Submission for the 2020 IAOS Prize for Young Statisticians

Big Data, Differential Privacy and National Statistical Organisations

Mr. James Bailie
Researcher
Australian Bureau of Statistics
www.abs.gov.au
Canberra, Australia

Abstract

Differential privacy (DP) has emerged in the computer science literature as a measure of the impact on an individual's privacy resulting from the publication of a statistical output such as a frequency table. In this paper, we provide an introduction to DP for official statisticians and discuss its relevance, benefits and challenges from a National Statistical Organisation (NSO) perspective. We motivate this study by examining how privacy is evolving in the era of big data and how this might prompt a shift from traditional statistical disclosure techniques used in official statistics – which are generally applied on a cell-by-cell or table-by-table basis – to formal privacy methods, like DP, which are applied from a perspective encompassing the totality of the outputs generated from a given dataset. We identify an important interplay between DP's holistic privacy risk measure and the difficulty for NSOs in implementing DP, showing that DP's major advantage is also DP's major challenge. This paper provides new work addressing two key DP research areas for NSOs: DP's application to survey data and its incorporation within the Five Safes framework.