Researchers’ Understanding of the Differential Privacy (DP) Approach and Their Attitudes Toward Its Implementation

Qi Zhong1, Robert W. Proctor1, Aiping Xiong2, Jeremiah Blocki3, and Ninghui Li3

1Department of Psychological Sciences, Purdue University
2College of Information Sciences and Technology, The Pennsylvania State University
3Department of Computer Science, Purdue University

Researchers in many fields rely on the U.S. Decennial Census and the American Family Survey for studies in various domains. Over the past decade, the protection of personal privacy has become an increasingly relevant issue as it has become clear that there are ways to identify individuals from de-identified data. Consequently, a decision was made to apply differential privacy, a method for preserving privacy, to the data from the 2020 Census. Differential privacy methods introduce noise into the data set so that group statistics are still useful, but the individual data are randomly perturbed. Concern has been expressed as to how much this will impact research based on the Census data. We conducted a survey for which the respondents were researchers who have published studies using the 2010 U.S. Census data or the American Family Survey data to gauge their understanding of and concerns about differential privacy. The results show that most of them were aware of the differential privacy plan but also had concerns that the impact of differential privacy on the data would lead to loss of information and accuracy and reported being hesitant to use the data sets with differential privacy.

Keywords: 2020 Census, American Family Survey, differential privacy, risk, survey research