

Books & arts



Many voters waited hours to cast ballots in US primaries this spring.

Democracy suffers when government statistics fail

The United States must change how it measures society, argues metrics maven Julia Lane. By Beth Simone Noveck

The Web was invented 30 years ago, yet 2020 is the first year that the United States' decennial census has allowed households to respond online. This shift came in the nick of time – given the need for social distancing – and has been a (qualified) success. As of September, more than 88% of housing units have been accounted for. Some 65.5% responded online, by phone or by mail, rather than in response to the conventional knock at the door.

The lack of innovation in federal statistical

agencies such as the Census Bureau contributes to the slow, bureaucratic and hidebound nature of government, argues New York University economist Julia Lane in *Democratizing Our Data: A Manifesto*. Many factors are to blame: a dearth of modern technology, absence of adequate training in new data-science techniques such as machine learning, outdated legal rules that preserve inflexible practices, and budgetary processes that create silos and impede collaboration.

If you think statistics and government

sounds like twice the tedium for half the price, you would be entirely wrong. *Democratizing Our Data* is an illuminating and powerfully argued case that the United States must change the system it uses to produce public statistics. Written pre-pandemic, its lessons are now the stuff of daily headlines – as crumbling systems stymie crisis management from epidemiology to unemployment to the logistics of the upcoming election.

As the nation's former deputy chief technology officer (under President Barack Obama), I found the book thoroughly persuasive with regard to statistical agencies – which are an exemplar of how government fails to innovate more generally. Broad-scale collaboration with universities and those outside the federal government will yield new ideas more quickly.

Lane has done path-breaking work to invent ways to measure the economic impact of public investments in science and technology. In her trenchant style, replete with personal stories, Lane asserts that the United States is failing to adequately track its population, economy and society. Agencies are stagnating. The census



Local-government workers in Bogotá in April get ready to check which families need food.

dramatically undercounts people from minority racial groups. There is no complete national list of households. The data are made available two years after the count, making them out of date as the basis for effective policymaking.

Until recently, the US federal government spent billions on science and technology research with no idea about the return on investment. Moreover, we lack skilled people in government to make use of this rich data store to inform how policy is made. Given that the average salary for a senior data scientist in Silicon Valley is twice that of a senior civil servant, this is no wonder.

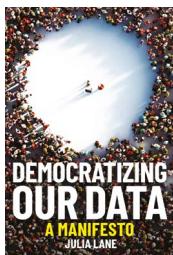
In the United States, there is no single national statistical agency. The process of gathering and publishing public data is fragmented across multiple departments and agencies, making it difficult to introduce new ideas across the whole enterprise. Each agency is funded by, and accountable to, a different congressional committee. Congress once sued the commerce department for attempting to introduce modern techniques of statistical sampling to shore up a flawed census process that involves counting every person by hand.

Both the US gross domestic product (GDP), arguably the most important measure of national economic well-being, and the national unemployment statistics are hopelessly flawed. Yet despite decades of criticism that we are failing to measure what we value, we do not change these measures. Paraphrasing Robert F. Kennedy, Lane writes about GDP: "We judge the United States by production – we count air pollution, cigarette advertising, locks for our doors and jails for people who break them. We count the destruction of the redwood, the production of nuclear warheads, but not the health of our children, the beauty

of poetry, the intelligence of our public debate or the integrity of our officials ... [GDP] measures everything, in short, except that which makes life worthwhile." While we may not want to measure the beauty of poetry, her point is that we are stuck without the ability to experiment with new forms of measurement.

The US statistical agencies still rely primarily on mailed surveys. By contrast, universities are able to experiment with sentiment analysis, looking at tweets or Google searches to understand trends. Although this technique has not always succeeded (searches for the word 'flu' did not effectively anticipate the number of visits to the doctor), it is only through trial, error and experimentation that better methods emerge. Lane provides a fascinating, albeit disheartening, rendition of what it takes to write, test, approve, train the staff for, administer and analyse a major national public survey. A former US chief statistician estimates that the process can take ten years.

By contrast, the best private-sector companies produce data that are in real time, comprehensive, relevant, accessible and meaningful. To produce comparable public data, Lane suggests we should learn from examples such as the Longitudinal Employer-Household Dynamics programme. This started as a university-based research project to measure



Democratizing Our Data: A Manifesto
Julia Lane
MIT Press (2020)

economic returns from on-the-job training. As it evolved, researchers – working in collaboration with the Census Bureau and the states – developed new measures of workforce dynamics as well as nifty visualizations. Over decades, this partnership between public officials and university researchers became a valuable national indicator of worker flows, job flows and worker churn using already-collected data in widespread use among workforce and transport planners.

Lane argues that Congress should create a "National Lab for Community Data" that operates outside the government. Like other federally funded national research labs, such as Lawrence Livermore in California, which accelerate important research in the public interest, it would enjoy access to well-trained talent from outside, and data from inside, government. And because it would be quasi-external, it would be able to innovate faster and be more responsive to citizens' needs, leading to the creation of more relevant data.

Lane builds the case for new legislation to create this independent national data authority. Presumably, because the desire is to make a non-partisan argument for innovation, she conspicuously avoids discussion of politics. We do not hear about how many talented public servants have left the federal government because dissent is not tolerated under President Donald Trump or about how, in 2019, the US Department of Agriculture decided to relocate the agency's widely respected expert research services from Washington DC to Kansas City, Missouri. Fewer than two-thirds of the staff agreed to relocate, thereby gutting the agency's data and research capacity.

If you are looking for another tell-all about the Trump administration, this is not it. Given how politicized the discussion around the census has become – with high-profile court battles over counting undocumented immigrants, for instance – and the broader national debate around the politicization of science agencies during the COVID-19 epidemic, more discussion about politics in *Democratizing Our Data* might bolster Lane's argument for external collaboration. Also welcome would have been a longer discussion about statistical agencies in other countries – what is working and what is not around the world – as well as further exploration of how to engage the public in these important debates.

For now, this pithy volume is a must-read account of what the US federal statistical agencies are, what they do and why public statistics are vital to democracy. If we cannot be counted, we cannot be heard.

Beth Simone Noveck is a professor at New York University's Tandon School of Engineering, where she directs the Governance Lab.
e-mail: noveck@thegovlab.org