Data for Good:

The $32 Billion Boost

The first of a three part series

This report was made possible by the generous support of our sponsors

AI associated with CRM activities will boost global business revenue from the beginning of 2017 to the end of 2021 by $1.1 trillion. The potential GDP boost for Canada is $32 billion.[1] What follows is the Canadian Chamber of Commerce's exploration of the critical intersection between prosperity, technology and privacy. The goal of this effort is to answer a critical question: **Will Canada act or be acted upon by the Fourth Industrial Revolution?**

This first release in our series, *Data for Good: The $32 Billion Boost*, is about context.

Over the past year, we have collected and distilled the thoughts of dozens of industry experts through round tables and personal interviews on why getting the data framework right is so important. Here, we present the argument that favours freedom to operate in the face of increasing pressure to regulate the collection, use and disclosure of personal information. Our intent is to demonstrate how personal data is used to innovate and create products and services that improve people's lives and how Canada can thrive in the global digital community.

---

[1] A Trillion-Dollar Boost: The Economic Impact of AI on Customer Relationship Management, IDC, June 2017 p. 16

The Age of Data

In the next two years, 40 zettabytes of data will be created, roughly equivalent to four million years of HD video or five billion Libraries of Congress.[2] In the last 15 years, the flow of data across borders has experienced a 45-fold increase and is predicted to grow another nine times by 2021. Data flows now exert a larger impact on GDP than trade in goods.[3]

Data, and the techniques and technologies employed to collect and analyze it, have the potential to allow Canada and the world to solve some of the world's most pressing economic, social and environmental problems. Data is now the engine of economic growth and prosperity, and countries that promote data's availability and use for societal good and economic development will lead the fourth industrial revolution and give their citizens a better quality of life. To seize this opportunity, people and businesses need to be able to share their data with one another.

Canada's Data Story

Canada is wired and connected with an internet savvy population. Ninety-eight per cent of the Canadian population are internet users. Canada ranks 14th in the World Economic Forum's Network Readiness Index (NRI). The NRI praises Canada for making it quick and easy to start a new business and for also fostering a highly-skilled workforce that compliments such a forward-thinking economy.[4] The shift from traditional trade in goods to digital trade has leveled the playing field for micro-sized businesses to compete on the global stage alongside multi-billion dollar multi-nationals. Shopify, an Ottawa-based leading multi-channel commerce platform, now powers more than 500,000 businesses in 175 countries around the world.[5]

The right framework for Canada's data economy is one that both creates trust and incentivizes innovation. We cannot ignore the role that companies that have not even yet been created will play, as creators. Canada's framework for the data economy must help SMEs grow and become world leaders in the digital space.

---

[2] https://www.weforum.org/agenda/2018/01/data-is-not-the-new-oil/

[3] https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/digital-globalization-the-new-era-of-global-flows

[4] http://www3.weforum.org/docs/GITR2016/WEF_GITR_Full_Report.pdf p.76

[5] https://www.shopify.com/press/releases/shopify-now-powers-over-500-000-businesses-in-175-countries

From Datum to Data

Datum is simply bits of information that have been collected and stored. It can be structured—names, phone numbers, email addresses stored in a customer relationship management program—or unstructured—conversations in an online chat forum, web-surfing habits, video content. The value of that information is totally dependent on context. Data is datum in context.

For instance, an email address might be valuable as a way to distribute advertising but only if the recipient has an interest in the product or service being advertised. Mass emails distributed to uninterested recipients can actually have a negative impact on a business's reputation.

The smartphone you carry around in your pocket is a treasure trove of data points. Sort of. The places you travel to, the roads you use to get there are important to car companies and retailers. The shops and websites you visit and what links you click on, the items you purchase and which search engine you used matter to marketers and advertisers. The stories and pictures you share on social media and the games you play and how often are immensely important to the businesses behind those shops and websites and platforms and games because they tell a story about you.

But in isolation, your story alone does not hold much value because mass production of products nd services are not built around the preferences of one individual. Analytics programs use statistics to arrive at conclusions or predictions and have shown potential for a wide range of data mining applications. For example, they have been used in speech recognition, digital forensics, protein sequencing and bioinformatics.

But first comes the hard part of crunching the data. Success here demands both quantity and data quality. Data, argues Hal Varian, Google's chief economist, exhibit *"decreasing returns to scale,"* meaning that each additional piece of data points become less valuable and at some point collecting more does not add anything. More important is the quality of the algorithms that crunch the data and the talent a firm has hired to develop them.
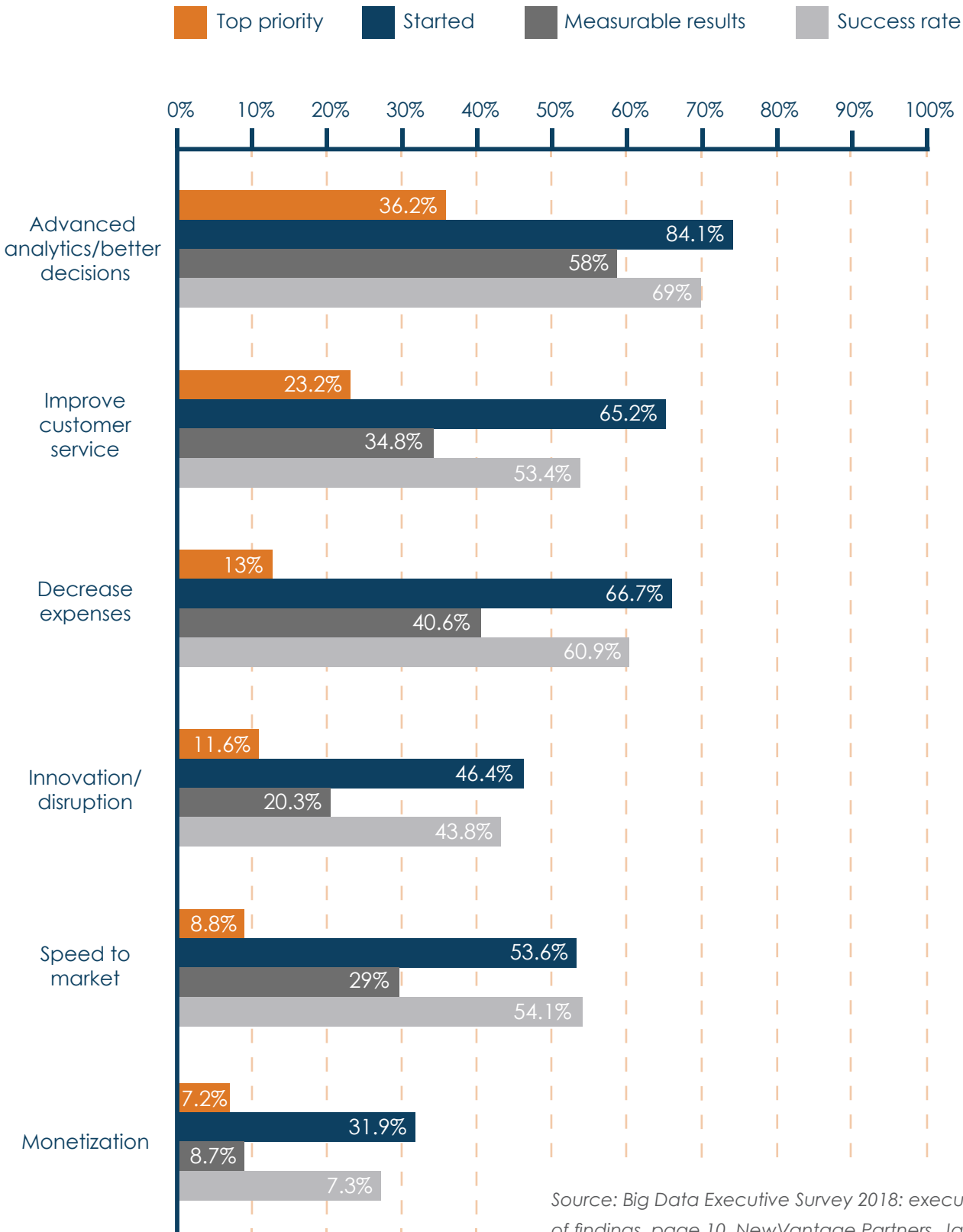
Here is what good data quality looks like:

- The data is as complete as possible.

- Each data record is relevant and unique.

- The data is accurate and current.

- The analytics model uses consistent patterns and terminology.

The Business of Data

By combining an email address with purchasing behaviour that is captured in a way to include the good data qualities, the distribution list becomes more relevant. Yet not relevant enough. To be truly relevant requires finding the patterns that emerge from the purchasing habits of tens of millions of people. Data points about when and where all those 10 million people shopped, what items they purchased together, what information they checked out online first, the length of time between purchases and the locations they purchased at are then parsed by demographic data, those detailed customer profiles become far more useful in product development. For instance, predicting what products to manufacture, the quantity needed in different parts of the world, how and why market share is shifting and how to focus the innovation effort.

Corporate Big Data Initiative Success Rates: U.S. and Worldwide 2018

■ Top priority ■ Started ■ Measurable results ■ Success rate

Advanced analytics/better decisions
- Top priority: 36.2%
- Started: 84.1%
- Measurable results: 58%
- Success rate: 69%

Improve customer service
- Top priority: 23.2%
- Started: 65.2%
- Measurable results: 34.8%
- Success rate: 53.4%

Decrease expenses
- Top priority: 13%
- Started: 66.7%
- Measurable results: 40.6%
- Success rate: 60.9%

Innovation/disruption
- Top priority: 11.6%
- Started: 46.4%
- Measurable results: 20.3%
- Success rate: 43.8%

Speed to market
- Top priority: 8.8%
- Started: 53.6%
- Measurable results: 29%
- Success rate: 54.1%

Monetization
- Top priority: 7.2%
- Started: 31.9%
- Measurable results: 8.7%
- Success rate: 7.3%

Source: Big Data Executive Survey 2018: executive summary of findings, page 10, NewVantage Partners, January 2018

Value for Value: The Exchange of Your Information against Personal Benefit

Direct Value – Loyalty Programs

Loyalty programs are a great example of the evolution of an exchange of personal data for something of value—from a card with stickers that lead to a free coffee to frequent flyer programs that track consumer behaviour in exchange for tangible goods in the form of points. Loyalty programs have been around for decades and started out as a way to retain customers.

Loyalty programs work because people like them and derive value from the exchange. The success of such programs is explained by the Conservation of Resources theory[6], which asserts that individuals strive to obtain and preserve resources they value, and the loss of resources results in stress. Peter Danaher applied this theory to loyalty programs and concludes that loyalty program points reflect a resource that is valued by an individual which they can use to obtain other objects of value.[7]

The shift from a stamp on a loyalty card at the local shop to digital data capture has enabled a massive amount of data points and an infinitely more sophisticated and robust platform with which to make predictions of consumer behaviour. The massive increase in data points and analytical accuracy are the drivers of value for product developers and advertisers.

Indirect Value – Search Engines

Search engines are another great example of the evolution of exchange of personal data for intrinsic value. Except this time, it is a service. A highly effective and convenient service, one that virtually every person in western society relies on every day to find information he/she needs. Search engines help us find educational material to research our burning questions, meet people, explore different cultures, comparative shop and help us collaborate. While the internet provides the conduit that enable us to communicate, it is the search engines that allow us to connect the dots—all seemingly for free.

Every request sent to a search engine has meaning which, to the company providing the search engine has value. Search engines store records of search terms, IP addresses, browser used and which items are clicked, all in an effort to better understand preferences and improve search results.

The volume of search engine data is staggering:

- There are over four billion internet users and close to two billion websites in the world.

- 10 computers and 50 smartphones are sold every second.

- There are over two trillion Google searches per day.[8]

- 16% to 20% of queries that get asked every day have never been asked before.

- Every query has to travel on average 1,500 miles to a data centre and back to return the answer to the user.

- A single Google query uses 1,000 computers in 0.2 seconds to retrieve an answer.

[6] Hobfoll, Stevan (1989). "Conservation of Resources. A New attempt at conceptualizing stress". The American Psychologist. 44 (3): 513–524

[7] http://sydney.edu.au/business/__data/assets/pdf_file/0006/328272/Dynamic-Management-of-Rewards_Peter-Danaher.pdf

[8] https://ardorseo.com/blog/how-many-google-searches-per-day-2018/

Percentage of Global Population Using the Internet
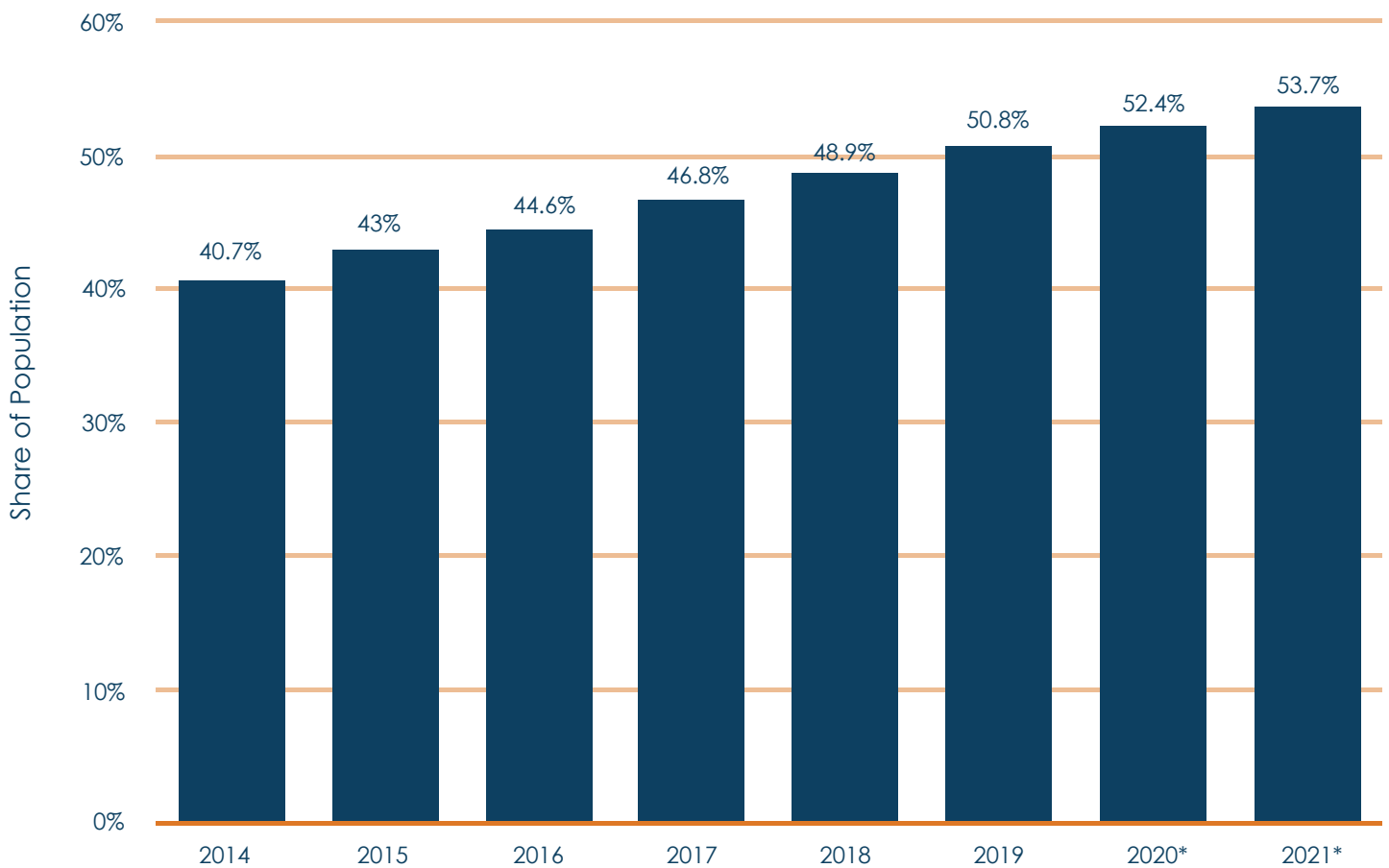
There are reasons for collecting and using this data that benefit user experience. Algorithm-driven approaches to things like spell check, translation and better search results rely on the data mined from user search patterns. Without the algorithms, products like Google Translate would not be able to support less-used languages like Catalan and Welsh. Data is also mined to craft countermeasures to web irritants like link farms.[9]

---

[9] https://arstechnica.com/tech-policy/2010/03/google-keeps-your-data-to-learn-from-good-guys-fight-off-bad-guys/

How Data Is Being Used for Good

Data provides the intelligence that product developers rely on to improve efficiency and efficacy. Communications, connectivity, infrastructure, safety and well-being are all enhanced by data. The world is more connected than ever. Not just internet users but also machines. Networks of cameras, sensors, vehicles and mobile applications are all feeding decision-making. Newspaper editors are making decisions about content and placement of stories that are informed by online reading habits, bringing the most relevant and interesting facts to your attention. The GPS on your phone can now provide you with alternate routes that avoid traffic jams because the system recognizes a pattern of congestion of mobile phones. Railway crossings are safer because machines are able to alert trains of unusual behaviour at intersections. Speech recognition has decreased wait times at call centres and enabled us to communicate our needs without a keyboard interface. Capturing driver behaviour has enabled vehicle parts manufacturers to design parts that are safer and more resilient.

Fraud and Risk Detection

Credit card fraud detection in milliseconds helps financial services firms protect their customers' security while reducing loss due to fraud. Pattern recognition and anomaly detection can identify fraud, waste and abuse for healthcare insurance companies.

Protect the Vulnerable

Simon Fraser University researchers have devised a way to use artificial intelligence to capture the essence of an individual in a video clip—much like an impressionist painting—without revealing the identity of the individual.[10] This makes the subject of an interview far more relatable to the audience than the typical blanked out or pixelated solutions currently in use for video journal interviews or court proceedings where the interviewee may be at significant risk of personal harm if their identity were revealed.

Self-driving Cars

By 2020, it is expected that 10 million self-driving cars will be on the road along with more than 250 million smart cars—cars connected to high-tech networks—sharing the road with them. There are already self-driving features on several vehicles that are available today thanks to Tesla, Mercedes and BMW and pilot projects from Blackberry QNX, University of Waterloo, Erwin Hymer Group, Continental Corporation, X-Matik, Magna and Uber.

We can send our gratitude for this innovation to machine learning and the complex system of sensors, cameras and software that help vehicles absorb data from their environments and learn and respond to what the data is telling them. Google's self-driving car project, Waymo, has already logged the equivalent of 300 years of driving experience on city streets since 2009 and promises to free up time, lower stress, make our roadways safer and improve transportation for us all. IBM and GM are also investing in this area. Even Uber is testing out and investing $300 million to further develop a fleet of self-driving vehicles.

*As an anchor for the Province's Autonomous Vehicle Innovation Network, the **University of Ontario Institute of Technology** has a leading role in developing, testing and validating autonomous vehicle system in North America. Included in this is the simulation of wireless environments that A/Vs will see in the real world along with the performance of Lidars, antennas and GNSS like systems in severe weather environments. UOIT has also taken a leadership role in quality assurance of software systems for A/Vs, critical for performance on the roads.*

---
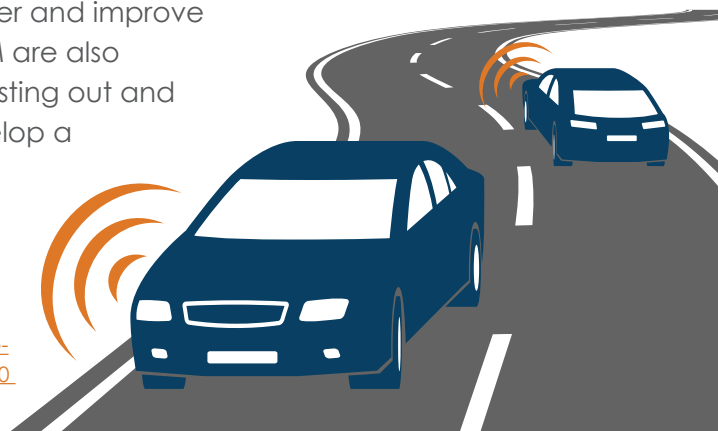
[10] https://www.cbc.ca/news/canada/british-columbia/anonymous-sources-blur-face-1.4775650

New Way to Detect Breast Cancer

A conventional approach to breast cancer diagnosis can often times have false positive results, leading to potentially unnecessary biopsies and surgeries. The most common cause of false positive results are high-risk lesions. However, with an AI system that uses a model with approximately 600 existing high-risk lesions that checks for patterns of different data elements, analyzes traditional risk factors, including family history, pathology reports, past biopsies and demographics breast cancer diagnosis predictions increased by 97% and benign surgeries decreased by almost 30%.

Spartan Bioscience

We may soon see a shift in how our health care system is administered, whereby consumers can accurately self-diagnose diseases like strep throat and STDs. Public health may be improved by accurate tests in food for toxic E. coli bacteria. Veterinarians can monitor pets for kennel cough, and hospital staff can determine which blood thinners to use for stroke patients. Brewers can even check for spoiled beer. All this and more will soon be possible using a personalized DNA testing machine called the Spartan Cube. Devices like the Spartan Cube are only now possible because of the abundance of available genetic data.[11]

Northern Biologics

Using patient biomarker data analysis, Canadian biotech spin-off Northern Biologics is developing a therapy of antibodies targeted at proteins that have been associated with cancer. Scientists hope to have a major impact on basic research in cancer biology, expand the options for cancer therapy and yield numerous commercial opportunities. Their data centric approach could enable these therapies to enter the market within a decade of their discovery, twice as fast as the typical timeline.

---

[11] https://www.theglobeandmail.com/report-on-business/spartan-bioscience-looks-to-make-dna-testing-accessible-with-portable-device/

DNA as Data—
We Need to Share

Advances in DNA sequencing have allowed researchers to accumulate vast amounts of human genomic data. In the right hands, this "big data" has many potential health applications, from increasing our knowledge of rare diseases and supporting the development of new cancer therapies to helping clinicians care for patients.

Can-SHARE launched in 2015 with $3.3 million in funding from multiple sources, including Genome Canada, CIHR, Génome Québec, Genome British Columbia and the Ontario Ministry of Research and Innovation. It aims to advance the objectives of the Global Alliance for Genomics and Health, an international coalition formed to enable the sharing of genomic and clinical data. *"Strong statistical power is often needed to demonstrate that a certain trait has a health impact,"* says Anne Marie Tassé, P3G's Executive Director. *"Data sharing allows researchers to obtain better quality results more quickly."*

An Opportunity for Canada?

In May of this year, Europe's new General Data Protection Regulations (GDPR) came into force with a number of new features, such as data portability and the right to be forgotten. The legislation has global reach—it applies to any company that collects and stores the personal information of a European citizen. The narrative in Canada recently has centered on maintaining adequacy with the GDPR. Privacy advocates have argued that the Personal Information Protection and Electronic Documents Act (PIPEDA) no longer meets this challenge, as it does not account for new technology.

Under Canada's PIPEDA, organizations are accountable for protecting the personal information under their control. They are responsible for identifying privacy-related obligations and risks and appropriately addressing them in developing their business models and related technologies and business practices and safeguards before they launch new products or services. They also need to minimize risks to their organization, their employees and their customers, as well as mitigate the effects of any privacy breaches. They do this by having an evergreen privacy management program that encompasses these considerations on an ongoing basis.

PIPEDA was once considered the global gold standard for achieving a balance between the legitimate interests of business and protecting the sanctity of personal information. It has been effective precisely because it is technology neutral and principled based. Those two distinctions offer a unique opportunity for Canada between the restrictive environment in Europe and the permissive environment in the United States. Canada could become the equivalent of the Swiss banking industry for data distribution.

Canada can achieve this goal by guarding the balance achieved in PIPEDA and by focusing resources on three key areas where Canada already has leadership: cybersecurity, homomorphic encryption and artificial intelligence.

Cybersecurity

Collecting, storing and using personal data now has significant risks. Data breaches are costly. Protecting personal data is imperative, and Canada is a current global leader in this space. In 2016, an independent report prepared by Deloitte (Harnessing the Cybersecurity Opportunity for Growth), prepared for the Toronto Financial Services Alliance and Ontario Centres of Excellence ranks Canada as the forth-largest innovation hub for cybersecurity in the world, behind the U.S., Israel and the U.K., based on venture capital deals in cybersecuritwy. The key strengths are in financial services and technology.

There are several pockets across the country where we are seeing significant investments:

- The Smart Cybersecurity Network, SERENE-RISC, based in Montreal.

- Ontario is a leader with a number of academic institutions, like the University of Ontario Institute of Technology and Algonquin College, offering advanced education in cybersecurity.

- Ottawa is becoming a hub for cyber, both for established companies like Blackberry (QNX) and for the start-up community, with a growing list of companies offering unique products in areas like multi-factor authentication and encryption technology.

- Most recently, we are seeing a significant presence in New Brunswick with the opening of Siemans Global Cybersecurity Centre of Excellence and the Canadian Institute for Cybersecurity – UNB and CyberNB.

Homomorphic Encryption

Building walls around data to keep predators out has limits. Making the data unusable without a key (encryption) is another effective way to deter miscreants. However, conventional encryption requires the data be decrypted to perform analysis on the data set. Homomorphic encryption is a form of encryption that allows computation on ciphertexts, generating an encrypted result which, when decrypted, matches the result of the operations as if they had been performed on the plaintext. The purpose of homomorphic encryption is to allow computation on encrypted data.
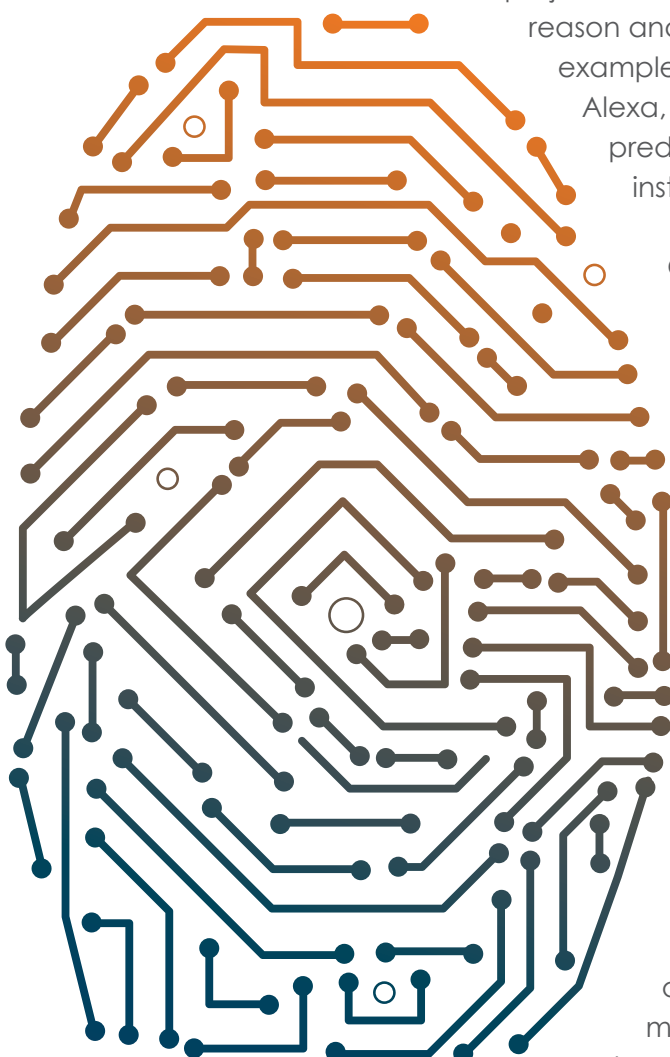
Artificial Intelligence

Artificial Intelligence (AI) is broadly defined as the ability of a computer or robot to perform tasks commonly associated with human beings. The term is frequently applied to the project of developing systems endowed with the ability to reason and learn from experience. Commonly adopted examples of AI include virtual assistants, like Siri and Alexa, video games, smart cars, Amazon's purchase prediction algorithms and analytics used by financial institutions to detect fraud.

Given the rapid growth in computational power, experts predict the next five years will bring more advancement than the past 30 years combined. The global market potential is estimated at $2.4 billion in 2017, and trillions of dollars are expected to be invested in the coming years.

By leveraging the pioneering and innovative work at three academic centres, the Montreal Institute for Learning Algorithms (MILA), the Alberta Machine Intelligence Institute (AMII) and the Vector Institute for Artificial Intelligence (VIAI), along with the work of cutting-edge start-up companies in Montreal, Toronto and Edmonton, Canada has emerged as a leader in the global race to develop artificial intelligence. This country has one of the largest AI talent pools in the world, making it a go-to-destination thanks to its ground-breaking success in this field.

The federal government has supported AI development with the 2017 Pan-Canadian Artificial Intelligence Strategy and this year's AI-Powered Supply Chains Supercluster, which is expected to generate $16.5 billion in GDP over the next 10 years.

Data: The Good

Artificial intelligence algorithms rely on vast amounts of high-quality data. The benefits are cross-cutting. Industrial processes, improving the detection and treatment of disease, energy efficiency, agricultural yield, public transit and mining are all ready areas where Canada has significant expertise and high potential.

The big-data prize lies in finding the balance between the protection of personal information and realizing the true economic benefit. Here again, Canada has an advantage with its current privacy laws that rely on an obligation to protect information no matter where it rests.

Data, and the techniques and technologies employed to collect and analyze it, will allow Canada and the world to solve some of the world's most pressing economic, social and environmental problems. Data is now the engine of economic growth and prosperity. Countries that promote data's availability and use for societal good and economic development will lead the fourth industrial revolution and give their citizens a better quality of life. To seize this opportunity, people and businesses need to be able to share their data with one another.

The right framework for Canada's data economy is one that both creates trust and incentivizes innovation. In this new world, we cannot ignore the role that companies that have not even yet been created will play, both as innovators and wealth creators. Canada's framework for the data economy must help SMEs start up, grow and become world leaders in the digital space.

Up Next?

In the second segment of this three part series, we will examine the emerging trends in technology that have put the collection, storage and use of personal information at risk. We will do this through an analysis of some of the major data breaches of the past few years that have driven the current narrative about the adequacy of our privacy laws and how business has responded.

Canadian Chamber of Commerce / Chambre de Commerce du Canada