Data Drives the First Digital Census

From address collection to mobile self-response, the U.S. Census Bureau is leveraging big data.

James Mersol

Fri, 10/25/2019 - 10:09

While the 2020 election already dominates news headlines, another major event coming in 2020 has largely flown under the radar — the largest data collection ever in U.S. history. It’s the 2020 census.

Photo Credit: GarryKillian/iStock
“This census is our census,” said Atri Kalluri, the senior advocate for response security and data integrity for the U.S. Census Bureau at Splunk’s .conf19 in Las Vegas this week. He added that the census bureau thinks about why the Constitution mandates the census every decade and also how it affects appropriations of funds for communities and how each community is represented in Congress.

Of course, the data and IT landscape is vastly different than it was a decade ago. The Census Bureau has already collected data from the U.S. Postal Service and over 40,000 local governments, Kalluri said, and it is using that data to help establish what addresses have changed over the past decade. This year, the bureau also has GPS data for every address, and while it keeps that data secure on an individual basis, it will add geospatial data to the aggregate results.

This will be the first census where households can respond to the census online or by making a phone call, rather than filling out paper forms and filling it in.

“This is the first time we’re using the internet to collect data for a Decennial Census,” said Zack Schwartz, deputy division chief of the bureau’s communications division.

Respondents can also fill out the census from their mobile devices without having their census ID number, making it easier than ever to self-respond, Kalluri said. Enumerators who visit nonresponding households will also collect responses on mobile devices rather than the traditional clipboard of forms.

“If you’re standing in line for your badge at this conference [next year] and you feel motivated, you can respond on your device,” he said.

“Security is in our DNA,” Kalluri added in an interview. His office plans to use automation to collect responses in centralized nodes to minimize the time individual responses stay on mobile devices, as data privacy is critical to national trust in the Census Bureau.
Data on native languages spoken across American households informed the bureau’s choices for language options. Households responding by phone, internet or mobile can enter their information in 13 different languages, and when the bureau releases the results, it will do so in 59 different languages to reach the broadest majority of Americans.

Another factor that has weighed heavily on the minds of Schwartz and his team is protecting the public image of the Census Bureau and maintaining public trust. Galvanized by the proliferation of misinformation and disinformation during the 2016 election, the team began working with social media companies years ago to counter falsehoods about the census and its purposes, he said.

Maintaining the positive reputation of the Census Bureau is also critical to minimizing costs and maximizing value. Self-response is the most cost-effective way for the bureau to collect data, especially when compared to the money and time spent on sending enumerators out to knock on doors.

Even with the motivational campaign, the Census Bureau currently predicts that most households will not self-respond and is leveraging datasets to reduce the cost of knocking on doors. Schwartz explained that the data from local governments also allowed the bureau to eliminate vacant or nonexistent addresses from its registries. The data also allow the bureau to plan optimized routes for enumerators to travel, cutting back on travel time and mileage fees as they collect data.

Finally, the agency stands ready to efficiently tabulate the data and release the census results. Making the aggregate data accessible for the American public while protecting individual households’ privacy is a priority, Schwartz said, and the bureau has tested and deployed an interface built upon user experience best practices.

“We want you to use our data for the work that you do,” Schwartz said.

View printer friendly version
Census
Data Analytics
privacy
Standard