## Abstract

Faced with declining budgets, declining response rates, rising data collection costs and increasing demand for rich, detailed and frequent statistics for informed decision making, National Statistical Offices (NSO) are increasingly exploring more creative ways to produce statistics than using traditional censuses and surveys.

Data integration is the act of combining data from multiple sources, including NSO's own data sources and new data sources. As they have arguably more public value than the sum of its disparate component sources, integrated data sets can provide rich data analyses and small domain estimates which are hard to generate from probability surveys without making significant modelling assumptions. As a result, integrated data sets are being relied upon as a new tool to complement the traditional methods to produce official statistics.

In this talk, we describe the type of data integration that utilises external information to improve the efficiency of NSO survey estimates.

After outlining the different data resources that may be available to the NSO for data integration, e.g. satellite imagery data, on-line panels data or administrative data together with a probability sample from the NSO, we describe the estimators that are commonly advocated in the literature to harness the information from such data resources for descriptive inference.

Finally, obtaining access to non-NSO data from data custodians, with few exceptions, cannot be taken for granted, and requires sustained investments from the NSO to build rapport and trust with them. We also briefly describe some experiences in gaining such access to administrative and scanner data in Australia.

Key words: Big Data, Calibration Estimator, Data Integration, Doubly Robust Estimator, Regression Data Integrator