

## The Differential Privacy Corner: What has the US Backed Itself Into?

O'Hara, A<sup>1\*</sup> and Brummet, Q<sup>2</sup>

<sup>1</sup>Georgetown University

<sup>2</sup>NORC

An expanding body of data privacy research reveals that computational advances and ever-growing amounts of publicly retrievable data increase re-identification risks. Because of this, data publishers are realizing that traditional statistical disclosure limitation methods may not protect privacy.

This paper discusses the use of differential privacy at the US Census Bureau to protect the published results of the 2020 census. We first discuss the legal framework under which the Census Bureau intends to use differential privacy. The Census Act in the US states that the agency must keep information confidential, avoiding "any publication whereby the data furnished by any particular establishment or individual under this title can be identified." The fact that Census may release fewer statistics in 2020 than in 2010 is leading scholars to parse the meaning of identification and reevaluate the agency's responsibility to balance data utility with privacy protection.

We then describe technical aspects of the application of differential privacy in the U.S. Census. This data collection is enormously complex and serves a wide variety of users and uses – 7.8 billion statistics were released using the 2010 US Census. This complexity strains the application of differential privacy to ensure appropriate geographic relationships, respect legal requirements for certain statistics to be free of noise infusion, and provide information for detailed demographic groups.

We end by discussing the prospects of applying formal mathematical privacy to other information products at the Census Bureau. At present, techniques exist for applying differential privacy to descriptive statistics, histograms, and counts, but are less developed for more complex data releases including panel data, linked data, and vast person-level datasets. We expect the continued development of formally private methods to occur alongside discussions of what privacy means and the policy issues involved in trading off protection for accuracy.

\*Corresponding Author:

Email Address: [amy.ohara@georgetown.edu](mailto:amy.ohara@georgetown.edu) (A O'Hara)

