

National Association for Business Economics: Economic Measurement Seminar

Maintaining the Quality and Integrity of U.S. Government Data

Director's Remarks as Prepared for Delivery

July 19, 2022

- Good afternoon, everyone. Thank you to the organizers for inviting me to participate in this panel. It's a topic that is near and dear to me.
- You know, U.S. government data is both important and ubiquitous. Financial markets eagerly await reports—sometimes with trepidation—on GDP, inflation, fuel prices and the like. These lead to decisions that affect people's lives and livelihoods.
- This session is about maintaining the quality and integrity of government data. So if the question is how we maintain quality and integrity, we need to first understand the threats.
- I think that federal statistical agencies face three major challenges to maintaining quality and integrity of government data:
 - First is public trust in government and specifically in the effort to collect and secure data.
 - Second is an adherence to generations-old conceptions of how government collects and publishes data.
 - Third is the effort to diminish the independence of federal statistical agencies.
- To me, trust is arguably the biggest challenge. Hey, many folks trust the government. But let's be real—there are parts of society that don't. And it's impeding our ability to secure data from those folks.
- We've researched this. In 2018 the Census Bureau conducted a survey called the 2020 Census Barriers, Attitudes, and Motivators Study. We found that trust in government was lowest among people who are historically hard to count. They tend not to participate in our censuses and surveys.
- But even before this study, response rates to our national surveys were trending downwards. It's well-documented in the survey research literature. And such downward trends have two consequences.
- First, the cost of securing a response increases due to the extra effort that's needed to get participation.
- The second consequence is an increasing risk of nonresponse bias. As the nonresponse rate increases, so does reliance on the adjustments that are needed to align observed responses to the entire population. And as the pool of nonresponding subjects grows, so does the likelihood that those you don't observe are systematically different than those you do observe in ways you can't adjust.
- And there's more. Among participants, there's the burden of providing accurate responses to questions that are increasingly difficult to answer.
- Take something as simple as monthly household income. For those of us with full-time salaried jobs, it's easy to provide accurately. But many people in this country work multiple part time jobs, engage in gig work, and work episodically, just to make ends meet. That makes reporting one's income very difficult.
- It seems like the people we need to know about most—the vulnerable population—may well be those we measure the least accurately. The bottom line is that multiple threats to data quality exist regardless of participation status.

- So how do we maintain the quality and integrity of our data? At the Census Bureau, our approach is to reimagine the underlying data collection paradigm. We've begun a transformation and modernization initiative to create an enterprise-level process of data ingestion, processing and dissemination. This will require a multi-year effort.
- We're focusing on new ways to find and collect data and envisioning a Bureau that adopts a data-centric lens. We are currently an organization that solicits data from siloed directorates to create siloed products that don't often connect with each other. A data-centric Census bureau flips this framework on its head by focusing on pooling and linking data from all sources to help address important policy questions.
- Besides our conventional censuses and surveys, we're expanding administrative data from government sources as well as from the commercial sector. And we're evaluating and improving current processes, infrastructures, and even our internal work culture.
- Our objective is to ensure equitable collection, analysis, and dissemination of data that reflects the changing nature of our nation's population and economy. Our new system will use common platforms for ingesting data, be they censuses and surveys, or administrative data sources.
- We've already begun the process of combining and linking some of our largest administrative data bases. Two other aspects of this transformation effort are important.
- The first is a new, intensified effort to engage stakeholders, partners and the public on a continuous, evergreen basis, rather than once in a decade. This will clarify data user needs. It also will provide an opportunity to relay the value proposition from participating in our data collections. Our outreach efforts have already begun in earnest and are expected to bolster trust in the Census Bureau.
- Secondly, we're fully integrating the principles of diversity, equity and inclusion—or DEI—into our workforce. Creating an inclusive and diverse workforce by living our DEI principles will add new perspectives. This will help spur opportunity, innovation, and excellence. Ultimately, DEI will lead to more effective research, better methods, and more accurate and relevant data.
- OK, so I've talked about how we're addressing trust and modernization. Next, I'd like to discuss how we maintain integrity of our data.
- Integrity flows naturally from the core values. These include scientific integrity, objectivity, transparency and DEI.
- Scientific integrity speaks to the objectivity of the data and adherence to rigorous standards. It also includes organizational independence that is free of political interference. Maintaining scientific integrity requires due diligence in exploring and communicating the strengths and limitations of the data.
- This speaks to the value of transparency, and the 2020 Census provides a great illustration. We've already released a number of reports on 2020 Census operational quality metrics. We commissioned three independent studies with outside experts. We released results from a Demographic Analysis and our Post Enumeration Survey. And there's still more to come.
- Like its predecessors, while the 2020 census wasn't perfect, it was fit for congressional apportionment and legislative redistricting. And because perfection is unattainable, knowing the data's strengths and limitations actually increases its utility.
- Let's now talk about agency independence.
- I believe that scientific integrity, objectivity and transparency can only be achieved by a statistical agency when it operates independently. That means statistical products must be developed, controlled and disseminated by career staff, and career staff alone.
- Let me conclude the way I started, by expressing my appreciation to NABE for the opportunity to share my perspectives on how the Census Bureau and government more generally maintains the quality and integrity of our data.
- Thank you.