C-51 and Big Data Analytics: The Perils of Security by Algorithm

Micheal Vonn, BC Civil Liberties Association

Introduction – the notorious "C-51"

Bill C-51 was passed into law in 2015 and became the *Anti-Terrorism Act, 2015*. As occasionally happens with bills, "C-51" was so high profile and controversial that the initial moniker became the de facto name of the legislation and the Act is still best known as "C-51". The general consensus is that it is the most radical and sweeping change to Canada's national security law landscape since the anti-terrorism legislation introduced immediately post-9/11.

C-51 has been soundly critiqued. Commentators contend that some of the legislation is not compliant with the *Charter of Rights and Freedoms*, none of it is sound security policy and, however well-intentioned, that it presents a real and genuine danger to Canadians[1].

One constitutional challenge has been filed and more are expected. In addition, the federal government has announced that it will introduce amendments to C-51. Discussions and public debate about C-51 are still very live issues in 2016, not only with respect to the law as it currently exists, but with respect to the oversight and accountability mechanisms that are so conspicuously missing from the law.

C-51 is an extensive, intricate, omnibus act. It creates whole, new, free-standing statutes and amends many others. This paper is about the intersection of C-51 and the use of big data analytics, which brings a particular focus to two aspects of C-51: the *Security of Canada Information Sharing Act* and the *Secure Air Travel Act*, which creates the new Canadian 'no fly' regime.

Security of Canada Information Sharing Act

The *Security of Canada Information Sharing Act* ("Info Sharing Act") was a major part of the concerted public push-back on C-51. The Info Sharing Act allows 17 federal government institutions to do wholesale accessing of personal information in the name of national security. This includes all of the federal security intelligence, border and policing agencies as well as the Canada Revenue Agency, Health Canada, Veterans Affairs, and others. The *Privacy Act* also governs these federal agencies and allows information sharing, but only under specific exemptions to the obligation not to disclose personal information. In contrast, the Info Sharing Act's authorization for the dissemination of personal information is vast.

Listed agencies are authorized to access, use and disclose personal information in relation to an "activity that undermines the security of Canada". These are not activities that "*threaten* the security of Canada" (the usual language of national security provisions, but rather that "undermines" the security of Canada.

[1] Laura Payton, Anti-terrorism bill C-51 'dangerous' legislation, 100 academics say, CBC News, 3 March 2015, at: http://www.cbc.ca/news/politics/anti-terrorism-bill-c-51-dangerous-legislation-100-academics-say-1.2975233

As legal scholars Kent Roach and Craig Forcese have observed, this definition is so "wildly overbroad" as to be "unprecedented in Canadian law."[2]

In the legislation, activities that "undermine the security of Canada" include:

- interference with the capability of the Government of Canada in relation to intelligence, defence, border operations, public safety, the administration of justice, diplomatic or consular relations or the economic or financial stability of Canada;

- changing or unduly influencing a government in Canada by force or unlawful means;

- interference with critical infrastructure;

- an activity that causes serious harm to a person or their property because of that person's association with Canada; and

- an activity that takes place in Canada and undermines the security of another state.

In all, there are eight listed "activities" in the "undermining security" definition and on this list the word "terrorism" appears only once. This highlights how C-51 is in fact not "anti-terrorism" legislation, but rather "security" legislation of the most sweeping kind.

While the Info Sharing Act states that the activities that undermine the security of Canada do not include "advocacy, protest, dissent and artistic expression", the Act still authorizes surveillance of these constitutionally protected activities because its purposes are framed so broadly. The purposes are "detecting, identifying, analyzing, preventing, investigating or disrupting" an "activities that undermine the security of Canada". And while this is not supposed to include "advocacy" and "protest" per se; it can and does include advocates and protesters.

The language of C-51 does nothing to diminish the current ambit of surveillance for national security purposes which includes monitoring non-violent protests, including those of First Nations and environmental groups. The federal Government Operations Centre has called on all federal departments to compile information on all protests in the country for "common situational awareness at the national level related to all hazards of national interest, emerging or occurring."[3] The results of freedom of information requests have shown that the monitoring of potentially hazardous events is so extensive as to include student rallies, demonstrations held in support of the International Day to

---

[2] Craig Forcese and Kent Roach, Bill C-51 Backgrounder #3: Sharing Information and Lost Lessons from the Maher Arar Experience, 16 February 2015, available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2565886

[3] David Pugliese, Government orders federal departments to keep tabs on all demonstrations across the country, Ottawa Citizen, 4 June, 2014, at: http://ottawacitizen.com/news/politics/government-orders-federal-departments-to-keep-tabs-on-all-demonstrations-across-country

Eliminate Racial Discrimination, anti-poverty gatherings and monitoring "Opponents of New Speed Limit Laws"'[4].

Given that information sharing in the national security context is nothing new and has long been achieved under existing legislation, there has been an implicit question about what the Info Sharing Act is *for*. The growing consensus among privacy advocates is that C-51 is not about say, CSIS going to another federal agency to ask for information about an individual of concern. Nothing in the Act requires that level of individualized suspicion or tailoring. The language of C-51 is wide enough to effectively allow for the commandeering of entire databases, and we assume that that is exactly what it is for.

For example, CSIS could request from Passport Canada the entire biometric database of Canadian passport holders' photos formatted for facial recognition technology, arguing that the information sought is for the purposes of "detecting, identifying analyzing, preventing, investigating, or disrupting" activities that "undermine the security of Canada."

And it is not just biometric databases that would be of interest to security intelligence. Simply put, there are no databases of personal information that are not of interest to security intelligence, no matter how remote from security-related concerns.

And that is because increasingly the analysis process of choice is not about bringing together files of relevant security-related information for assessment by human analysts. Increasingly, the process compiles vast and disparate databases of no obvious relevance and not so much data mining, as data fracking – looking for correlations and relationships that are not even suspected and only dredged-up by powerful data analytics. This is the essence of 'big data'.

Watch lists provide an excellent example of the change to a 'big data' approach. The public is familiar with the notion of 'terrorist watch lists' of various kinds. Our sense of this, up until recently, is that staff of national security agencies review files and decide who is suspect/dangerous enough to be put on a list.

Watch lists are inherently problematic. Part of the bureaucratic risk logics that have always been at play in the creation of inevitably bloated and inaccurate watch lists is that there is no institutional risk to the vast over-inclusion of names. That is, the agency accountable for the list only has a problem if it *doesn't* add a person who should be on the list and that person proceeds to do something terrible. Having people on the list who should not be on the list is no problem for the agency. That is a problem for those people. Security logics are concerned with false negatives, not false positives, which is one of reasons that security logics are in constant tension with individuals' rights.

The New Canadian No-Fly Regime

[4] https://www.documentcloud.org/documents/2192551-goc-800demonstrations.html

Recently, the Public Safety Minister responded to a flurry of media reports about children having difficulties boarding airplanes. Although the media reports said the children were on the 'no fly' list, that is not technically the case. C-51 authorizes two different lists, which could be termed the No Fly List and the Slow Fly List. The children discussed in the media are on the Slow-fly List.

People on the No Fly List are denied boarding. People on the Slow Fly List may still be able to board, but at a minimum are subject to delay because of special screening.

As a threshold matter, no fly schemes are highly contested as security tools. These are watch lists of persons who are prevented from flying, on the basis that they are too dangerous to fly, but too innocent to arrest, even on charges of conspiracy. Too dangerous to fly, but not too dangerous to go anywhere else: train, ferry, subway, school, shopping mall, etc.

Commentators have long argued that if the aim of the government is to keep suspected terrorists off of planes, that that can be accomplished by use of the currently existing criminal law. But rather than evaluate the need for a no-fly scheme, C-51 has expanded the program and made it more difficult, and in some cases impossible, to effectively challenge being included on the lists.

This discussion of the 'no fly' regime very importantly circles back to the Info Sharing Act.

Pretty obviously "the Minister" who is officially responsible for putting people on these lists probably doesn't know any of the people. It's the security agencies putting the lists together. And as can be seen in the extraordinary secrecy protections that can be invoked if a listing is challenged, they are not keen to share information about their grounds and evidence for justifying a listing.

Because so little is known about how the No Fly List is compiled, the public has mostly educated guesses. But there is more known publicly about what is going into the decisions about Slow Fly, if only because the Canadian Border Services Agency (CBSA) has told us in recent media reports.

The CBSA says it does "special screening" on the basis of risk scoring, which is not described.[5] Almost assuredly this is risk scoring based on the analysis that is generated by a secret algorithm applied to as many data sets as the security agencies can get their hands on, because that is what it looks like everywhere else.[6]

This is not a new approach. The US and European authorities are doing it and the approach has been soundly critiqued. We are 'post-Snowden', so the fact that there is a global surveillance operation alliance should be news to no one. However, long before the explosive revelations of Edward Snowden, and before him, whistleblowers like William Binney and Thomas Drake, privacy commissioners and privacy advocates had been pushing back against the dissemination of PNRs – Passenger Name Records.

---

[5] Thousands flagged for scrutiny by Canada's new air passenger screening system, Truro Daily News, 14 January, 2016, at: http://www.trurodaily.com/News/Canada---World/2016-01-14/article-4403897/Thousands-flagged-for-scrutiny-by-Canadas-new-air-passenger-screening-system/1

[6] See: https://papersplease.org/wp/2014/09/22/gao-audit-confirms-tsa-shift-to-pre-crime-profiling-of-all-air-travelers/

PNRs are records created by airlines and travel agencies relating to travel bookings. The question is whether this information, generated for business purposes, belongs in the hands of security agents.

Not even addressing the metadata, PNRs can include contact information, passport information, itineraries, information on travel companions, hotel and car reservations, credit card details, dietary information, information on disabilities, etc.[7]

The question of who gets PNR and how it can be used is an old fight on the privacy front. State actors have long fought for this information for the purposes of national security surveillance. This was initially understood as being information that would be vetted against watch lists. That is, you would identify that a person on your watch list was about to enter your country on the basis of the match between your watch list and the PNR information.

But matching names of passengers with names of persons who are "wanted" or otherwise "flagged" is not the only use for PNR data for national security purposes now.

As described in a slide used in a presentation by NSA to the 2011 annual conference of the Five Eyes (the intelligence alliance of the US, Canada, the UK, Australia and New Zealand), there is a "New Collection Posture" in respect of national security intelligence data gathering.

Here are the stages that the slide outlines[8]:

> Collect it all
> Process it all
> Exploit it all
> Partner it all
> Sniff it all
> Know it all

This "new posture" is not about data matching. It is about the ever-wider demands for mass, dragnet surveillance of all available data for the purposes of linking and profiling.

We are rapidly moving from looking for known persons to "risk scoring" everyone on the basis of data analytics that are conducted on any data sets that intelligence agencies can get their hands on. This is an approach that is well advertised in US security programs with names like Total Information Awareness. There are lots of generic terms that relate to this, "intelligence-based" policing and security being a popular one.

And while "prevention" typically sounds like a good idea in all kinds of realms, detecting and disrupting "pre-crime" presents formidable concerns. These dangers have been thoroughly canvassed and

---

[7] Office of the Privacy Commissioner of Canada, Checking In: Your privacy rights at airports and border crossing at: https://www.priv.gc.ca/resource/fs-fi/02_05_d_45_e.pdf

[8] The slide is reproduced in Glenn Greewald, *No Place to Hide: Edward Snowden, the NSA and the Surveillance State*, 2014, p. 97.

explored in a June 2015 report from the Council of Europe entitled "Passenger Name Records, Dating Mining and Data Protection: the Need for Strong Safeguards" (the "Report")[9].

Predictive Analysis and "Security" by Algorithm

The Report looks at the "preventative" or "predictive" profiling of individuals occurring through data analytics applied to large, diverse data sets. The short answer to the question of what benefits are derived from this approach is: none. If the question is 'does it work?' the answer is 'no'.

> There is no serious, credible evidence that untargeted suspicionless data mining and profiling in general, or the use of PNR data in such activities, are effective in detecting [or] ("identifying") terrorists or other serious criminals.[10]

And the 'not working problem' is a situation that is not going to change. The "technology" is not going to get better and allow us to do what we can't do now. There is a problem known in statistics as the "base rate fallacy"[11]. This is the mathematically unavoidable fact that when you are looking for very rare instances in a very large data set, you always end up with either excessive numbers of "false positives" or "false negatives", or both. This is a mathematical inevitability no matter how well designed your algorithm. And every program that deals with screenings under these conditions has to address this problem.

So, for example, if you are doing medical screenings programs for age-related cancers, you have to restrict your screening program to the age demographic in which there is a sufficiently high incidence of the disease that you aren't generating huge numbers of 'false positives' with all the harms that come with that.

In breast cancer screening programs, for example, there is a highly contentious debate about what the appropriate age is to begin screening. Some of this is an ethical debate and some a political debate. But at its foundation, it is a mathematical problem.

The policy considerations will vary according to context. Screening in a medical context is clearly different from screening in a national security context. The Report comes to the conclusion that:

> ... profiles should never be used in relation to phenomena that are too rare to make their application reliable, such as trying to "identify" (real, let alone potential) terrorists from a large data set.[12]

---

[9] Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Process ing of Personal Data, Council of Europe, Passenger Name Records, data mining & data protection: the need for strong safeguards, June 2015, prepared by Douwe Korff, available at: https://www.coe.int/t/dghl/standardsetting/dataprotection/TPD_documents/T-PD%282015%2911_PNR%20draft%20report%20Douwe%20Korff%20&%20Marie%20Georges_15%2006%202015.pdf

[10] *Ibid*., 94

[11] *Ibid*., 24

[12] *Ibid*., 25

In addition to asking whether the program 'works', we must also ask what it costs. In a nutshell: it is perpetuating the creation of "suspect communities"[13] and discrimination by computer.

This takes people aback very often, because computers are "scientific" and science, it is often thought, is not biased like we fallible humans. But it does not work like that. We have science *about* these technological processes. These processes can be investigated, and they have been. What we know on the basis of those investigations is that profiles generated by algorithms very readily perpetuate and reinforce social inequality and discrimination against 'out-groups'.[14]

Additionally, systems that are operating in this sphere are not static, they are so-called "dynamic" systems. That means that they create feedback loops that continually modify the underlying algorithms. As a result, it is not possible to investigate or interrogate the 'blueprint' for the analysis that is used to generate profiles. In a dynamic system, the 'blueprint', that is, the criteria that are used in the assessment, cannot be effectively extracted. As the Report says, "the analyses are based on underlying code that cannot be properly understood by many who rely on them, or even expressed in plain language".[15]

A dynamic system cannot be reverse engineered to effectively "explain" its own functioning for the purposes of assessing the relevance of the criteria being used. And not only does the complexity of the algorithm resist understanding, but dynamic systems are particularly susceptible to reinforcing, by amplifying, engrained biases in the program.[16] In essence, for all intents, the algorithm is "unaccountable". And accountability is at the heart of our due process rights.

You challenge an administrative decision by being provided the reasons for the decision and being able to address the criteria, evidence and logic that was used by the decision maker. This is essentially impossible when the decision is made on the basis that "the computer said so".

And it simply doesn't do to say that humans are part of the ultimate decision-making – like the Minister being "responsible" for listing individuals on the no-fly list. If it is algorithm-based decisions that lead to the human sign-off, it is still a decision based on essentially unknowable calculations.

This is not good in any scenario, but the almost insurmountable barriers to challenging the algorithm are compounded in the national security context by the inevitable institutional demands for near-total secrecy.

If profiling in this way creates rights violations that are without effective remedy or redress, involving progresses that are neither transparent nor accountable, not only is the situation fundamentally unfair, but unfair in a way that undermines the Rule of Law.

---

[13] *Ibid*., 86
[14] *Ibid*., 27
[15] *Ibid*., 28
[16] *Ibid*., 28

C-51 is contentious for many reasons, but the intersection of C-51 and ‘Big Data,’ and the perils of “security”-by-algorithm may prove to be one of its most troubling aspects.