

**OPENING KEYNOTE BY ANIL ARORA
CHIEF STATISTICIAN OF CANADA**

— — — — —
**INFONEX: BIG DATA AND ANALYTICS
FOR THE PUBLIC SECTOR**

Opening Keynote by the Chief Statistician of Canada Infonex: Big Data and Analytics for the Public Sector October 1, 2019

Introduction

Good morning. I am Anil Arora, Chief Statistician of Canada. I am thrilled to be here today to talk about Big Data and Analytics.

As you know, we are all part of, and contribute to, a data-driven society. Numbers, metrics and the endless ways we have come to quantify the many aspects of our society are now a part of everyday life. We use our phones, or other devices, to track how many steps we take, how many calories we consume and how much money we spend. Pacemakers have embedded AI, which warn and adjust for optimum performance data and literally help save lives today. This endless stream of data surrounds us, and as a society we recognize now how valuable a commodity data has become.

Today, I would like to speak to you about five topics:

- Volume does not equal quality;
- Using data as a strategic asset;
- Exploring new and enabling technologies responsibly;
- Collaborating to support innovation; and
- Recognizing the critical role of official statistics for our economy and our society.

Volume vs. quality

Let's talk about volume. The sheer amount of data is becoming increasingly prolific. According to Forbes, 2.5 QUINTILLION bytes of data are created each day. It's hard to wrap your head around this figure. The IST will only increase this amount. More than 3.7 billion people use the Internet.

Which makes you wonder, how do we leverage all these data into useful information?

Well, when we are talking about Big Data, we need to recognize that volume does not equal quality. You can have boatloads of information, but it is not useful to anyone until it is analyzed. It is only when data are analyzed that they can be used to make decisions.

It is easy to get caught up in data and metrics, especially for businesses. In fact, a recent article in the Harvard Business Review explains how *too many leaders confuse numbers with strategy*. It warns leaders not to rely too much on metrics as they can lead us astray from our strategy. This is a really interesting point, because it shows how a

lack of analysis can negatively impact businesses or any other type of organization. Too much data without the means or knowledge to use it periodically could be detrimental.

Data as strategic asset

Sound analysis gives real insights into data. These insights support informed decision making as well as the development of policies on real issues that affect our day-to-day lives. Anyone, and in fact a device, can pump out numbers, but you need to include that element of human expertise to unlock the true value of data – to use it as a strategic asset.

So, what is the role of Statistics Canada, the national statistical agency, in all this? Well, that expertise is exactly what we do best. We think of ourselves as the people, or the geeks to some, behind the numbers.

One clear example of data as a strategic asset is the need for data on the opioid epidemic. As you may be aware, this is a widespread problem in Canada, with severe physical, psychological, social and economic repercussions. We collaborated with the City of Surrey to get data from first responders on people who unfortunately lost their lives due to this crisis. We were also able to get data from provincial organizations that showed what happened to these individuals six months or a year before these individuals passed away, such as interactions with health services or the judicial system. The result is, data can be used to break down stigma and truly understand what the crisis is about and what policies and programs would best equip us to address it, to inform us about timely and thoughtful interventions.

In such cases, the contributions of national statistical offices, what we call NSOs, are very apparent. To support the information needs of society, NSOs cannot function solely as data providers. We must provide the expert analysis to solve today's complex problems in society.

I would like to take a few moments to touch on some of the ways we at Statistics Canada have been adapting and realigning ourselves in the era of Big Data to continue to fill this role.

Changing expectations and user needs / new technologies

Our agency has been releasing facts for over 100 years. We recognize that times have changed and that Canadians' relationship with data has changed. Expectations continue to rise. They want more information, they want it tailored to meet their individual needs, while at the same time having assurance that their information will be properly safeguarded.

We have moved beyond talking about the need to modernize and have taken action to adapt to this new environment. We have firmly embedded modernization into our way of

working, our way of thinking ... into our culture. It is what propels us forward, puts us in a better position to respond to the changes around us.

Because it's not just about modernizing our processes. It's about changing our perspective on how we serve people like you and Canadians involved in the public policy process, in business, academia, associations, and others who contribute to making our communities better.

Continued improvement, for an agency that has been around for 100 years, is a must as we recognize that we are not the only players in town. Many businesses, governments and organizations are delving deeper into the world of data, and producing or providing their own data. We at Statistics Canada see this as an opportunity to transform ourselves to best serve the public and business sectors.

We have been improving the timeliness of publicly available data by finding new information sources, developing new methods and processes and collaborating more closely with partners. We continue to focus on producing core insights that data users rely on, such as the Consumer Price Index or the Gross Domestic Product. And we are providing Canadians with new tools that make the use of statistics more relevant, such as our recently launched International Trade Explorer Tool—a suite of four interactive data visualizations to discover Canada's trade relationships and how they change over time.

And while we are committed to producing high-quality data, we remain committed to protecting the privacy and confidentiality of Canadians. Maintaining the trust of Canadians is the core of our business and everything we do. This is our currency and has been for over 100 years. We take the privacy of Canadians very seriously and we safeguard the data we collect. Data collected from new sources are only ever used for statistical purposes and treated with the same degree of scrutiny, confidentiality and security as data collected through traditional methods.

Administrative data

One of the ways we are improving the timeliness of data is by responsibly using new data sources, such as administrative data. In fact, Statistics Canada has been using and protecting administrative data since its creation.

These types of data offer an opportunity for collaboration between producers of such data and us at Statistics Canada. Admin data can be a "win-win" for everyone as we can provide insights faster and more frequently. We can also reduce response burden and costs by moving away from traditional surveys. These opportunities for collaboration enable us to benefit from each other's expertise. In addition, it allows us to be more user-centric, while maintaining our role as the trusted broker of key information for Canadians.

AI and machine learning

We are also embracing AI automation and machine learning:

- AI was used to analyze 1.1 million comments on the 2016 Census to inform future censuses.
- We see functions such as coding of survey responses being fully automated with greater accuracy and efficiency in the future.
- We can also enrich data through record linkages.

Necessity and proportionality

Another way we are meeting this challenge is through our work in consultation with statistical and privacy experts from around the world to develop a new methodology framework based on the principles of necessity and proportionality.

First, we establish necessity. Statistics Canada is mandated to produce data for Canada—data that are essential for governments, municipalities, businesses and individuals. We measure our society, economy and environment.

Then we use proportionality, which means that when we plan surveys, our experts work to balance the volume and the sensitivity of sources of data with the need to reduce the response burden on individuals, all while maintaining protection of your privacy

The framework expands on the principles that have always guided Statistics Canada. We are now modernizing our approach by expanding on these principles to develop a scientific framework. This framework explicitly assesses proportionality and data sensitivity, while ensuring statistical values, such as privacy and confidentiality are respected.

Collaboration / innovation

Now that you've heard a little about how we are modernizing, I would like to demonstrate how this value has truly embedded itself in our agency's culture, and manifested itself in various ways across the agency. One might say that modernization has paved the way for innovation and collaboration.

Innovation Ecosystem

To foster a strong culture of innovation and continuous improvement, Statistics Canada has created an Innovation Ecosystem.

This Ecosystem has been designed:

- to empower employees to experiment, to connect ideas and people;
- to develop and share expertise;
- to provide infrastructure, tools, coaching and support; and
- to leverage innovation to deliver on our goals.

The Ecosystem is comprised of a Digital Innovation team, an Innovation Centre and Lab, an Innovation Radar, monthly hackathons, centres of expertise and a Research and Development Board.

LEAN culture, cloud sandbox, hackathons

We have also begun a *LEAN Culture* Initiative. The goal here is to promote open and transparent communication with partners and stakeholders, grow a culture of innovation and encourage innovative mindsets and behaviours through recognition. To support this, we have a dedicated LEAN coach and in-house LEAN training. For those who might be wondering what LEAN means, it is the state of an organization that always strives to deliver value to clients.

We recognize that inclusion is a primary success criterion for innovation. Successful experimentation engages and leverages diverse and multidisciplinary talent, knowledge and experience. Whether it's from hosting regular hackathons or creating an Innovation Cloud sandbox, we continue to build our ecosystem.

We are doing this in a way that focuses on engagement with Canadians. For example, our Gamification hackathon challenged participants to come up with unique ways to “gamify” our data to make it fun and accessible for more people to use.

Governance

So, you might be wondering, how does a large organization manage all these changes and modernization initiatives with stringent timelines and fixed resources?

Another step we are taking to support modernization is to re-think our information technology infrastructure. Statistics Canada has organized its IT workforce to focus on a cloud-first strategy to form—in close collaboration with Shared Services Canada—the basis of a modern and secure IT infrastructure, to migrate away from legacy data centres to operate on the Cloud.

Components of the data strategy (domestically and internationally)

The Statistics Canada Data Strategy outlines five capabilities we are focusing on:

1. Data Governance
2. Data Stewardship, which includes:
 - Data Discovery
 - Data Digitization
 - Data Interoperability
 - Data Management
3. Data Resources
4. Data Trust Framework
5. Data Leadership

The high-level goals of the Strategy are to ensure that data are findable, accessible, interoperable, reusable, reproducible and open.

And some of you may know, all of this aligns closely with the Government of Canada's own data strategy: *A Data Strategy Roadmap for the Federal Public Service*. Statistics Canada has been at its forefront, and here to support it.

This strategy offers a huge opportunity to create unified and open data resources across the federal government. Our modernization work aligns well with the key challenges and elements of the data strategy and will be a major driver for many of the key elements public-service wide.

Through the processes and discussions we held to determine how we could best contribute to this federal strategy, we noted three key lessons:

1. Be a proactive and adaptive organization
2. Engage Canadians and Canadian organizations
3. Foster strategic partnerships – think national statistical system

Priorities for Data Governance in Canada (domestic and international collaboration)

As we continue to move forward, our priorities for data governance are:

- A governance structure that is more responsive to the ever-changing environment, including the data revolution and legislative changes like the legalization of cannabis;
- Collaboration with partners, nationally and internationally, with the intent of promoting interoperability;
- Alignment with the Data Strategy for the Federal Public Service.

A priority at the national level is to establish a governance structure to develop and revise standards that align with the needs of government to assist with evidence-based decisions.

In the past, Statistics Canada followed a five-year life cycle for the principal classification systems. What we discovered was that as our data series approached the five-year mark, the data themselves did not align with the socio-economic landscape at that time. In order to adjust to this rapidly changing environment, Statistics Canada decided to be proactive and initiated an evergreen approach.

Cannabis

For example, this way of doing things was used to address the need for data to measure the impact of the legalization of cannabis within Canada. Revision to nine of the major classification systems were released to correspond to the legalization. There are now specific categories within these classifications that will assist in measuring the

impact on industries, products, occupations, imports/exports, and educational programs to name a few.

National Occupational Classification

Another example of where the evergreen approach will assist is in the changing occupational landscape. As Canada becomes more information-based and as technology moves towards more AI and machine learning, we are hearing from our data users that the National Occupational Classification needs to be updated more frequently to support new occupations. Statistics Canada has been a collaborative partner with Employment and Social Development Canada in developing and managing the National Occupational Classification, a systematic classification structure that categorizes the entire range of occupational activity in Canada. Occupational information is of critical importance for the provision of labour market and career intelligence for numerous programs and services and we look forward to maintaining this relationship and working closely to meet our data users' needs.

High-level Group for the Modernization of Official Statistics

It is not only the standardization of the classification systems that are important but also the information models to help define the processes from collection to dissemination of data and the structure of the information that define the business. Statistics Canada plays an active role and has a strong international presence regarding standardization. Currently I chair the High-level Group for the Modernization of Official Statistics – managed by the United Nations (UN) Economic Commission for Europe. The High-level Group is responsible for determining the annual international collaboration projects that align with the UN statistical modernization programme.

One of the four modernization groups that reports to the High-level Group is the Supporting Standards Group, which is responsible for the development and maintenance of information models needed for statistical modernization. These modes include the Generic Statistics Business Process Model, the Generic Statistics Information Model, the Generic Activity Model for Statistical Organizations and the Common Statistical Production Architecture. Statistics Canada is active on the Supporting Standards group and on three of the four supporting task forces. These information models are used in developing systems across processes that allow for interoperability.

Data Strategy for the Federal Public Service

Alignment with the Data Strategy for the Federal Public Service is another key priority. Government of Canada departments and agencies are faced with many of the same challenges with respect to data governance. Enterprise data governance is needed to support the implementation of the Data Strategy for the Federal Public Service and to drive cultural change and the strategic use of data.

The Data Strategy summarizes some of the Government's key challenges as being: absence of horizontal governance for strategic direction on data issues, lack of data literacy and cultural reticence to break silos, lack of adequate digital infrastructure and a complex rules framework, and the challenges of acquiring, governing and managing large volumes of data.

There is a need to start looking at the government as a single entity, collaborating and balancing the need to share across organizations in a timely manner while respecting the privacy of and ethical responsibilities to Canadians. We need to investigate trustworthy alternative sources that can enhance our current work and fill current data gaps. This will require a framework to assess quality along with standardization of concepts to allow for the integration of information, while protecting privacy and maintaining public trust.

A proactive rather than reactive public service which responds to the evolving data ecosystem and the needs of Canadians will require strong governance systems that can easily communicate with one another, standard concepts and information models used to provide the foundation that will allow for the integration of reliable information.

Canadian Data Governance Standardization Collaborative

As you may have heard, 90% of the world's data was generated in the last two years. As the supply of data increases exponentially, it only makes sense to adopt standardization strategies at this stage.

This spring, Canadian subject-matter experts began planning a Canadian Data Governance Standardization Collaborative to speed up the development of industry-wide governance standards. The goal is to foster coordination and collaboration on data governance standardization issues, including definitions and classifications. Statistics Canada is a member of this collaborative which will include industry; governments and bodies that have a legally recognized regulatory function; public interest; academic and research bodies; and standards development organizations.

Our agency works with stakeholders to determine the type of standards needed and the process for developing the standards. Methods for governing the data need to be modernized to benefit Canadian organizations and citizens.

Clearly, there is a need for classifications and other metadata to be revised more frequently to adjust for the changes in the Canadian economy and society.

And while we begin work in this area to improve information flows, interoperability and increase the availability and timeliness of information, we must ensure that privacy is embedded into the process and not an afterthought or seen as a trade-off!

Canadian Research and Development Classification

Statistics Canada is collaborating with the Canada Foundation for Innovation, the Canadian Institutes of Health Research, the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada on the development of a new single Canadian Research and Development Classification 2019. Among other things, this shared standard will be used by the federal granting agencies for collecting and managing administrative data on research and development programs or grants, and for statistical purposes by Statistics Canada. It will also align Canada with international research and development classification standards.

NAICS

Standard classifications are important at the international level. Statistics Canada, along with the statistical agencies for Mexico and the United States, developed the North American Industry Classification System. This system was designed to provide common definitions of the industrial structure of the three countries and a common statistical framework to facilitate the analysis of the three economies in response to the implementation of the North American Free Trade Agreement. We use this system when industry classification is required in our programs. It is also used by Canada Revenue Agency, including in tax software.

We will continue to work with international partners to ensure that data are comparable across countries.

And why are we doing all of this? Why are we making significant changes to re-align ourselves internally while at the same time forging ahead with innovative collaborative approaches? Because there are enormous benefits to be had for our society and our economy if we find ways to unleash the power within that data.

So is this all good theory?

The Power of Big Data: Usefulness of statistics for society / economy

Allow me to give 5 concrete examples of Statistics Canada's initiatives from across the public sector that use Big Data for data analytics. They almost always include a collaborative approach where we share information with our federal partners to support their policy work.

Example 1 (Immigration): Integrating big structured data, such as administrative data, to better understand the outcomes of immigrants.

- Did you know that, since 2016, Canada welcomes on average 300,000 newcomers to our country per year – policy makers need sound analytics to ensure they are developing good policies that support successful integration.

- Statistics Canada is using integrated administrative data to support analytics in this space.
- For example, home ownership is a key milestone in the social and economic integration of immigrants. New housing data from the Canadian Housing Statistics Program was used to track home ownership among immigrants. Results show that:
 - In both Toronto and Vancouver, immigrants were less likely to own single detached housing compared with Canadian-born homeowners
 - Home ownership among resettled refugee families is generally lower than among Canadian-born families – in Vancouver, 50% among resettled refugees compared with 61% among Canadian born families – lower rates are largely due to lower incomes.
- Immigrants also contribute to the economic engine of the country. Using integrated data from Canadian Employer-Employee Dynamic Database, researchers at Statistics Canada found that immigrant-owned companies were younger, more likely to be high-growth firms, and contributed a higher share of net job growth.

Example 2: Predictive analytics to guide options for cancer screening

- In an era of increasing health care costs, providers are looking to make the most effective use of resources – predictive models can help to test the impacts of different policy and program options.
- Working with the Canadian Partnership Against Cancer, Statistics Canada created OncoSim, a predictive model using a range of data sources that estimates the impact of different options for cancer screening and treatment.
- The lung cancer model was recently used to estimate the impact of introducing a lung screening protocol to screen those aged 55-74 years of age who have smoked the equivalent of one pack per day for 30 years or more.
- Results of the model show that over time, the organized approach would save health system costs by reducing the need for lung screening scans and diagnostic procedures by more than 80%. This could be done without compromising health impacts, namely the rate of cancer incidents.
- The model is being used by cancer programs across the country to plan better services for patients.

Example 3 (Education): Understanding labour market outcomes of post-secondary grads: Can we find out who is employed following graduation?

- Investments in post-secondary education are significant for students and families, so understanding the labour market outcomes following graduation is important.
- Statistics Canada is using integrated education and tax data to track the labour market outcomes of graduates.
- Results reveal that, among students graduating in 2014, 66% of them entered the labour market and 20% returned to school full-time.

- The areas of study with the highest median employment income two years following graduation include architecture and engineering and health and related fields, as well as business, management and public administration.

Example 4 (Municipalities) : Using open data to help cities meet their Sustainable Development Goals as established by the United Nation’s 2030 Agenda for Sustainable Development.

- One of these Sustainable Development Goals focuses on making cities inclusive, safe, resilient and sustainable.
- The goal also focuses specifically on measuring the proportion of the population that has convenient access to public transportation – meaning that bus stops can be found within 0.5 kilometres of the places citizens live, work, study and shop, and that there is frequent service during peak travel times.
- To help cities meet these goals, Statistics Canada is working with open data, specifically the *General Transit Feed Specification*, a protocol that holds live-time fleet management that report the GPS position of their assets; as well as make publicly available the latitude and longitude of static transit stops and their respective schedules.
- Using these open data with our own data, we are able to estimate the proportion of the population that has access to convenient public transit.
- This information will help cities to track how they are doing and to identify areas where improvements could be made, resulting in better transportation services for citizens.

Example 5 (Agriculture): Using satellite data to provide timely information on crop yields

- Another example of how Statistics Canada is using innovative data collection methods and technology, while adapting to better meet the needs of users, is Ag-Zero.
- Ag-Zero is an exciting initiative for our Agriculture Statistics Program that aims to collect the program’s required data through administrative data sources and satellite imagery by 2026.
- This is another example where my colleagues at Statistics Canada have shown that they are listening to feedback from Canadians who want timely, accurate and detailed data, while completing fewer surveys.
- This modernization initiative leverages alternative data sources, such as Earth Observation technology and historical crop insurance data for Manitoba, and advances in data processing techniques to get the information to farmers while reducing the time they need to take to respond to statistical questionnaires.
- Statistics Canada will be piloting this innovative approach which can also benefit other organizations in the value chain.
- For example, imagine how the data insights of AgZero can help public policy makers and companies work faster with farmers after a significant weather

event impacts crops, and governments and insurers need to quickly deliver support to farmers.

Like other national statistical offices, we account for changes that result from legislation, and adapt greatly to support policy-related information needs.

Take the legalization of cannabis. We looked at ways to get data on production, consumption and prices, but cannabis was illegal at the time, and this information was outside the scope of our statistical systems.

So, we had to use some pretty out-of-the-box methods.

- One was crowdsourcing:
 - We launched a site called **StatsCannabis** and asked people to tell us how often they consumed cannabis, and what they paid.
 - We got over 20,000 responses in the first month! This is one of reasons why trust matters for a statistical agency. Canadians trusted Statistics Canada with this sensitive information.
- We also began using municipal wastewater analysis to learn about consumption in different municipalities.
 - We did this through a technique called wastewater-based epidemiology.
 - This new technique allowed us to tackle the challenge of under-reporting. This is something we always face when trying to produce data through things like surveys.
- We're also starting to integrate cannabis sales into retail sales data.

United Nations Global Platform for Big Data for Official Statistics (collaboration)

At a higher level, there is also recognition for the need to standardize and collaborate. A global collaboration to harness the power of big data to better lives demands that we align to facilitate our work internationally.

Take for example, the UN Global Working Group on Big Data for Official Statistics, established in 2014. It is mandated to give direction on the use of big data for official statistics. The UN Global Platform allows us to work and learn together—we have the same challenges and can collaborate on capacity-building and training. We can benefit by having access to new data sources, technologies and expertise.

Canada is a member of this working group, along with 27 other member states and 16 international organizations. The working group saw the need for a platform to provide researchers a place to collaborate, share ideas and work together on projects, while using trusted data and trusted methods. Such collaboration drives innovation, especially since knowledge sharing has traditionally been limited to presentations and papers. With the digitization of the information and communications technologies and the rise of virtual communities across the world, Big Data is emerging as an important opportunity

for evidence-based policy making, which can be especially relevant on the international scale.

This Big Data working group serves governments, business, academia, media and other data users with the goal of providing better data for a better life.

By providing members with direction on the use of new data sources and corresponding new technologies, services and applications for official statistics, globally, the group aims to keep official statistics relevant in a fast-moving data landscape by making them timelier, more granular and more frequent; and strives to find world class data solutions for the compilation of statistics and Sustainable Development Goal indicators by connecting people, global data, new methods and algorithms, and the latest cloud and software technology services.

The ultimate goal is to better support evidence-based decision making at local, national and global levels for the sustainable development of the economy, environment and society.

Canada is involved in several task teams, including the Task Team on Scanner Data. This team is looking at how to increase the use of scanner data in official statistics. For example, using scanner data from retailers to help in the calculation of price indices.

Conclusion

I would like to take this opportunity to thank each of you for being here today to partake in this exciting conference. We are going to delve deeper into Big Data and what it means for you.

Our speakers over the next two days will talk about a diverse range of topics, from AI, machine learning and bot technology, to IT auditing, lifestyle analytics, data visualization, open data, and so much more.

I am looking forward to presentations from two of my colleagues as well. Eric Rancourt, the Director General of the Methodology Branch at Statistics Canada, who will give insights into different types of data—their possibilities and their limitations—as well as the approaches we take at Statistics Canada to assess data quality.

We will hear from Monica Pickard from the Data Science Division, who will talk about how to move forward confidently with big, unstructured data.

In conclusion, I would like to leave you with a few take-aways...

- Recognize that volume does not equal quality. There is a whole process behind producing good data that can support evidence-based decisions.
- Use data as a strategic asset. To gain real insights, leverage the knowledge of experts at Statistics Canada.

- Explore new technologies. As user expectations continue to rise, these new possibilities will enable us to keep up.
- Collaborate to support innovation. Work together to find innovative solutions to challenges, whether that is by collaborating with stakeholders or embracing data literacy as a skill.
- Recognize the usefulness of statistics for our economy and our society. Expert data analysis can lead to better health services, job creation, smarter businesses...The possibilities are not yet fully known.

As we all know, the world of data is changing so rapidly that it can be hard to keep pace. Let's all take this opportunity to learn from one another as we discover the latest innovations, the latest technologies and the latest challenges surrounding Big Data. I know I am excited to hear more about this topic, and I hope you are too! Data is about opening your mind up to possibilities, and I look forward to exploring those possibilities with you.

Thank you.