FUTURE CHALLENGES AND OPPORTUNITIES FOR STATISTICS CANADA

DOUG NORRIS

Doug Norris, PhD, is a Senior Vice President and Chief Demographer at Environics Analytics and one of Canada's leading experts on the census.

Statistics Canada has come a long way from its origins as a paper-based organization that did tabulations manually. Today, it is collecting and disseminating increasingly complex data and analysis electronically.

Over the past 100 years, the statistical agency has overcome many challenges, but with every test, it has evolved and grown stronger. That experience will be crucial in the coming years as it is about to be tested once again. The rapid advances of technology and the associated rise of "big data" have ushered in a new era in which evidence-based decision-making is the new mantra.

For many organizations, data and analytics are increasingly expected to shape policy and business decisions. Adapting to a wide range of increasing demands won't be easy, but it could mean an even stronger agency.

NEW CHALLENGES

One of Statistics Canada's primary roles is to help all levels of government craft better policies by supplying them with vital information about Canadians. While that mandate has not changed, the questions our leaders are being asked to answer have changed — and they're becoming increasingly complex as Canadian society undergoes a significant shift. For instance, Canada's population is aging, while rural and small towns across the country are shrinking. Overall, the country is becoming more diverse and our leaders are trying to cope with the challenges related to inclusion.

To compound matters, the economy is undergoing a change due to the rapid growth of the digital economy and the adoption of automation, which will compel the future labour force to adopt new skills. This economic change could also exacerbate

inequality in the population.

Governments need higher quality data, delivered in a timely manner and with more granularity, to address these multifaceted issues. It's not just governments who need these data; not-for-profit organizations and private sector businesses are also looking to data to help them assess the changing needs for their products and services, as well as to improve the way they serve their customers. This increased awareness of the power of data has resulted in a renewed focus on "data gaps," which was recently highlighted in an extensive series in the *Globe and Mail*.[1]

The pressure to deliver data in a timely manner will be a particular challenge for Statistics Canada. Historically, the data development cycle was a long, often multi-year process. At times, data only became available after some unexpected event thrust an issue onto the public agenda. Technology has improved timeliness, but the organization still has a long way to go before it can deliver data in near real-time.

Fortunately, the agency is heading in the right direction. There is a recognition that non-traditional data sources can also play a role. A recent example occurred leading up to the legalization of cannabis. To help governments and businesses prepare for the policy change, there was a surge in demand for cannabis data. In response, Statistics Canada conducted a new survey and used "crowdsourcing" techniques to estimate the price paid for cannabis.

A second example involved the demand for data on the foreign ownership of condominiums, which were thought to be a significant factor contributing to the inflation of housing costs. Statistics Canada has been working with the Canada Mortgage and Housing Corporation (CMHC) and municipalities to develop new data on this issue quickly.

DATA COLLECTION

The increasing demand for data is happening at a time when traditional data collection approaches are being called into question. Much of the Statistics Canada social statistics program has been built around household surveys. However, response rates to surveys have been falling and it seems clear that it's time to consider a new approach.

The solution is to look to administrative or other data sources collected for reasons other than statistical purposes. For instance, large federal government administrative data sets, such as tax records and immigration records, have been used for some time as part of some statistical programs, but these represent only the tip of the iceberg.

Governments, the private sector and other organizations collect an enormous amount of data, not the least of which are social media data. These data present statistical challenges, including the coverage of the data set and the concepts and definitions used. Expanding data sources can also introduce new issues around comparability between data sets. This is a particular challenge for data related to health, education, justice and labour. Many of

1 *Globe and Mail*, "In the dark: The cost of Canada's data deficit," Jan. 26, 2019

those data fall under provincial jurisdiction where programs and privacy requirements differ from province to province. Municipalities also collect data that would be valuable, although the comparability issues are much greater.

Despite these challenges, the use of administrative data has several advantages. In many cases, the data will be more accurate. For example, the shift to using tax data for all respondents in the 2016 Census significantly improved the accuracy of the income data, especially at a small-area level. The use of postal codes or address coding also allows for the production of small-area level data — something that has been in very short supply outside of the census — which is important for businesses and local governments. Moreover, the use of administrative data makes it possible to combine data from different sources or over time, which allows for the development of a much richer data series.

While administrative data can provide great opportunities, all data development initiatives must be carefully balanced against issues of privacy, which are becoming more critical with each passing day. We are fortunate that Statistics Canada has a stellar reputation for safeguarding the privacy and confidentiality of personal data while maintaining or even expanding access to data to inform public policy. Privacy and security expectations will continue to rise, especially as our lives are stored electronically. StatCan needs to be more proactive in informing the public about the importance of the data it collects and its applications, but also about all of its safeguarding mechanisms. They need to earn the public's trust.

ACCESS AND DISSEMINATION

In addition to the many challenges and opportunities on the data development side, Statistics Canada will also have to continue to improve access to and dissemination of data. Analysts have already seen some major improvements in this area in the post-cost recovery era. Today, all non-confidential data produced by Statistics Canada are available free of charge online.

The introduction and growth of Research Data Centres, which give researchers the ability to access detailed data in a controlled environment, has been a significant advancement in making data more accessible. A future step, on which some work is already being done, is to provide access to detailed microdata in real time to researchers with, of course, the appropriate safeguards on confidentiality and security of the datasets.

While technology has greatly facilitated data dissemination, it also presents larger challenges to maintaining privacy and confidentiality. For instance, while using technology to link datasets can make data more useful, it makes disclosure more complex. It is interesting that the U.S. Census Bureau recently decided to strengthen its data disclosure approach by moving to a technique called differential privacy,[2] which requires all data tabulations to include some random noise. Furthermore, at the time of writing, the Census Bureau has stated that because of this change it may not be able to

2 See https://assets.ipums.org/_files/mpc/MPC-Working-Paper-2018-6.pdf

produce some of its detailed tabulations and public use data files from the 2020 Census.

STAKEHOLDERS

A valuable by-product of the widespread focus on data is that there is a growing number of active stakeholders creating data. With much more data available, the opportunities for data development are greatly increased, but they will require new methodological techniques and approaches.

There needs to be much more cooperation among the many data players. Historically, much of Statistics Canada's work has been directed at meeting the needs of the federal government at the national and, to a lesser extent, provincial level, with less attention given to smaller area geographies. The involvement of more stakeholders means that mechanisms need to be established to better understand data needs, including sharing some of the data development work as well as the resulting output of those data.

Another challenge will be deciding how to prioritize data needs. In some cases, Statistics Canada may have to stop producing some of its lower priority data, thereby creating data gaps for some, to free up resources for higher priority needs.

Many opportunities and challenges lie ahead for Statistics Canada and more generally for all data players. Hopefully, the newly announced Canadian Statistics Advisory Council can provide Statistics Canada with guidance. More generally, the government and Statistics Canada have embarked on a data modernization program that not only looks for ways to replace existing data, but will also search for ways to develop new and more detailed data to address issues of the day.

We, at Environics Analytics, are excited about this government's commitment to big data and innovation. We believe that more collaboration between the data geeks in the public and private sectors will accelerate the process. We, together with our clients, some of the best and brightest data users in Canadian business, are committed to helping put data in action and make Canada a data-driven economy and society.