Data Opportunities and Risks: The Dynamic of Public, Personal, and Commercial Interest

James Capotosto*

Public responsibilities such as community safety, government security, health care, education, and welfare are increasingly being looked at through the lenses of big data. Tensions between the use of data to improve these large and complex portfolios and issues of personal confidentiality persist. Yet, people freely serve up their personal information to private companies for the smallest of benefits. This commentary explores this dynamic in terms of current capabilities, opportunities, and constraints.

The sheer abundance of data from smart devices, browsing histories, and social media interactions makes it easier to understand how people interact with their neighbours, communities, and societies. People are increasingly interconnected with each other and their environments through the internet of things. Hundreds of millions of photos, texts, videos, and social media updates are sent daily, while businesses collect data on consumers' preferences and purchases (Khoso, 2016). Data are growing at breakneck speed, and by 2020 almost two megabytes of new information will be generated every second for every person on the planet (Marr, 2015).

Companies like Alphabet Inc., Apple, Amazon, Twitter, and others monetize these vast stores of information, taking every opportunity to use personal data to their advantage. Facebook's average revenue per user was $4.83USD as of December 31, 2016, which is up 29 per cent from the same period one year ago (Facebook, 2017). Last August *The Washington Post* reported that Facebook collects 98 data points on each of its 1.86 billion monthly active users (Dewey, 2016). Among these are age, education level, income, wealth, ethnicity, relationship status, family composition, employment status, religious preferences, frequent locations, and many more pieces of personal information.

Most of this information is provided voluntarily through users' profiles and posts of things like family photos. Pictures need not be posted by users themselves, as Facebook's facial recognition software sifts through every picture available to identify and build profiles on everyone featured (Halpern, 2016). Facebook knows much of its users' browsing histories because it is alerted every time users load a page with a like or share button (Dewey, 2016).

Facebook also collaborates with major data brokers to access information from government and public records, consumer contest, surveys, warranties, and private commercial sources like pharmacy records and pay stubs (Dewey, 2016; Halpern, 2016). Municipalities also sell sensitive data like voter registrations, motor vehicle details, death notices, and foreclosure declarations (Halpern, 2016). Facebook is not alone in these types of practices as many other companies, especially those that operate in the social media space, pursue similar opportunities. Several years ago Target was in the limelight for correctly inferring that one of its customers was pregnant. Using information from a customer's purchase history, Target made this deduction even before the teenager told her parents (Duhigg, 2012).

People freely provide information to these companies regardless of how the potential release of their information may affect them in the future. Even if people are aware of the potential consequences, they show little concern. Facebook's monthly active user base has grown nearly 850 per cent in the last seven years (Facebook, 2017; Facebook, 2013). Within Canada and the United States, growth was only 106 per cent for the same period; there were 231 million monthly active users as of the end of 2016. In other words, nearly 65 per cent of the combined population of these two countries chose to freely share their personal information with Facebook at least once per month. With the exception of Canada's last general election, more people use Facebook regularly than show up to vote. Voter turnout in Canada's last five general elections ranged from 59 to 68 per cent (Elections Canada, 2017).

APPLYING DATA TO POLICING AND CRIME

Data are transforming everything from retail to health care and traditional news to social media. Data are becoming more and more important for community safety and well-being. Police departments are turning to data-driven insights in an effort to boost their effectiveness and improve social outcomes. Known as "predictive policing", the practice involves analyzing data on the time, location, and types of past crimes in conjunction with other information like geography, weather, and a host of demographic information (Gibbs, 2015; Lynch, 2016). The idea is to gain insight into where and when future criminal events may take place, as well as who might be involved. Armed with this information, police officers can take proactive action before events occur.

Correspondence to: James Capotosto, Senior Economist & Policy Advisor, Ministry of Justice, Corrections and Policing, Government of Saskatchewan, 1200 – 1874 Scarth Street, Regina, SK, Canada, S4P 4B3.
E-mail: james.capotosto@gov.sk.ca

Although never available at the volume, variety, and velocity available to private sector businesses today, data have been used in the administration of justice for many years. Even the notion of predictive policing—or more simply, crime forecasting—has been around since the 1930s (Hvistendahl, 2016). Sociologist Clifford R. Shaw and criminologist Henry D. McKay introduced the concept when they explored the persistence of juvenile crime in specific neighbourhoods (Clifford & McKay, 1969). Scientists have endeavoured and experimented to better understand risk levels and the likelihood of criminal events ever since (Hvistendahl, 2016). Leveraging the potential of data for policing became far more mainstream with the introduction of Compstat, short for computer statistics, to the New York City Police in the early 1990s.

Jack Maple, then Deputy Police Commissioner for Crime Control Strategies, introduced Compstat as the first large-scale dynamic approach to crime reduction through information sharing, increased accountability, and data-driven insight (Bureau of Justice Assistance [BJA], 2013). Using comparative statistics and other information at their disposal, experienced police officers identify crime spikes and deploy resources for targeted intervention. Compstat includes four generally recognized core components: timely and accurate information or intelligence; rapid deployment of resources; effective tactics; and relentless follow-up (BJA, 2013). It is a tool that provides nearly real-time actionable intelligence for police to identify trends and prevent further crime and violence.

Technologies and applications in this space are only getting better with time and effort. Police departments are using techniques to draw intelligence from disparate datasets to reveal underlying patterns, meaning, and trends (Schlehahn, Aichroth, Mann, et al., 2016). This in turn improves our understanding of the factors that can trigger criminal events, increases situational awareness, and provides information to support more effective planning and execution of crime prevention measures. Applications of predictive policing can increase awareness of the potential prescriptive measures that could be taken to yield tangible outcomes and reduce crime. If employed, police departments would know that placing two resources at the corner of Portage and Main will reduce crime in downtown Winnipeg by three per cent.

Endeavours like PredPol, HunchLab, and the CrimeScan program are leading the way in this space. They provide crime prevention platforms geared toward improving community safety. All three, *mutatis mutandis*, combine expertise in mathematics, behavioural science, and other disciplines to better understand data and provide forward-looking recommendations as to where and when crimes could occur (PredPol, 2017; HunchLab, 2017; Hvistendahl, 2016). The idea is that, by anticipating where and when future crimes may occur, these platforms allow officers to pre-emptively deploy resources to stop crimes from occurring. PredPol uses crime type, location, and date, while HunchLab and CrimeScan also include things like weather, socioeconomic indicators, and 911 calls to produce predictions about the places and times that crimes are most likely to occur.

Current predictive techniques rely on vast amounts of data and typically fall within two categories: forecasting when and where criminal events may happen, and determining the likelihood of who will commit a crime or become a victim (Hvistendahl, 2016). Predicting crime hot spots should follow three steps: identifying patterns and correlations within past crime data; predicting when and where criminal events are likely to happen; and, deployment of resources to areas where the criminal events are likely to occur. Predicting offenders or victims also follows three phases: network analysis to analyze social networks that have been engaged in past criminal activities; using machine learning and algorithms to identify probable perpetrators or victims of violence; and, home visits where people are informed that they are considered at risk and offered some options (Hvistendahl, 2016).

Tools like those mentioned above are not as popular in Canada, but efforts are being made. The Institute for Canadian Urban Research Studies (ICURS) at Simon Fraser University has undertaken several related projects. They have used computational criminology and predictive policing to better understand crime patterns, and leverage applied mathematics, computer science, and criminology to address problems (Simon Fraser University [SFU], 2017a). ICURS developed CourBC, a database, which can help identify the type and complexity of issues likely to use considerable court resources, as well as the impact of bail/remand decisions and drug courts. ICURS crime attractors and generators project helped explain the paths offenders normally follow in both routine activities and journeys toward criminal activity (SFU, 2017a). The project provided baseline information for predicting offending patterns in order to support investigations and target prevention programs. ICURS is also doing work in social network analysis of offender networks, which enables examination of the spatial aspects of organized crime and gang related activity (SFU, 2017a). ICURS continues to develop analytic tools to facilitate policy assessment in fields like the economics of policing and the intersection of policing and mental health (SFU, 2017b)

Saskatchewan has created the Saskatchewan Police Predictive Analytics Lab (SPPAL), in a joint venture with the provincial government, the University of Saskatchewan, and the Saskatoon Police Service. SPPAL formally opened in January 2016 and is housed in a secure facility at the Saskatoon Police Service headquarters. It is the first of its kind in Canada because it is the first initiative to compile and analyze identifiable information. The intent is to inform interventions that support individual and public safety in compliance with legislation. SPPAL is exploring how historical information on missing children, collected over the last 10 years, can assist with current cases in reducing the amount of time children or youth are missing. The idea is to develop actionable intelligence. For example, if it is determined that a person usually does a, b, c, and d before something negative happens, then the police may be able to intervene at some point before d occurs.

Many of these activities are still in their infancy. Researchers and data scientists are just scratching the surface of what could be possible. It is important that they keep an open mind as they progress and assemble the right information about how people engage with police and vice versa. Taking data further and applying fields of study like behavioural economics and nudging from among others will have important implications for the future of policing.

APPLYING DATA TO COMMUNITY SAFETY AND WELL-BEING

Twenty-first century policing must move away from purely crime-focused practices to the emergence of a community safety and well-being service model. Evidence suggests that police calls for services are becoming increasingly unrelated to criminal code violations (Mazowita & Greenland, 2015). According to the Canadian Association of Chiefs of Police, roughly 70 to 80 per cent of all calls for service are related to broader social matters and issues like mental health disorders and substance abuse (Canada … Standing Committee on Public Safety and National Security, 2013). In this paradigm policing services must be delivered in collaboration with other human services like health care and welfare. Policing and the broader set of human services are facing unprecedented challenges, raising expectations and spiraling costs to the point that the current delivery models are unsustainable (Canada … Standing Committee on Public Safety and National Security, 2014). Gaining a better understanding of how Canadians are accessing and using public services will be critical to improving living standards, as well as reducing strain on taxpayers.

Data and predictive applications are needed to address the underlying challenges that bring people in to contact with police and other public services in the first place. Applications like those mentioned above present a wide array of possibilities to advance society and improve community safety and well-being. Data-driven decisions could significantly improve public policy, while ensuring the most efficient and effective use of taxpayer dollars. Unfortunately, all this potential rests on the foundations that people can use this information ethically and appropriately and this is where the crux of the problem lies.

Predictive policing and other data-driven initiatives must focus on proactively improving community safety and outcomes for everyone, including those people engaged in criminal activity. With appropriate governance and transparency, the trend toward evidence-informed policing should lead to less-biased decisions and increased fairness, resulting in better relations between police officers and the communities they serve (Lynch, 2016). This is especially important when considering movements like Black Lives Matter. Predictive models grounded in salient data improve the likelihood of delivering fair and equable justice.

Concerns regarding predictive policing and leveraging the value of data in Canada and other jurisdictions centred on the ethical use of information are warranted. Data scientists have an expression that seems particularly relevant in this context: 'garbage in; garbage out'. It refers to the idea that incorrect or poor quality data will always produce faulty output. Algorithms are only as good as the information used to create them. Those created under biased assumptions and poor quality data will perform poorly.

Data scientists must consider the potential consequences of how their algorithms and the tools they create may be used. They must also be open to feedback and iteratively improving upon their work. This cautionary note is particularly important for decision-makers to consider when making choices about people. How might suspicion of crime impact peoples' livelihoods or living standards? A prime example of the unintended consequences associated with information sharing played out in Ontario when United States Customs and Boarder Protection Officers blocked entry for Canadians with histories of mental illness after accessing police records of non-criminal mental health incidents (Bridge, 2011). In this case no one was physically harmed, but the Canadians involved were certainly affected. Faulty and improper use of analytics in law enforcement risks increasing discrimination, negative profiling, and eroding of personal rights (Schlehahn et al., 2016). Consequences could also include poor outcomes for people making credit applications or those seeking employment and educational opportunities.

CONCLUDING REMARKS

There is no question that data and the potential insights gained from good data and analyses may be used to improve outcomes in everything from retail to community safety and well-being. Realizing the potential of data clearly is the next frontier in innovation and delivering better value to the public. Yet, misuse of data can lead to poor outcomes and, in some cases, do more harm than good. Fortunately, the potential for misuse and harm can be mitigated through transparency.

Companies like Facebook and Target use information to influence decision-making and maximize annual returns for shareholders. Should the public not leverage similar information and analytic techniques to deliver better outcomes in health care, justice, education, and welfare? In the latter case the potential benefits are much higher, but the potential risks are quite similar. Are you more comfortable with organizations like Facebook using your personal information than public agencies like the police that are mandated to support the public good?

Rather than shy away, the public should embrace the full potential of data-driven decision-making to help people and communities. Insights that can benefit the public and drive the delivery of better human services should be pursued. With the appropriate checks and balances in place, in coordination with prudent governance and oversight, the tradeoff between better outcomes and privacy is worth taking. Citizens should demand openness and transparency, but ensure that privacy never trumps safety.

CONFLICT OF INTEREST DISCLOSURES
The author declares that there are no conflicts of interest.

AUTHOR AFFILIATIONS
*Ministry of Justice, Corrections and Policing, Government of Saskatchewan, Senior Economist & Policy Advisor.

REFERENCES
Bridge, S. (2011, September 9). Canadians with mental illnesses denied U.S. entry. CBCNews [online]. Retrieved from http://www.cbc.ca/news/canada/canadians-with-mental-illnesses-denied-u-s-entry-1.1034903

Bureau of Justice Assistance [BJA]. (2013). Compstat: Its orgins, evolution, and future in law enforcement agencies. Washington, DC: Bureau of Justice Assistance.

Canada. Parliament. House of Commons. Standing Committee on Public Safety and National Security. (2013). Evidence. Meeting 67, January 31. 41st Parliament, 1st Session. Available: www.parl.gc.ca/HousePublications/Publication.aspx?DocId=5950911

Canada. Parliament. House of Commons. Standing Committee on Public Safety and National Security. (2014). Economics of Policing. 1st Report. 41st

Parliament, 2nd Session. Available: www.parl.gc.ca/HousePublications/Publication.aspx?DocId=6583312

Clifford, R. S., & McKay, H. D. (1969). Juvenile delinquency and urban areas, rev. ed. Chicago, IL: University of Chicago Press.

Dewey, C. (2016, August 19). 98 personal data points that Facebook uses to target ads to you. The Washington Post. Retrieved from www.washingtonpost.com/news/the-intersect/wp/2016/08/19/98-personal-data-points-that-facebook-uses-to-target-ads-to-you

Duhigg, C. (2012, February 16). How companies learn your secrets. The New York Times Magazine. Retrieved from www.nytimes.com/2012/02/19/magazine/shopping-habits.html

Elections Canada. (2017, January 9). Voter Turnout at Federal Elections and Referendums. [Information Post]. Retrieved from www.elections.ca/content.aspx?dir=turn&document=index&lang=e&section=ele

Facebook Inc. (2017). 2016 Facebook Inc. annual report. Retrieved from https://investor.fb.com/financials/?section=secfilings

Facebook Inc. (2013). 2012 Facebook Inc. annual report. Retrieved from https://investor.fb.com/financials/?section=secfilings

Gibbs, P. (2015, April 29). The benefits of data in criminal justice: improving policing. Sunlight Foundation. Retrieved from https://sunlightfoundation.com/2015/04/29/the-benefits-of-data-in-criminal-justice-improving-policing/

Halpern, S. (2016, December 22). They have, right now, another you. The New York Review of Books. Retrieved from www.nybooks.com/articles/2016/12/22/they-have-right-now-another-you/

HunchLab. (2017). Next generation of predictive policing. [Information Post]. Philadelphia, PA: Azavea. Retrieved from www.hunchlab.com

Hvistendahl, M. (2016, September 28). Can 'predictive policing' prevent crime before it happens? Science Magazine. Retrieved from www.sciencemag.org/news/2016/09/can-predictive-policing-prevent-crime-it-happens

Khoso, M. (2016, May 13). How much data is produced every day? [Information Post]. Northeastern University. Retrieved from www.northeastern.edu/levelblog/2016/05/13/how-much-data-produced-every-day/

Lynch, C. (2016, April 24). Is predictive policing the law-enforcement tactic of the future? The Wall Street Journal. Retrieved from www.wsj.com/articles/is-predictive-policing-the-law-enforcement-tactic-of-the-future-1461550190

Marr, B. (2015, September 30). Big data: 20 mind-boggling facts everyone must read. Forbes. Retrieved from www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read

Mazowita, B., & Greenland, J. (2016). Police resources in Canada, 2015. (Catalogue No. 85-002-X). Ottawa, ON: Statistics Canada.

PredPol. (2015). About PredPol [Information Post]. Retrieved from www.predpol.com

Simon Fraser University. (2017a). Computational criminology and predictive policing [Information Post]. Retrieved from https://www.sfu.ca/icurs/research-areas/computational.html

Simon Fraser University. (2017b). Institute for Canadian Urban Research Studies: Research Areas [Information Post]. Retrieved from www.sfu.ca/icurs

Schlehahn, E., Aichroth, P., Mann, S., Schreiner, R., Lang, U., Shepherd, I. D. H., & Wong, B. L. W. (2015, September 7-9). Benefits and pitfalls of predictive policing. Paper presented at 2015 European Intelligence and Security Informatics Conference, Manchester, UK.