions about government data. Data are important for government administrators managing public programs, analyzing results, preparing government statistical reports, and evaluating policy and program effectiveness and costs. In addition, information that emerges from government data or government-supported research and analysis also makes people's lives better in obvious and less obvious ways. For example, people use data for information related to their personal and public health, weather readiness, traffic management, trip navigation, and other issues that involve some government data. How data are collected, analyzed, and shared greatly affects their usefulness to the government and to the public.

## Ways Data Are Used

Box 1 describes some major ways governments and others use data and translate data into evidence.

---

## BOX 1

### Data Uses

Quality data and data procedures aid the following types of uses:

- **discovery** of what is going on in the community, nation, or world that may need attention and action from governments and others, as well as what may be important or interesting to understand better
- **description** of conditions or contexts, such as harmful conditions or incidents; characteristics of program beneficiaries; situations facing affected individuals, businesses, or regulated parties; potential service delivery partners and suppliers; and operating context including risks and opportunities
- **priority-setting** of where to focus in the long and short term and what areas should get higher-priority attention
- **identification** of factors governments may want to affect to improve results or of correlated indicators that can be useful as warning signs
- **predictive analysis** to inform planning, priority-setting and the timing of actions and locations for actions
- **statistical analysis** of causal factors influencing conditions and outcomes, diagnosing problems, and identifying opportunities to inform or modify interventions
- **identifying and prescribing** actions (treatments) to improve results

These multiple uses of data support decisions made about where to focus and what practices, products, and services to use to improve outcomes, operational quality, and government accountability.

---

## Government Data Decisions

Governments cannot, of course, afford to collect information about everything all the time. Decisions must be made about how to collect, store, analyze, share, and use data. Governments make decisions about data strategies by asking and answering several questions:

- Why are data being collected? For what uses and purposes?
- What data items should be collected and how should collected data be described and categorized?
- How should data be collected, including who collects them, when and where they are collected, and what means of data collection should be used?
- Who can access data and data analyses?

## Why Are Data Being Collected?

Government data have many different purposes and uses: to measure government results, improve outcomes and operational quality, make planning decisions, and expand general awareness and understanding of government services and results. Data are used at every phase of the policymaking and policy implementation process, including for program planning, managing daily operations, conducting monitoring or after-action reviews, preparing statistical reports, and conducting research and evaluations. The media and public also use government data for general information and for making various personal decisions.

## What Data Are Collected?

The purposes and uses of data determine what specific data items should be collected and at what level. Useful data can be gathered that include the full universe of topics of interest or focus on a sample from that universe. For program management and performance measurement, activity and outcome data are specified at a program level and often can be disaggregated down to state, local, office, or other levels and aggregated up to a state, regional, or national level. For federal statistical reports, research, and many evaluations, samples that represent the universe are commonly used.

## How Are Data Collected and Analyzed?

Laws, policies, and program guidance often specify who collects government-related data and where and when they collect the data, which in turn relates to how the data can be analyzed and used. Decisions about these matters not only affect how easily data can be used, but also how willing people are to provide and use collected data and how readily they will trust the resultant analyses.

Most government data are now collected and maintained electronically, although a substantial amount of data is still produced manually and on paper. There are pros and cons associated with each method that affect cost and quality. Individuals in households, businesses, organizations, and agencies increasingly collect and transmit data to governments by machines in fully electronic ways. For example, when managers or staff implementing a public program collect and transmit data about their own program electronically, it may take less time and effort than preparing and submitting reports on paper. Once data are collected at any level of government and using any method, they must be reviewed for quality, validated, and often clarified or “cleaned” before analysis can be done. Data and reports prepared on a computer can be reviewed electronically; however, the details may get detached from a paper trail, complicating efforts to check for accuracy. The paper trail may be easier to review with manual reports, but that review may be quite labor intensive.

When data are collected can also make a difference. For example, K–12 test scores collected at the end of a school year do not help current-year teachers or their students, yet sharing scores with next-year teachers runs a high risk of triggering a well-documented phenomenon known as the Pygmalion effect (Boser, Wilhelm, and Hanna 2014). This effect occurs when teachers challenge lower-scoring

students less than they challenge higher-scoring students, depressing the long-term performance of lower-scoring students and exacerbating disparities.

Often a data source, on its own or combined with various data sources, may be analyzed and tailored for different purposes. Managers and performance staff track and analyze both aggregated and disaggregated program activity and outcome data to understand whether programs are moving in the desired direction and whether changes to practice may be warranted. Administrators may also analyze that data to inform funding and staffing decisions. Evaluators may access underlying data in the management information system and analyze both individual-level and aggregated data to estimate effectiveness or formulate theories of change that can be tested to identify ways to improve results. Government statistical agencies may combine the program data with economic, survey, or other data to produce regular reports on the program as a whole in a broader context.

## Who Can Access the Data and When?

Government officials also make decisions about how, when, and with whom data and analyses are shared. In some cases, the actual data are available publicly. In other cases, data files prepared or analyzed are accessible to particular users or interested parties. Specific laws in different jurisdictions and for different programs often allow or limit what and how data can be shared. Data-sharing decisions must also be made about what data not to share to protect personal privacy, national and business security, or confidential business information.

Important questions around data sharing, therefore, include who gets what data when and how data and data analyses get shared. For example, returning relevant analyzed data in a timely manner back to those who submitted data could improve subsequent data quality if people or organizations supplying data feel that analyses of their data, possibly combined with data from others, can help them find ways to improve their own outcomes, operational quality, transparency, or priority setting.

Once data are collected and turned into useful evidence, the evidence needs to be communicated to key users in a thoughtful and timely manner so the information can be used when needed and so evidence users can readily find, access, understand, accurately interpret, appropriately absorb, and apply data-informed insights (Metzenbaum, Nightingale, and Katz 2022).

The way data, data analyses, and other evidence get used requires attention, too, to encourage continual improvement and minimize fear, resistance, and unfair use of data and evidence. On occasion, performance data linked to a threat of penalty or promise of reward encourages those being measured to try to manipulate the measurement system. Years ago, for example, a nontrivial number of teachers in Atlanta gamed student test scores.<sup>1</sup> Using data to prevent this sort of measurement manipulation requires greater investment in data validation and manipulation prevention efforts and well-designed incentive structures. Fortunately, informed by experience and evidence, policymakers can change performance incentive structures to promote improved results. For example, many federal grant programs offer funding or extra funding to encourage program grantees to generate, share, and use

data and evidence. When thoughtfully designed, these federal incentives help build program evidence capacity, improve data quality, and expand adoption of effective strategies.

## Decisions That Make Data More Useful

The following examples highlight the kinds of data-handling decisions governments make that can make data useful:

- **Capture details about data relevant to program objectives and operations.** Federal, state, and local government agencies all maintain data on their programs—such as data about how many people or consumers are served, services received, program results, and expenditures. Capturing details about collected data enables federal, state, local, and private sector actors to make more informed decisions to improve results. For example, every traffic fatality is reported using a standardized data collection framework that captures information about operator, equipment, physical environment, and jurisdictional characteristics before, during, and after each fatality. This information helps federal, state, local, and private sector parties make more informed decisions about which drivers, vehicles, locations, and times of the day and year warrant more attention to try to reduce crashes and fatalities.
- **Collect data frequently and note timing.** Noting when and where services are provided or events happen and when measurements are reported tends to make data more useful, as does more frequent data collection. For example, geo-coding in smartphone navigation systems that frequently collect information about when and where drivers are located has proven valuable for helping people find the best route to their chosen destination. Researchers have similarly learned that time-stamping when actions such as surgeries and judicial decisions happened can reveal time-linked variations in the effects of those surgeries and judicial decisions. More frequent data, both in terms of geography and time, tend to be more useful than less frequent data when measuring items with spatial dimensions and time-linked variations such as water- and air-quality readings and labor market conditions.
- **Share and collect data in a timely manner.** Just as the timing of data collection is important, so is timely data communication. More timely data sharing boosts the usefulness of that data. Police realized, for example, that collecting, sharing, and analyzing crime data weekly instead of quarterly made crime data far more useful both to precinct captains and citywide for deciding where to focus and what to do to reduce crime than did previously reported quarterly crime data, especially because the quarterly data were not reported to a central office until six months after the close of the quarter.
- **Collect adequate details associated with data.** Data are most useful if they are directly relevant to those producing and receiving the resulting reports and analysis, and to other users. Granularity is a term that is generally used to refer to the extent to which a description or observation of a program or system of interest is broken down into small parts. More granular data can reveal important information such as differences in services to specific groups of

people or geographic areas, differences in intermediate outcomes such as vaccination, enrollment, or program completion rates, and differences in end outcomes such as employment, illness, or property damage.

- **Apply technology, but carefully.** Technological advances raise new challenges in some ways, but they also make it easier to collect, analyze, and share data to produce useful evidence and potentially improve privacy protections. Some advances include electronic data transmission, ground and sky-based sensing, computing devices embedded in everyday objects, crowdsourcing, interoperable cloud computing, big data analytic methods, data visualization software, mapping, digitized photos, edge computing, Application Programming Interfaces, easier access to transactional and administrative data from multiple sources, and other developments. For example, the online platforms that evolved when the COVID-19 pandemic increased health risks through close human contact made it possible to engage and inform more people quickly and conduct more transactions and data-informed decisionmaking remotely. More detailed data can heighten concerns about privacy and confidentiality protections, though. This is an increasingly important aspect of data use and data systems in public policy areas, and technological strategies are constantly being improved to reduce data privacy disclosure risks and protect confidentiality.<sup>2</sup>

## Summary

Government policymakers, program administrators, and staff use data and data analyses for multiple purposes, including description and analysis (e.g., of conditions, impacts, costs, supply and demand trends), diagnoses, predictions, and prescriptions. In addition, data analyses help individuals and organizations choose among products, services, providers, and practices and even decide whether something such as air quality, swimming water, or drinking water is sufficiently safe. Program administrators make many decisions about data that enable more useful information for operations, management, performance, evaluation, and other analyses,

Much remains to be learned as knowledge, technology, and the world evolve. Preventing problems while tapping the benefits of advances in technology and analytic methods can be aided by building and sharing evidence and expertise about wise data handling. Programs, agencies, and cross-program efforts to make progress on specific outcome goals are aided by many governmentwide councils, including councils for performance improvement officers, chief evaluation officers, chief data officers, chief technology officers, chief information officers, and the Federal Privacy Council.<sup>3</sup> Working together, agency administrators, program offices and researchers and management experts continue to build federal agency knowledge and capacity about ways to collect, store, analyze, and share data and data analyses to improve outcomes, operational quality, and government transparency and accountability.

Careful decisions need to be made about how much data are collected, how to protect privacy, and how to share collected data and analyses. Otherwise, data collection just becomes burdensome and

annoying, requiring work without generating action-informing insights. Rapidly evolving technologies and analytic methods make wise data handling more feasible than ever before, and much can be learned from past and current practices about more and less effective data-handling strategies for different purposes.

## Notes

- <sup>1</sup> Valerie Strauss, “How and Why Convicted Atlanta Teachers Cheated on Standardized Tests,” *Washington Post*, April 1, 2015, <https://www.washingtonpost.com/news/answer-sheet/wp/2015/04/01/how-and-why-convicted-atlanta-teachers-cheated-on-standardized-tests/>.
- <sup>2</sup> For a discussion of some data privacy statistical methods and issues, see Bowen and Liu (2020).
- <sup>3</sup> See the Federal Chief Data Officer Council website: <https://www.cdo.gov/> (accessed November 8, 2022).

## References

- Boser, Ulrich, Megan Wilhelm, and Robert Hanna. 2014. *The Power of the Pygmalion Effect: Teachers Expectations Strongly Predict College Completion*. Washington, DC: Center for American Progress.
- Bowen, Claire McKay, and Fang Liu. 2020. “Comparative Study on Differentially Private Data Synthesis Methods.” *Statistical Science* 35 (2): 280–307. <https://doi.org/10.48550/arXiv.1602.01063>.
- Metzenbaum, Shelley, Demetra Nightingale, and Batia Katz. 2021. “What Is Evidence? Basics of Evidence Brief #1.” Washington, DC: Urban Institute.
- . 2022. “Communicating Evidence: Basics of Evidence Brief #2.” Washington, DC: Urban Institute.

## About the Authors

**Shelley Metzenbaum** runs The BETTER Project to Bring Everyone Together to Enhance Results. She previously served as the Office of Management and Budget’s associate director for performance and personnel management, founding president of The Volcker Alliance, head of regional operations and state/local relations at the US Environmental Protection Agency, capital budget director in Massachusetts, head of Boston’s Washington office, and economic development specialist for Arkansas. She teaches and writes about using goals and measurement to improve outcomes in effective, cost-effective, and equitable ways and not using them unfairly and in ways that make people afraid, compromising performance gains.

**Batia Katz** is a research analyst in the Center on Labor, Human Services, and Population at the Urban Institute, where she conducts research on workforce development. Katz’s previous research includes studying the science labor market, the impact of personality traits on employment outcomes, and gender and family in the workforce. Katz has a BA in economics from Haverford College.

**Demetra Nightingale** is an Institute fellow and director of the Federal Evaluation Forum at the Urban Institute and was the chief evaluation officer at the US Department of Labor from 2011 to 2016. She conducts research on employment, labor, welfare, and other social and economic policies. She also

previously taught graduate courses in program evaluation and in social policy at Johns Hopkins University and the George Washington University's Trachtenberg School of Public Policy and Public Administration. She is the author or coauthor of five books and numerous articles, a fellow of the National Academy of Public Administration, and serves on many task forces and advisory panels. She received her doctorate in public policy from the George Washington University.

## Acknowledgments

This brief was funded by the Annie E. Casey Foundation and the William T. Grant Foundation, through the Urban Institute's Federal Evaluation Forum. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at [urban.org/fundingprinciples](https://www.urban.org/fundingprinciples).



500 L'Enfant Plaza SW  
Washington, DC 20024

[www.urban.org](https://www.urban.org)

### ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.

Copyright © December 2022. Urban Institute. Permission is granted for reproduction of this file, with attribution to the Urban Institute.