STATSCAN LOOKS TO THE FUTURE: 'PRESSURES AND OPPORTUNITIES THAT ARE IMPOSSIBLE TO IGNORE'

A CONVERSATION WITH ANIL ARORA

Anil Arora is Canada's Chief Statistician and the head of Statistics Canada, the key federal agency tracking the social and economic characteristics of the country — and a vital institution for envisioning Canada's future. In recent years, controversies over the previous Conservative government's decision to scrap the long-form census and the current Liberal government's strategy for centralizing information technology in Shared Services Canada prompted Mr. Arora's two predecessor chief statisticians — Munir Sheikh and Wayne Smith — to resign their posts. Mr. Arora assumed leadership of StatsCan in September 2016 after serving in senior management positions with Natural Resources Canada and Health Canada between 2010 and 2016. Earlier, between 1988 and 2010, Arora had held a variety of positions at Statistics Canada, eventually serving as director general responsible for all aspects of the 2006 census, and as assistant chief statistician of social, health and labour statistics from 2008 to 2010. In 2018, Statistics Canada marks the 100th anniversary of its birth — the May 1918 founding of the Dominion Bureau of Statistics. Mr. Arora spoke in mid-May with *Canadian Issues/Thèmes Canadiens*.

Q: Tell us about what's planned to showcase StatsCan's 100th anniversary this year.

A: We started off a few weeks ago with the Governor General (Julie Payette), who came in and addressed the entire institution... We're showcasing the importance of this institute both domestically and internationally by highlighting some of its achievements, some of its firsts, some of its history, some of its leaders. Every single month, we're trying to get people to take pride in what an important democratic institution and world leader Statistics Canada is...

We're modernizing, which is why this institution is world-leading, because it's always looking at where it comes from, where it is and — knowing the winds and changes in demographic, technological and other factors — where it needs to be. It's not only a year to celebrate the journey that we've come from. It's a turning point for us for a number of reasons. We have an amended, strengthened Act (Bill C-36)

and we've got a fairly substantive modernization agenda that we're launching...

Q: Could you elaborate on the mandate change and the modernization?

A: Bill C-36 received royal assent on Dec. 14, 2017... The first big change is that in law now there's a differentiation between the role of the Chief Statistician and the role of the government/minister through which Statistics Canada reports to Parliament. Essentially it separates the *what* from the how. This is an agency that lives and dies by it being relevant to the changing and evolving needs of this nation — its economy, its society, its environment, you name it... So it's entirely reasonable for the government to ensure that the agency is at the table, involved in policy discussions, involved in even the initial scoping and forward-looking agenda items. And then, there's the how. How we go about providing good, high-quality, timely data, and how we communicate to Canadians the operations of this agency based on professional statistical standards, is now — in law — the purview of the Chief Statistician.

Q: If the present Act – the new Act – had been in place when controversy erupted over the Conservative government's scrapping of the 2011 long-form census, would that have unfolded in a different way?

A: Yes, I think it would have made very clear and transparent who made the decision and why it was being made, and then the accountability would have essentially lie where it ought to lie. By convention, we've operated as an arm's length institution. Now, even if there is an exception to that, that would be very clear... It also establishes a Canadian Statistics advisory council in law, with an obligation about the advice it gives the Chief Statistician or the minister, and to publish an annual report. This is another body in the check-and-balance, and for transparency. The other change is in the term and conditions around the tenure of the Chief Statistician — (the appointment is) no longer 'at the pleasure' of the government, but it would be for a five-year renewable term.

Q: Tell us about the modernization agenda.

A: We're an institution that essentially has its roots in survey-taking. You can go back to 1666, when (New France's intendant of justice, police and finance) Jean Talon took the first census here in Canada, around Montreal. We have hundreds of surveys that we've undertaken. We then started to bring in administrative sources of data to supplement and in some cases to substitute for surveys where it makes sense. (Editor's note: for example, income data is not collected via census survey but via tax files).

If you look today, you've got your Fitbit, your fridge, your thermostat (and other data sources). Then there is so much regulatory/administrative data being harnessed. There's so much information people provide over the Web. Look at, today, this explosion of data that is coming at us from all sorts of different sources. And juxtaposed with that is the need for far more (understanding of) complex interactions between social phenomena. For example, what are the implications of housing on health? Or the crowding conditions in a household on a child's educational outcomes and even their life expectancy? All of a sudden, the kinds of questions that

are coming at us are getting more sophisticated and sharper. And if we've only got limited resources to spend, where would they be best spent to get the best outcomes? And I don't know if you've got kids, but my kids aren't going to sit and give you a survey response for a half an hour. Forget it. If it doesn't fit on their iPhone screen and part of something fun, they aren't going to do it. So, there are a number of changes happening...

And that's not to in any way negate people's strong concerns for privacy and confidentiality protection. They want it all. You only want to be bothered (to give information) when it's absolutely necessary and perhaps the least amount of burden. They want us to actually harvest the information that we already have, at the same time protecting privacy and confidentiality. There's this explosion of data; on the other hand, you've got the real sophisticated need for timely information – almost real-time information. People are saying: 'Why are you giving me statistics from a month ago, or a year ago or – God forbid – three years ago? I want to know what's happening today.' Or if a government brings in a policy, they want to know the next day what is the implication of that policy, so they can adjust and tweak.

So, for a statistical agency, these are pressures and opportunities that are impossible to ignore. Our modernization efforts are to react to those pressures, which are very real — and by the way, not unique to Canada. (Editor's note: Arora heads a working group on the modernization of official statistics for the 60-nation Conference of European Statisticians under the auspices of the United Nations).

In Canada, I've launched a five-pillar modernization approach. We need to be far more user-centric; we're a national statistical agency, in law, given the mandate to coordinate, integrate, un-duplicate and look at data that exists out there across the system. We have a legal obligation to protect the privacy and confidentiality (of Canadians). We need to use novel methods and technologies that exist today. Crowd-sourcing is an entirely acceptable and reasonable way for us to get information in real time; web-scraping to get at things like prices and how they're changing over time is entirely reasonable – using scanner data instead of sending people to retail outlets to get at what the (consumer price index) might look like. I could keep going on and on — (another example:) harvesting the mountains and mountains of regulatory and administrative data that exist...

We are not the only game in town – we know that. We're going to have to share and collaborate with those who are producing and disseminating data and information. We are going to have to build the numeracy and literacy skills that are going to be necessary in this context today of the data-driven society and economy. We need to make sure that there's responsible production and use — and informed use — of good data, and knowing where the limitations are and knowing where the biases might reside in various data sources (and) the quality gaps... That's really important as we get into artificial intelligence and machine learning algorithms and so on...

The last pillar, and perhaps for us in the agency the most important one, is a modern workforce in a modern workplace. We can't have the kind of technology and flexibility (we need) if we can't get people to get out of their cubicles and talk to their

colleagues. Like I said, housing is related to health which is related to income which is related to so many other factors. If our people can't get around and talk to others and access data and keep it confidential and move around, then we are not going to attract the expertise and the talent. We are going even beyond that. We're working with partner institutions – the universities and colleges out there... to make sure that data literacy, numeracy skills, good data management are part of their curriculum. And we think we even need to go beyond that to the K-12 system, so that as people graduate (from high school) they have these basic skills that are going to be critical going forward.

Q: How does StatsCan become an even more data-driven organization while making sure it doesn't have too much data about each and every one of us?

A: For 100 years, this is what we've been doing. We're the best in the world at taking two pieces of data collected through different sources with identifiable information (and figuring out) how you bring them together, how do you anonymize them, how do you link them and then get at the insights. This is what we do day in, day out. We're the world's best at those kinds of methodologies. The privacy commissioner has said essentially that this is the gold standard.

Q: What happens to that every-five-year head count? If information about Canadians can be collated from administrative and other sources, will the census possibly no longer happen?

A: A very interesting question, obviously... Can we see, one day, us moving to a synthetic or real-person register?... This is what happens in Scandinavian countries, and in many European countries, where every individual is given a unique identifier and that's how the system then is tracked. But that's not something a statistical agency would impose upon Canadians. But (if a system like that was in place), and we had, then, the authority to go in and keep it up to date, and nurture it with other pieces of information, then certainly we could see one day where we could substitute going to Canadians to ask that information. But that's not a decision for the statistical agency to make on its own...

The second thing to that is that even in our short form, we ask a number of questions other than just simply your name and date of birth and your gender. We also ask the language questions, etc. For some of those questions, it's not clear yet where we would get the equivalent amount of information. I still think that some form of going out and getting information from Canadians in a cost-effective way is still going to be there for quite some time.

I essentially was the leader of the 2006 census where we redesigned the whole mechanism by which we take a census in Canada. Many other countries have leveraged on that methodology since then... Now we don't even mail out questionnaires to the majority of Canadians – we send them a letter with a code. And in this last (2016) census, just under 70% of Canadian households filled out the questionnaire online. Plus, we have always continued to find the most cost-effective way to get that kind of information. So, in that equation, we also have to look at the overall cost-effectiveness of getting information. Information needs continue to grow. Every census, when we go out to consult — you won't believe — we get at least a thousand

new topics where people want us to ask a question of Canadians. If truly the appetite for information from Canadians is insatiable, we have to look at what is the best mechanism to get that. The census kind of fulfils a need.

Q: What do you think will be the biggest difference between how information is collected and used now and how it will be in 2067?

A: We're really talking about a crystal ball here. But let's look at those faint signals out there that could potentially be the predominant forces. We are seeing the next generation putting a lot more value on that quid pro quo — that value proposition (between gaining insights from data and protecting personal privacy). They're willing to give a little to get a little. We all do that, too — e.g. online banking.

I do see a day when, instead of us putting information out that says, 'Look we were able to give the correlation between an overcrowded house, the income of the family, and the education, health and income-earning outcomes,' as a *nice-to-be-able-to-do* (analysis), to really where the population is going to say, that is the expected (analysis)...

Just like the technology of ride-sharing (e.g. Uber) or accommodation services (e.g. Airbnb), business models have evolved around data essentially being the currency, and even shaking traditional institutions and businesses. I think that will just accelerate over time. I think governments will be challenged with their traditional roles. (*Editor's note: Arora referenced here the Sidewalk Labs-led redevelopment of part of Toronto's waterfront in a new private-public model for designing urban neighbourhoods*).

I think it's raising questions about what is the role of government. I see those kinds of models are challenging institutions, whether academic or government... That's why I think our modernization plans are going to be so important for us to be able to look at those (challenges) and make sure we are adding value. It's now moving less to production (of raw data) and more to the expertise and analytics, being able to communicate and being able to explain.