Statistics Canada's Experience with Crowdsourcing during COVID-19: Expanding the Collection Tool Kit.

Éric Rancourt[1]; *François Brisebois[1]

[1] Statistics Canada

Abstract:

The COVID-19 pandemic has suddenly turned the data collection strategies of statistical agencies upside down. With the elimination, or at least the restriction, in conducting field collection activities, obtaining responses to traditional surveys posed a greater challenge. At the same time, faced with the urgency and the unknown created by the pandemic, the demand for statistical information quantifying the extent of the effect of the pandemic on the population and the economy then exploded.

Among the solutions provided to remedy this emergency situation, Statistics Canada turned to the collection of citizen data generated, more specifically to the use of collection by crowdsourcing. Crowdsourcing had been tested in the past at Statistics Canada in a very different and limited context; here, with the emergency situation caused by the pandemic, the crowdsourcing approach offered a simple, fast and inexpensive way to hear and communicate the pulse of citizens. From a statistical data quality perspective, this new tool brilliantly ticked the boxes of relevance and more importantly great timeliness in this period of social and health crisis. However, due to its non-probabilistic nature, such a data collection method also presented its share of challenges, with in the foreground the strong potential for participation bias and therefore a limitation in its ability to produce statistical information representative of the specific population of interest.

The presentation will describe the path taken by Statistics Canada in this experience with crowdsourcing, through the quick adoption of the approach at the very beginning of a pandemic, the measures put in place to mitigate the limits and risks associated with using such data, to an overview of the agency's research agenda aiming at a better integration of these citizen generated data in its statistical programs.

Keywords:

Citizen generated data; Non-probabilistic data; Data quality; Data integration