

## Potential privacy lapse found in Americans' 2010 census data

February 17 2019, by Seth Borenstein



This March 23, 2018, file photo shows an envelope containing a 2018 census letter mailed to a U.S. resident as part of the nation's only test run of the 2020 Census. The Supreme Court will decide whether the 2020 census can include a question about citizenship that could affect the allocation of seats in the House of Representatives and the distribution of billions of dollars in federal money.(AP Photo/Michelle R. Smith, File)

An internal team at the Census Bureau found that basic personal information collected from more than 100 million Americans during the



2010 head count could be reconstructed from obscured data, but with lots of mistakes, a top agency official disclosed Saturday.

The age, gender, location, race and ethnicity for 138 million people were potentially vulnerable. So far, however, only internal hacking teams have discovered such details at possible risk, and no outside groups are known to have grabbed data intended to remain private for 72 years, chief scientist John Abowd told a scientific conference.

The Census Bureau is now scrapping its old data shielding technique for a state-of-the-art method that Abowd claimed is far better than Google's or Apple's.

Some former agency chiefs fear the potential privacy problem will add to the worries that people will avoid answering or lie on the onceevery-10-year survey because of the Trump administration's attempt to add a much-debated citizenship question.

The Supreme Court on Friday announced that it would rule on that proposed question, which has been criticized for being political and not properly tested in the field. The <u>census</u> count is hugely important, helping with the allocation of seats in the House of Representatives and distribution of billions of dollars in federal money.

The 8 billion pieces of statistics in <u>census data</u> are supposed to jumbled in a way so what is released publicly for research cannot identify individuals for more than seven decades. In 2010, the Census Bureau did this by swapping similar household information from one city to another, according to Duke University statistics professor Jerome Reiter.

In the internal tests, Abowd said, officials were able to match of 45 percent of the people who answered the 2010 census with information from public and commercial data sets such as Facebook. But errors in



this technique meant that only data for 52 million people would be completely correct—little more than 1-in-6 of the U.S. population.

He said the 2010 census used the best possible privacy protection available, but hackers since then have become more skilled in reconstructing data. To counter their growing abilities, the agency has completely changed the system for 2020 and will offer the "gold standard" of privacy regardless of the fate of the citizenship question, Abowd said.

People "want to know that statistical tables aren't going to come back and haunt them," Abowd said at the American Association for the Advancement of Science's annual meeting. "I promise the American people they will have the privacy that they deserve."

Georgetown University provost Robert Groves, who headed the 2010 census, said the count had the proper privacy and that every census improves. He lauded the new steps.

Former agency chief Kenneth Prewitt, a professor of policy at Columbia University, said the basic information such as age and ethnicity, even if publicly revealed, isn't as big a deal as other data breaches.

"There is a widespread privacy anxiety out there that is very much related to Facebook and Google and so forth," Prewitt said. "I'm much more worried about the fact that my iPhone follows me around every day."

In a statement, Apple's Fred Sainz took issue with such privacy concerns: "The iPhone doesn't follow you around all day long—Apple has no idea where you are nor do we care. And Apple does not sell information to companies." He noted, however, that consumers can choose apps that know their location.



Abowd said "the 2020 census will be the safest and best protected ever. And this is not as easy as it sounds."

The new system involves complex mathematical algorithms that inject "noise" into the data, making it harder to get accurate information and providing "a very strong guarantee" of privacy, said Duke University computer sciences professor Ashwin Machanavajjhala.

This increases privacy while lowering the accuracy for researchers who use the statistics. Think of it as one set of knobs being dialed up while a second is dialed down at the same time.

The decision on the official privacy/accuracy setting for 2020 hasn't been set. Abowd said policy officials, not engineers or scientists, will make that call.

The Census Bureau tried this system in a 2018 survey using an ultrastrict privacy setting that, while not directly comparable to Google or Apple, is hundreds if not thousands of times more secure for privacy than what's now being used on data from searches using Google Chrome or Apple's iPhone, Duke's Reiter said.

Prewitt suggested the public might not understand the extra efforts underway for the 2020 count but would be spooked by the disclosure about the privacy vulnerability, making people more reluctant to comply with the next census.

If the administration succeeds in adding the citizenship question, "there will be a huge evasion of it (the census) and some selective misuse of it," Prewitt said.

Whether some avoid the survey because of it or lie, neither is a good outcome, making the data less usable, Prewitt said.



Groves said technical experts have serious problems with the citizenship question because it hasn't been tested in the field, as all census questions usually are. He compared it to putting a new drug on the market before the necessary testing.

"Very subtle wording and positional changes in a thing like the Census can have enormous impact way beyond what we as humans can predict," Groves said

© 2019 The Associated Press. All rights reserved.

Citation: Potential privacy lapse found in Americans' 2010 census data (2019, February 17) retrieved 1 May 2024 from <u>https://phys.org/news/2019-02-potential-privacy-lapse-americans-census.html</u>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.