PERSPECTIVE

Addressing Data Quality Problems

BY KATHERINE BARRETT AND RICHARD GREENE

While the power of data to guide government management and policy is immense, that power is greatly diminished when quality issues emerge. "If the data are bad, you're going to get bad results, whether you're doing evaluations or whether you're trying to do program improvement," said Jonathan Ladinsky, an expert on performance management systems and data quality in federal, state, and local government.

As Doug Robinson, executive director of the National Association of State Chief Information Officers, has said, there are "lots of consequences to dirty data." That includes ineffective program delivery and potentially misguided decision making. "All these initiatives about evidence-based decision making are predicated on having data. But you waste taxpayer dollars and other revenue resources if your data quality is poor."

The consequences of having poor quality data have now expanded with increased use of generative artificial intelligence (AI). The dangers of AI searches that grab information from the internet are increasingly well known—but even when the AI systems are using only data from an individual government, the problems don't disappear. As Stephanie Deitrick, chief data officer for the City of Tempe, Arizona, said, "You are no longer just worrying about data that comes out of databases. You need to be thinking about what's on the website and what's documented in files."

Solutions

There is no shortage of advice on the basic ways to improve, though experts counsel to avoid thinking about perfection. Of course, perfection isn't achievable, at least not in 2025. "The only way you get perfect data is if you don't serve anyone," Ladinsky said. "You want to think about the data you really need. You need to think about what is most important for you for managing your program. You need to prioritize."

Still, worthwhile efforts abound. For example, over recent years the importance of data governance and data standardization has been a critical component of the Bloomberg Center for Government Excellence at Johns Hopkins University. That translates into technical assistance on governance policies, the importance of coherent and consistent data leadership, and an ongoing effort to ensure that departments use similar definitions and that data in one division can be shared meaningfully with another.

At the most basic level, data standardization means that an address, for example, will be recorded in the same way no matter who is listing it. It also means that there is a common definition for how employees are counted or how to record a date.

Creating standardized processes early on is, of course, far easier than doing it after multiple departments have plunged into their own ways of doing things. For example, when the City of Tempe began to require standardization of addresses about six years ago, "it was painful," says Deitrick. "We had five different address data sets in the city and none of them matched."

It took the city two years to combine datasets and remove obvious duplicates or other data inconsistencies, but that still left conflicts that needed review. Even now, the city is still working out technical considerations involved in correcting addresses that don't conform to the standardized process—for example, those that include additional details that help police respond to emergencies.

In general, data errors can often be traced to human rather than technical mistakes. But once information is solidly implanted in computerized memory it can live on forever without any notice. One key to avoiding this phenomenon is ample training for all employees whose hands touch the data.

In our 2020 book, *The Promises and Pitfalls of Performance-Informed Management* (Rowman & Littlefield Publishers: 2019), we quoted Beth Blauer, currently associate vice provost for public sector innovation

at Johns Hopkins University, about the importance of spreading data knowledge throughout the organization. Or, as she said, "doubling down" on employees' data skills.

At GovEx, the organization Blauer used to head, the focus on data quality has long included a strong emphasis on training—not just for supervisors, managers, department heads, and city leaders but for individual employees whose understanding of how data will be used, and its importance, is key to having data collected and input accurately and successfully utilized.

Deitrick noted that one way to ensure that data is input correctly is to design the forms that are used to reject potential input errors. As an example, a form can be designed to prevent the start date of an individual or program to occur later than the end date. Or, it can make sure that a dollar figure—for a contractor's payment, for example—cannot exceed a pre-set agreed-upon amount.

Many errors crop up in data thanks to problems during input, so this is yet another area where improvements can be made. "We always need to be thinking about the need for collection to be easy for people who are doing data entry in the field," Deitrick said.

That means looking at the individual challenges of employees in the field and the situations they confront. For example, asking for too many data elements from a fieldworker who is collecting data outside on a blisteringly hot Arizona day may result in too many missing elements. That means that data quality also relies on a constant winnowing down of what you're collecting and siphoning off the material that isn't needed. "You need to challenge the notion that you have to keep collecting the data that you've always collected and think about what the repercussions would be of not collecting it," Deitrick said. "What we're collecting should have a purpose."

Given the promise of big data to change management and shift policy, Deitrick also emphasizes that public-sector employees and the residents they serve should recognize that data, and the analysis of data, is not foolproof. The analysis of data and the correlations that are drawn "have a level of uncertainty that we just have to be comfortable with."

> The consequences of having poor quality data have now expanded with increased use of generative artificial intelligence (AI).

Building better data infrastructure

For many local and state governments, particularly smaller ones, a sizeable resource gap exists between the data demands made by decision makers and the kinds of technology that would make collecting and retrieving data easier. "You're increasingly finding legislatures, individual legislators, and city and county councilmen, who have very high expectations of what government data systems are and should be able to produce," said Joshua Baker, vice president of Mathematica's state and local government division.

But for organizations that are still working in a mainframe or Access database or an Excel-data base environment, there are often difficulties in meeting those expectations. "There are real and sustained resource gaps, particularly for small governments that don't have the budgets to build the infrastructure necessary to collect, manage, and clean data," Baker said.

Baker also noted that he's seen local governments, particularly in the public health area, working to connect with state government to take advantage of federal technology modernization funds that are connected to Medicaid. A Centers for Disease Control and Prevention modernization initiative that was part of American Rescue Plan Act funding has also been helpful in improving local public health data and use. When local governments approach states with their own matching funds and an understanding of the financing mechanisms, it can help draw down needed resources that aren't available locally.

With the federal drive to cut back, however, these past opportunities may be less available in the future. With that in mind, Baker advised connecting and standardizing data collection with other governments to facilitate benchmarking and potentially collaborating on data quality improvements. Working with municipal or county associations can also be helpful, he said. "You really need to look outside of the bubble." This also can mean tapping universities, foundations, and think tanks.

That's what Jefferson County, Alabama, and the City of Birmingham, Alabama, have built through the Birmingham-Jefferson County Justice Governance Partnership, which is designed to address local violence that has contributed to a high Birmingham homicide rate. Sheila Tyson, the co-chair of the partnership, explained in a Birmingham-Watch article that by working with the Aspen Institute's Criminal Justice Reform Initiative, it was able to access data it was lacking on the root causes of youth crime.

As the article reports, Tyson spoke of the county's own data as being collected sporadically and incompletely.

"None of it was done correctly, none of it was stored properly, and none of it was continually being tracked," she said. "This is the first solid data that we can see and keep up, and it works because it's worked in other states. It has a proven record."

Katherine Barrett *and* ***Richard Greene*** *are principals of Barrett and Greene, Inc.*