HOUSE OF COMMONS
CHAMBRE DES COMMUNES
CANADA

Standing Committee on Access to Information, Privacy and Ethics

ETHI • NUMBER 099 • 1st SESSION • 42nd PARLIAMENT

EVIDENCE

Tuesday, April 17, 2018

Chair

Mr. Bob Zimmer

Standing Committee on Access to Information, Privacy and Ethics

Tuesday, April 17, 2018

● (0845)

[*English*]

The Chair (Mr. Bob Zimmer (Prince George—Peace River—Northern Rockies, CPC)): Welcome back, everybody.

We will call to order the Standing Committee on Access to Information, Privacy and Ethics. This is meeting 99. Pursuant to Standing Order 108(2), we are studying the breach of personal information involving Cambridge Analytica and Facebook.

Today we have some witnesses via teleconference and in person.

In person, we have Daniel Therrien, Privacy Commissioner of Canada, and Barbara Bucknell, Director of Policy and Research.

Via teleconference, we have Chris Vickery, Director of Cyber-Risk Research at UpGuard.

Welcome to all.

Mr. Therrien, you have the floor

Mr. Daniel Therrien (Privacy Commissioner of Canada, Office of the Privacy Commissioner of Canada): Good morning.

[*Translation*]

I would like to thank the committee for the invitation today to discuss the privacy implications of online platforms and appropriate legislative responses to the concerns of citizens about how their personal information is being used.

As you are aware, I received a complaint about this matter and announced some weeks ago that my office is conducting a formal investigation into how personal information on Canadians has been impacted by the activities of Facebook and Aggregate IQ.

Due to my confidentiality obligations under the law, I'm not in a position to discuss the details of this investigation with you today. I cannot prejudge our findings.

What I can share with you, however, is some perspective on the wider context that may assist you as you begin your study.

[*English*]

Canadians want to enjoy the many benefits of the digital economy, but they rightly expect they can do so without fear that their rights will be violated and their personal information will be used against them. They want to trust that rules, legislation, and government will protect them from harm.

In the recent Facebook matter, what happened, as acknowledged by CEO Mark Zuckerberg, was, quote, a “major breach of trust”. As recognized by the CEO of another giant tech company, Tim Cook of Apple, the situation is so dire that it is now time to develop well-crafted legislation to regulate the digital economy. The time of self-regulation is over.

In Canada, we of course have privacy legislation, but it is quite permissive and gives companies wide latitude to use personal information for their own benefit. Under PIPEDA, organizations have a legal obligation to be accountable, but Canadians cannot rely exclusively on companies to manage their information responsibly. Transparency and accountability are necessary, but they are not sufficient.

To be clear, it is not enough to simply ask companies to live up to their responsibilities. Canadians need stronger privacy laws that will protect them when organizations fail to do so. This was a major conclusion of my annual report to Parliament last year, and a point I made during your recent study of PIPEDA, Canada's private sector privacy law.

Significantly, given the opaqueness of business models and complexity of data flows, the law should allow my office to go into an organization to independently confirm that the principles in our privacy laws are being respected—without necessarily suspecting a violation of the law.

The time has also come to provide my office with the power to make orders and issue financial penalties, helping us to more effectively deal with those who refuse to comply with the law.

Strengthened legislation does not need to be an impediment to innovation. We know that personal information plays a key role in the digital economy, including advances in the field of artificial intelligence, which are necessary for Canada's social and economic development. We need legislation that ensures, as a general rule, that Canadians provide meaningful, informed consent for the collection and use of their personal information. But consent will not always be possible in the world of big data and artificial intelligence, where personal information may be used for multiple purposes not always known when it is collected.

This is why we recommended that Parliament examine exceptions to consent. We believe such exceptions, subject to conditions that would offer other forms of privacy protection, are preferable to relying on an interpretation of consent that is so broad as to become meaningless. We prefer narrower, specific exceptions, but we recognize that one option could be a European-style legitimate interest exception.

I'm of course very pleased that your committee recently issued a report calling for comprehensive changes to the federal private sector privacy law, which included several recommendations I had made but also others that would significantly improve the privacy rights of Canadians. Your report has shown that you are attuned to the issues stemming from the dated state of federal privacy laws in Canada, and you have actively called upon the government to make comprehensive changes.

● (0850)

Many in society, particularly in the last few weeks, are making similar calls. Even leaders of the tech industry now see the need for enhanced regulations.

If there was ever a time for action, I think, frankly, this is it.

[*Translation*]

Another area ripe for action concerns privacy protections and political parties.

As you are aware, no federal privacy law applies to political parties; British Columbia is the only province with legislation that covers them.

This is not the case in many other jurisdictions. The UK, much of the EU and New Zealand all cover political organizations with their laws.

In point of fact, in many EU states, information about political views and membership is considered highly sensitive, even within existing data protection regimes, requiring additional protections.

There are also now—in the digital environment—so many more actors involved: data brokers, analytics firms, social networks, content providers, digital marketers, telecom firms and so forth.

So while I am currently investigating commercial organizations such as Facebook and Aggregate IQ, I am unable to investigate how political parties use the personal information they may receive from corporate actors.

In my view, this is a significant gap.

Some independent authority needs to have the ability to review the practices of political parties and to assess whether privacy rights are being truly respected by all relevant players.

This gap requires addressing in one statutory form or another, either in privacy laws, in the Canada Elections Act or in a specific statute.

In conclusion, I would again highlight the urgency to act, as well as the stakes involved.

The integrity of our democratic processes—as well as trust in our digital economy—are both clearly facing significant risks.

I cannot think of more relevant questions for legislators to confront, and I applaud you for doing so.

Thank you again for your invitation, and I would welcome your questions.

Thank you.

[*English*]

The Chair: Thank you, Mr. Therrien.

We'll go to Chris Vickery, who is in sunny California today.

Mr. Vickery.

Mr. Chris Vickery (Director of Cyber Risk Research, UpGuard, As an Individual): Good morning.

It is a pleasure to be appearing before you. I am grateful for the opportunity. I believe the matter before us is one of very great importance. Facebook is certainly one of the core elements involved, but I would urge all of you to keep an eye towards the very focused efforts of others who rely on Facebook as a pillar of their operations but not solely on Facebook; others who are tending to cause direct harm to what I believe is the institution of democracy itself as sort of an end goal of what they're working towards here.

In case you don't know anything about me, I am somewhat uniquely situated to speak on the topic. The majority of my work can be described as hunting down data breaches. I openly call myself a “data breach hunter”. Over the last several years, my reputation has grown to be one of a leading authority on the prevalence and causes of data breaches as well as common patterns of incident response by the affected entities. Please note, though, that the data breaches that I locate and secure are not the result of actual computer exploitation or malicious acts. This is just data that has been left out in the open for whatever reason, and nobody realized it until I came along and found it. You may think there probably wouldn't be that much of that, but you'd be surprised. There is quite an epidemic of misconfigurations out on the Internet.

Some examples of data that I've secured stem from Verizon; Viacom; Microsoft; Hewlett-Packard; the United States Department of Defense; the Mexican national institute of elections, the INE; a couple of international terrorism blacklists; as well as the 2016 Trump presidential campaign website. They were leaking a bit of information as well.

The sum total of the efforts I've undertaken has resulted in the safeguarding of nearly two billion records containing private information, so I am well versed in this stuff. I look forward to answering any questions you may have.

More on point, I would like to point out that two data breaches that I came across in December 2015 involved the United States voter registration in its entirety, all 50 states plus DC. The second time, in the December that I found it, they were more enhanced. They had private details about people, with various pieces of personality and behavioural things—whether or not somebody was a gun owner, whether or not they lived a biblical lifestyle.

Six months later, in 2016, I came across another nationwide U.S. voter registration database, this one even more enhanced, having details on whether or not somebody watched NASCAR, whether or not they were anti-abortion sentiment holders, or whether or not they likely owned a gun.

Then another set of nationwide records came to my attention. I downloaded them after finding them in June 2017. This would be the third round of complete U.S. voter registration records that I came across. This was 198 million records, ranking as the largest U.S. voter data breach in known history. I would like to point out that at the time of the discovery, not a single one of these database breaches were protected with even a username or a password. They were simply out in the open. If you knew where to look, anyone in the entire world could find them.

The AggregateIQ situation that brings me here today is one that first started on March 20 of this year—not that long ago. I didn't know who AggregateIQ was until March 20. I was fiddling around on an open public website called GitHub where the developers collaborate and publish open source code.

● (0855)

I saw a reference to @aggregateiq.com in relation to some SCL Group code that was out there and just available to the public. I followed the bread crumbs, figured out what AggregateIQ was, and noticed they had a sub-domain called GitLab. When I viewed gitlab.aggregateiq.com, it occurred to me that the registration was available, and they were in essence inviting the entire world to register for an account on their collaboration portal.

I proceeded to register an account, it let me in, and all of these tools, utilities, credentials, scripts, employee notes and issues, and merge requests were all present before me. I very quickly realized the importance of this and that there would be likely heavy interest from regulators, governments, and the populace of several nations, so I began downloading. Normally, I go to great efforts to protect anybody who may be affected by this type of thing, but the overwhelming public interest in knowing the truth behind what Cambridge Analytica, AggregateIQ, and SCL Group have been doing is a compelling factor in this particular situation. I don't want you to think I just run out there and hand out everyone's dirty laundry when these things are found. This is a different situation.

Again, keep in mind that anyone in the entire world with an Internet connection could have found the same thing, gotten an account the same way I did, and downloaded the exact same things, regardless of what nation they were in or what loyalties they might feel. This was completely exposed with no manner of protection whatsoever. A malicious actor could have taken it a step further in that there were, and are, database passwords, usernames, credentials, keys, and authentication methods documented in these files that I did not take advantage of. I did download them, but I did not go the extra step and use those passwords to access the additional databases.

If it were found by someone else, and they were of the persuasion that would take advantage of it, it could have been, and may be, a much more serious data breach than has been mentioned. They could be completely infiltrated. Every bit of data that has ever crossed through AggregateIQ's hands could be in the hands of anyone who had found this same exposure.

There are a few remaining questions that I have not been able to decipher fully that I believe your investigation should figure out. While I am still looking into quite a bit of the data, I have not come to the exact final conclusion as to what AggregateIQ's relationship is to SCL Group and Cambridge Analytica. The walls of the separation between those entities are very porous. It's clear that code access permissions and data have traversed between the three of them, and other groups, so I would implore you to get to the bottom of that.

The second question is to what extent, if any, restricted political and private data has been utilized by AggregateIQ or AggregateIQ employees for commercial profit-seeking ventures. I have found evidence of ad networks being developed under the same domain, one notably called Ad*Reach network—and there are a few Ad*Reach networks on the Internet, so make sure you're looking at the right one before going after anybody in a questioning manner—as well as aq-reach. One of the employees who was working at AIQ was doing simultaneous work for an ad company called easyAd Group AG, which is based in Switzerland and has subsidiaries in the U.S. and in Russia. I would love to know what work was being done and if any of the data travelling through AIQ was utilized in any of those ad campaigns or set-ups that the employee was working on at the same time.

● (0900)

The Chair: We're at 10 minutes, so if you could wind up your testimony, that would be great.

Mr. Chris Vickery: Yes. I have one final point.

There is also a cryptocurrency token aspect to this. Exactly one comment within the GitLab commentary section was marked with flag that later I noticed was a confidential flag. That comment had to do with the Midas token. I looked into it. The Midas token was a project they were working on, and it was tagged to a website selling cryptocurrency at a $10,000 minimum buy-in.

The website has gone down since this was made public, and it feels very fishy to me. If you could figure out why somebody was developing a cryptocurrency on the AggregateIQ GitLab instance, for sale to the public, and why they would possibly not want anyone to know about this, I think it would be worth the investigation.

Thank you. I look forward to answering any questions.

● (0905)

The Chair: Thank you, Mr. Vickery.

Just for the committee's knowledge, California is about three hours behind us, so he's up at 5 o'clock in the morning.

Thanks again for appearing.

We'll start with Nathaniel Erskine-Smith.

Mr. Nathaniel Erskine-Smith (Beaches—East York, Lib.): Thank you very much.

There are a lot of moving parts to what we've heard. My first question is just to clarify.

Based on everything you've reviewed—obviously, you haven't been able to review everything, just given the sheer volume of the information you've been able to access—it's your view that information that was collected across a number of different campaigns for specifically political and more public purposes has been clearly used for commercial profit-seeking ventures.

Mr. Chris Vickery: It's highly likely that this has occurred. I have the tools. I don't have the ingredients that those tools mixed with, because that would have involved taking the additional step of going into databases and such. From what I see, there's no reason to have these tools in this way, and the documentation as it is, if you are not going to mix the political data for commercial reasons.

Mr. Nathaniel Erskine-Smith: Okay.

Can you give us an example for those of us who are less experienced? You mentioned gun ownership. You mentioned living a biblical life and a few other examples. What's the most personal information you found?

Mr. Chris Vickery: The most personal information specific to the voter data breaches or just generally?

Mr. Nathaniel Erskine-Smith: When you're talking about using different databases to combine a profile for an individual, how detailed a profile are we talking about?

Mr. Chris Vickery: The most detailed message you have sent to a loved one through any chat app could very easily be logged, archived, tied to your name.

Mr. Nathaniel Erskine-Smith: You have come across examples like that?

Mr. Chris Vickery: Yes, but let me clarify. There is a separate Facebook-related incident that has not been reported at all yet—I'm working with a journalist right now to bring it to everyone's attention—that is not involved with Cambridge Analytica, as far as I know, but the number is 48 million people on that one. It does involve messages. The degree of privateness they were sent...is not quite determined yet, but they do get pretty personal.

Mr. Nathaniel Erskine-Smith: You said that a number of different databases are being amalgamated, I guess, in some ways, to create these profiles. Can you give us the significance of these databases? Presumably, some of these are from the election databases you talked about. Are there other examples you can give us?

Mr. Chris Vickery: Yes. In their documentation, Aggregate IQ go into detail about their system. It starts with being bootstrapped by the RNC's Data Trust data vault, which is the Republican National Committee here in the United States. I had actually found the Data Trust database before it was part of the find in June 2017. It's quite extensive. It contains data as they merged with i360, which is a Koch brothers-backed political information company. Data Trust deleted a blog entry where they claim to have merged their data with i360.

There's also L2 Political. They provided data to this whole beast of a machine. That was admitted to on Cambridge Analytica's website recently.

Facebook is obviously part of it. The documentation by AggregateIQ goes on to explain that commercial databases are involved. I know that Experian is one that contributed data toward the RNC Deep Root Analytics data briefs that I found in 2017. I know that because there were Experian IDs being lined up to each voter ID with all the consumer habits being tied onto everybody.

AggregateIQ also states that candidates can bring in their own sources of volunteer and supporter and donator information. They'll aggregate all that into the main “database of truth”, as they call it. State voter files then corroborate what the RNC has on file.

So there's really no end to what they can plug into here.

● (0910)

Mr. Nathaniel Erskine-Smith: You mentioned there are porous walls between some of these entities, like AIQ, Cambridge Analytica, and others. When you access the AIQ information, can you give us an example of what that porous relationship looks like? What are some key examples where you see players across different companies accessing the same information?

Mr. Chris Vickery: Well, one example that is very appropriate, because it illustrates both the original discovery and the whole nature of this relationship, is an employee named Ali Yassine. I usually try not to name people, but I feel that it's important for you to know this for the purposes of looking into it. He was a full stack developer for SCL Group. On his public GitHub page, he had code that came out of AggregateIQ. I know this because I found it within AggregateIQ's code base, and it was marked as being authored by an AggregateIQ employee named Koji. So you have SCL and AggregateIQ that supposedly have no relationship but both working with the same code base. Then, further on down in the code base, there is a field that says “client”, and written in there is “Cambridge Analytica”. Now, I can't see why SCL Group would be saying that Cambridge Analytica is a client of theirs. They basically own Cambridge Analytica. SCL Group is the mother ship on top of that. The only reasonable explanation to me is that AggregateIQ would have been the one putting Cambridge Analytica as the client, then the code being passed to SCL Group, and that just not being changed immediately. There's a little triangle going on there.

I can also tell you that the GitLab logs very clearly show that with the Ripon project, which was primarily developed for Ted Cruz's 2016 campaign, the very initial seeds of it were downloaded from the domain scl.ripon.us, placed in the GitLab, and developed and evolved from there. Scl.ripon.us is a domain underneath Alexander Nix's name. He's the one who is registered under the WHOIS records. That's another example of code flowing from one to the other.

Also, there are examples that Cambridge Analytica has put forward, through their public statements, of data that they used. More recently, I guess they felt pressure to be transparent about where the data came from. They admitted that they got the RNC Data Trust data. The RNC IDs are all over the place in the fields, categories, targeting scripts, and parsers that are present in AggregateIQ's repository as well as in their documentation. So if data [*Technical difficulty—Editor*] directly from one to the other, they are certainly dealing with the same type of data.

The Chair: Thank you, Mr. Erskine-Smith. We'll have another round. We have two hours.

We'll go on to seven minutes for Mr. Kent.

Hon. Peter Kent (Thornhill, CPC): Thank you, Chair.

Thank you, Commissioner Therrien and Mr. Vickery, for being with us today.

Commissioner, I know you can't discuss the specifics of your formal investigation into Facebook and AIQ in the Canadian context, but I wonder whether you could share with us your anticipated timeline for completion of the investigation and an eventual report.

● (0915)

Mr. Daniel Therrien: It's difficult to say. There are many factors at play here. Under the law, we have one year to complete our investigation. We'll obviously try to do so before then.

When you look at the allegations made, you see a fairly complex web of interactions between a number of players. These we will need to clarify. That could take a bit of time. We're also working in concert with other commissioners or data protection authorities. Of course, we're doing so with the Province of British Columbia, with which we're jointly doing this investigation, but we're also in contact with others, including in the U.K. but not limited to the U.K. There is, then, a bit of coordination.

What I'm saying is that this is somewhat complex, which may add to the time, but we have certainly at the most a goal of doing this within a year, and we'll try to do that before then.

Hon. Peter Kent: Thank you.

Now, your emphasized remarks again today calling for amendments to the Privacy Act to cover political parties' use of personal information carries significant new weight, given the information before us and the public regarding attempts, and perhaps some successes, to interfere with the democratic process in the recent U.S. election and in the Brexit vote in the United Kingdom. Certainly we have questions waiting for Mr. Wylie regarding his employment by the Liberal Party of Canada under two leaders between 2007 and 2009, his termination for what was described by one of the leaders as invasive elements of the work he was doing or was proposing be used, and then his re-employment by the Liberal research group after the 2015 election—in 2016—and payment of $100,000. Those are questions for another day.

But your request is that political parties be brought under legislation and regulated either under the Privacy Act or under the Elections Act of Canada. Which would you suggest would take priority?

Mr. Daniel Therrien: I would say probably both, actually. The situation currently is that most federal political parties have privacy policies—internal codes of conduct, so to speak, in their relationship with the people with whom they interact and from whom they collect information. That's a start.

I think, first of all, the substance of these policies could be improved, from what we have seen. One common element missing from the privacy policies of federal parties is the right of individual electors to have access to the information that parties have about them. That's a huge flaw. There is, then, the issue of the substance. But these are voluntary codes, and no one independent of the parties examines whether the parties actually live up to the promise they're making in these policies. That leads me to a very important reason that political parties should be governed by legislation: to ensure that whatever substantive rules exist, hopefully better than what they are now, are verified by an independent third party.

Should that independent third party be the Privacy Commissioner, the Chief Electoral Officer, a third person? That can be discussed, but I think that what this leads to—leads me to, at least—is that there are at least two types of issues at play here. There's the issue of privacy and whether parties treat the personal information of individuals properly, which is a privacy issue that would make me, perhaps, the best person to look at the question. Then, the allegations that we have been seeing in the past few weeks lead to a mix of the use of personal information and privacy on one hand and political purposes on the other, which is more the domain of the Chief Electoral Officer. Ideally, I would say, the two institutions would be able to verify what is happening so that the expertise of each is put in common.

● (0920)

Hon. Peter Kent: I assume you watched Mr. Zuckerberg's testimony in Washington last week.

Mr. Daniel Therrien: We read about it.

Hon. Peter Kent: I wonder whether I could ask you for your impression of what he said. Were any of your concerns eased or heightened by his responses?

Mr. Daniel Therrien: Not particularly. There have been many media reports. We're investigating Facebook, so again, what I'm saying relates to media reports and facts other than those I'm investigating. But I think it's fair to say that the public record is clear that Facebook has made many promises over the years to its users to rectify this or that, to put them in control of their personal information. This has been done year after year for a number of years, and Facebook is not the only company that acts this way.

Hon. Peter Kent: No, no, I realize that.

Mr. Daniel Therrien: This leads me to accountability. Responsibility on the part of companies is necessary, but it is not sufficient. There needs to be an independent person to look at whether they're truly accountable.

Hon. Peter Kent: Thank you.

Let me ask a quick question to you, Mr. Vickery, and I certainly have more as we go through these two hours. Facebook has been maintaining, perhaps for reasons of liability, that this was not in fact a breach but simply a user abuse of the service conditions. We know about the breach associated with the Equifax scandal, for example, but would you consider this a breach of another sort?

Mr. Chris Vickery: I have been asked that, and my answer would be, yes, I would. However, I need to explain that in my work, I draw a line of classification between a malicious breach and a non-malicious breach. I would classify this as not necessarily a malicious breach, but it was a violation of the expected way in which this data would be handled in that it was gathered under the guise of academic research not to be utilized for commercial or other purposes—and clearly it was. It did cross that boundary. Facebook asked that it be deleted, etc. We all know that tale.

I would call it a data breach, but only [*Technical difficulty—Editor*] the difference between a hack and a different type of data breach.

The Chair: Thank you, Mr. Kent.

Next up, for seven minutes, is Mr. Angus.

Mr. Charlie Angus (Timmins—James Bay, NDP): Thank you, gentlemen. This has been very, very eye-opening.

I'd like to start with you, Mr. Therrien. In 2008, CIPPIC, the Canadian Internet Policy and Public Interest Clinic, launched its complaint, with your predecessor, on Facebook. At that time, they identified the issue of third party applications as a threat to privacy.

In the world of 2008, there was much a feeling, and I was very much in that world, of a deregulated Internet—you know, let them build—and Facebook was a fun place to meet former people from high school. Ten years later, it has morphed into the primary source of news—false news, real news—and has become the major, dominant player in many of the elections around the world.

Looking at the European data protection supervisor who says that the result of Facebook's dominant control are growing political extremism and isolation and political points of view, I want to ask you, looking back on that 2008 review, about those third party applications. Would it have made a difference if the Privacy Commissioner had come down harder? Did you have the tools at that time to address those breaches? And now, in light of what we're seeing with Cambridge Analytica, do we need really much stronger tools to be able to address these issues?

Mr. Daniel Therrien: I should start by saying that one of the issues we'll be looking at this time is whether Facebook in 2018 continues to respect the conditions to which they agreed to in 2008 and 2009. In 2008-09 my office, of course under another commissioner, was satisfied that Facebook had done certain things to comply with recommendations that the OPC had made. You're right to say that there is some similarity in the issues between the investigation then and the current set of facts and that wording with the use of information by third party applications. All of this is to say that we'll look at that question again.

Would stronger powers have made a difference? At the time, and still today, all that the OPC can do is make recommendations, not order anything. To be able to order certain conduct would certainly have helped. Would it have prevented this? Perhaps not.

I think the combination of a number of measures, with clearer rules around consent—clearly, the rules around consent are extremely unclear, which I've addressed in my report and you've addressed as a committee in your report—is part of the solution. Another important part of the solution is that the regulator, the OPC, be able to inspect the activities of companies proactively, not only when a complaint is made, to ensure that they are truly accountable. To wait until complaints are filed means that the problem needs to have been identified by an individual. Because of the opaqueness of the system, that will be rare. That's why I'm saying that part of the solution is also the authority to inspect without grounds, so that we can verify, and order-making and fines would have made a difference. Would it prevent everything? No.

● (0925)

Mr. Charlie Angus: No. But I guess my concern is that given what we've seen with the allegations coming out of Myanmar, and allegations or concerns being raised in Iceland about Facebook's election day app identifying people to go vote, it has a huge impact on the voting system. We have a provincial election in Ontario. My Facebook feed is full of ads that I can tell are not from any political party, but somebody's putting them out there.

Whether or not the frame of privacy is enough, Facebook seems to see itself as beyond jurisdiction. Do we need to have a larger, more robust form of legislation that involves perhaps the electoral commission, perhaps media standards in terms of the proliferation of fake news? The concern about privacy here is about the ability to target individuals and to then feed them fake information. This is the allegation that came out of Brexit and Nigeria: that those people can significantly move voters through their friend circuits.

You don't have the powers to handle all of those. How do we as a nation bring massive data monopolies to the table to be accountable?

Mr. Daniel Therrien: You're right that there are many issues, or I would call them regulatory areas, to cover. Whether they should be covered in one piece of legislation or in several I leave to you, but I certainly agree that the big tech giants play on a number of consequences regulated under many laws. You have identified privacy and elections, and I think also perhaps monopolies: are they a utility offering a public service as opposed to a company that gathers information to give a service and in the meantime makes certain profits? That's another question.

All of these questions are relevant. They are all relevant, they all need to be addressed. I'm not sure mechanically how it works, but I think generally they should all be looked at. The regulators should be able to talk to one another, because these issues intersect.

Mr. Charlie Angus: Would you suggest that this is something our committee needs to look at in terms of how we actually ensure that we have legislation that defends the democratic integrity and rights of citizens in the face of digital monopolies? Is that something that you think our committee could take up to provide recommendations on?

● (0930)

Mr. Daniel Therrien: Sure. Yes.

Mr. Charlie Angus: Thank you.

The Chair: Thank you, Mr. Angus.

Next up for seven minutes is Mr. Saini.

Mr. Raj Saini (Kitchener Centre, Lib.): Good morning, Mr. Therrien and Mr. Vickery. Thank you very much for being here, especially you, Mr. Vickery. I'm not really a morning person, so I can only imagine 6:30 a.m. in California.

Mr. Therrien, I want to start with you, please. There have been differing reports as to the number of Facebook profiles that were affected. Mr. Wylie has said 50 million. Facebook has said 87 million. Can you tell us how many Facebook accounts were affected by this breach in Canada?

Mr. Daniel Therrien: We are relying at this point on the data that Facebook is providing to us. Facebook explained the variation in numbers essentially by saying that they themselves don't know precisely. The 87 million number I think is a function of how many people used the application, the quiz, that was at the origin of this, according to the allegations.

Mr. Raj Saini: So that quiz you're talking about....

Sorry.

Mr. Daniel Therrien: Then they make assumptions as to how many friends this leads to. There is a chain at play here. So even Facebook does not know precisely how many people have been affected.

Mr. Raj Saini: According to the reports, the company that did the survey had downloaded or had access to 270,000 people who filled out the survey, and it looks like the average was 322 or 332 friends to come up with this number of 87 million. So you're relying on Facebook, but you're not 100% guaranteed of how many Canadian profiles were actually affected.

Mr. Daniel Therrien: Not at this point. We'll of course try to nail the number down in terms of the number of Canadians during our investigation.

Mr. Raj Saini: Facebook has stated that they would be sending out notifications to the people in Canada who were affected by this breach. Do you know whether all Canadians have been informed, yes or no?

Mr. Daniel Therrien: I don't know that. It has begun, but has it been completed...? It has begun.

Mr. Raj Saini: Okay.

Now, you've also stated in the media that you will be joining the investigation in B.C. with the privacy commissioner's office there. We know that in the last few weeks, the U.K. information office in London, England, has raided the offices of Cambridge Analytica, so a lot of information will be retained or discovered through that process. Do you have any ability to work with the U.K. information office to make sure that with regard to the information that has been discovered, you would have access to that to help in your own investigation in Canada?

Mr. Daniel Therrien: We actually have robust authority to share information with other data protection authorities, either in Canada or internationally, in the conduct of investigations. What we're missing, as we all know, at the end of the investigation is the ability to make orders and impose fines when needed, but the powers to actually compel the production of evidence and to share information with other privacy regulators is adequate.

Mr. Raj Saini: Do I have some time, Mr. Zimmer? Okay.

Mr. Daniel Therrien: On the point raised by Mr. Angus, though, who says there are many regulatory areas, including perhaps competition, there are holes there. I can share with the U.K. privacy office, but I cannot share with the Canadian Competition Bureau.

Mr. Raj Saini: Okay. That's a good point.

Mr. Vickery, I have a lot of questions, but unfortunately I have limited time. Let me just start off with one.

There are stories about Cambridge Analytica retaining AggregateIQ in order to get around certain British laws. Can you comment on this, and perhaps comment generally on some of the outsourcing work that may have been done to get around the laws of particular countries, especially when you talk about foreign campaigns? Kenya is an example also.

Mr. Chris Vickery: What I can say on this is that there were not invoices, receipts, or things of that nature included in the GitLab repository that I downloaded. Exact smoking-gun receipts and pieces of paper saying “we paid this much to this person”—that's not present. However, I can clarify my current understanding that is based upon everything that I have read and looked into and believe. A good example of money flowing between Cambridge Analytica and AggregateIQ is the development of the Ripon platform that was developed during the early time of Ted Cruz's 2016 presidential campaign, where Ted Cruz's campaign believed they were paying Cambridge Analytica for this product, development, or whatever service, but in actuality it was AggregateIQ that was doing the developing, creating the product, and basically being the workhorse on it, while the cheques were going to Cambridge Analytica.

Does that give a good example of the flow there?

● (0935)

Mr. Raj Saini: You mentioned cryptocurrency. Could cryptocurrency be used as the way to hide payment between entities who may be using the information?

Mr. Chris Vickery: The possibility is there. I want to emphasize that I have no current reason to believe that any money laundering has occurred, but I think it is worth looking into.

Mr. Raj Saini: Okay.

I know that UpGuard produced a report on deciphering AggregateIQ's activities. I tried to read the report, but unfortunately, I don't know code. Can you kind of summarize the report for us?

Mr. Chris Vickery: Which one? We've put out four now.

Mr. Raj Saini: I think it's the fourth part.

Mr. Chris Vickery: That's the most recent one regarding Canadian politics and such. That report confirmed the names that are present. It doesn't necessarily mean that because a candidate's name appears in a project that AggregateIQ was working on.... It doesn't mean that candidate was necessarily doing anything nefarious or that anything unlawful occurred. They could have been hooked up with AggregateIQ through no malicious intent whatsoever; somebody just suggested them or whatever. But we did see projects with names, including Todd Stone, Andy Wells, and Doug Clovechok. The B.C. Greens had some folders in there. I believe a lot of this has been sussed out by the Canadian media already. There should be several articles explaining the various levels of involvement, if our reporting didn't give you a good enough view.

The Chair: Thank you, Mr. Saini.

Next up, for five minutes, we have Monsieur Gourde.

[*Translation*]

Mr. Jacques Gourde (Lévis—Lotbinière, CPC): Thank you, Mr. Chair.

My question is for you, Mr. Therrien.

About two weeks ago, in an interview that you gave to a national French-language media outlet, I heard you say that a gray area surrounds the information collected by the political parties. I would like some clarification on that.

All politicians and political parties receive the voters' list, which includes each citizen's first and last name, full address, permanent voter number, and polling station location. Everyone has access to that, not only the political parties, but also the candidates running in a constituency, whether they are independent candidates or not. However, to the great dismay of all those politicians, the list has no phone numbers.

When we were all younger, it was relatively easy to find a phone number using a phone book, because 80% of subscribers to a fixed telephone network were listed in it. When we wanted to call someone, we just had to look up their name in the directory. Then we could add their phone numbers to the voters' list.

Mr. Therrien, are Canadians' telephone numbers now considered information covered by privacy? Should they not be accessible to political parties, or is this an example of a gray area? We have fixed telephone networks and we also have cellular networks. However, cellphone numbers are becoming more difficult to find. A phone number in a fixed network is public, but a cellphone number is not.

Mr. Daniel Therrien: There is obviously nothing wrong with political parties wanting to communicate with voters. However, to answer your specific question about whether the parties should have access to telephone numbers or other personal information, I would say that the notion of consent should come into play, given the principles of privacy. If an individual's phone number is not public and that person does not want to disclose it to anyone, including a political party, it should be possible for them to keep that number confidential.

● (0940)

Mr. Jacques Gourde: So there is a distinction to be made between the telephone numbers in a fixed network, which can be found in any telephone directory, and the ones in a cellular network.

Can we assume that, when a telephone number is listed in a telephone directory, it means that it has been authorized in advance?

Mr. Daniel Therrien: It's public information.

Mr. Jacques Gourde: So a cellphone number is not considered public information?

Mr. Daniel Therrien: Some are public and others are not; it depends on an individual's decision whether or not to give consent.

Mr. Jacques Gourde: To obtain someone's consent, you have to phone them. If we do not have a phone number to reach that person, what can we do? Should we go and see them?

Mr. Daniel Therrien: Um, yes.

Mr. Jacques Gourde: You will understand that that is a great challenge: we have to reach 85,000 voters in the 40 days of an election campaign, not to mention the fact that people are not there year round. That's about 2,500 doors a day to knock on, which is physically impossible, even with a team of 15 people. You really have to take the time to talk to people.

The basic problem for people wishing to engage in politics is that they cannot access a minimum of information about voters. So we use technological means to get an idea of their allegiance. We do not hide it: if we want their phone number, it is so that we can call them, even if we can always go see them. At the end of the day, the information that politicians want is whether they can count on their support. If the person clearly says that they support a certain candidate, the candidate will keep the information, and then make sure that the person goes to the polls on election day. Political party lists get longer over the years, and they can still be used if, of course, the information is up to date. But there is a margin of error, nevertheless.

If Canadians decide to get into politics and do not have access to a minimum amount of information, can we blame them for using tactics that will save them time and show them voting patterns as early as possible?

Mr. Daniel Therrien: I would actually tell you to tackle the problem of access to a minimum of information. As I understand it, you are saying that the voters' list does not allow you to obtain the minimum amount of information you need to be able at the very least to communicate with someone and to check whether they are going to support you or not. Revisiting what constitutes a minimum of information seems to be the solution to me, not finding other ways to communicate with people.

At the end of the day, we have the concept of consent, but I fully understand that the desire of parties to want to communicate with voters is extremely legitimate and that it may be necessary for them to obtain a minimum of information in order to do so.

Mr. Jacques Gourde: Let me go back to the gray area. Would it not be fairer for the legislation to allow the numbers to be distributed to all political parties and to all those running for election in the ridings? We could perhaps take the phone numbers out of the gray area and make the information accessible to everyone.

For example, someone running as an independent candidate, who has never done the research, can use the list of first and last names and addresses, but they will never have the time to find the phone numbers. So they are at a real disadvantage compared to all those who have been representing political parties for 25, 30 or 40 years.

Do you not think it would be fairer to at least give the same basic information to everyone so that everyone is on the same footing at the beginning of an election campaign?

Mr. Daniel Therrien: That seems to be the case. I am not an expert on the subject, but I understand that all parties have some information that comes from the voters' list.

You say that the information does not give you the minimum amount of information you need to communicate with electors. So that is a valid question that can be studied. You could look into the matter.

On the other hand, when I talked about a gray area, it was in the sense that, since I do not have the jurisdiction to verify how the parties use the information, I do not know what is going on. Parties have privacy policies in place to ensure a minimum of rules in their dealings with voters. However, neither I nor any other independent person can verify what is happening. So that's what I meant by “gray area”, an area with no independent arbitrator who can ensure that the rules in place are followed.

[*English*]

The Chair: Thank you, Mr. Gourde.

Next up is Mr. Baylis for five minutes.

Mr. Frank Baylis (Pierrefonds—Dollard, Lib.): Thank you, Mr. Therrien and Mr. Vickery, for being here.

It seems that we have a fundamental question to ask as a society, and that is, when we come to data, what will we allow and what will we not allow? It really falls on the government to make the rules and not allow each company to decide how and when they use data in whatever manner.

I'm going to put that question to both of you, in order to understand. Before I do so, I'll say that we've always had targeted marketing. I was just looking up Michael Dell from Dell Computers. Before he became a computer mogul, he used to sell newspapers. He would look at the database of people who were newly married or people who had just moved. He was extremely successful as a teenager doing that. I could get that data now from, say, Facebook. If I wanted to sell newspapers the old-fashioned way, tell me who has just moved and who just married. We've allowed targeted marketing before.

Now, selling of data—nothing to do with Facebook, again; I give charitable donations, and I know that some of these charities share my data with other charities, because it's a good way to hit someone up again. Sometimes they ask permission to share the data, and sometimes they don't. That sharing of data for commercial reasons, that targeting, has been allowed. Both of these things have been allowed in the past; Facebook makes it far more efficient. If I were a political party, let's say the Green Party, I'd say that whoever's posting a lot about environmental issues might be a good person for me to target to get a donation or to convert.

I want to ask you this fundamental question. What should we allow, knowing that these things have already happened, and what should we not allow? How should we as a government put parameters around this behaviour?

I'll start off with you, Mr. Therrien, and then we'll go to Mr. Vickery.

● (0945)

Mr. Daniel Therrien: I don't think there's a short answer to that question, but if there were one, I would say that we have tried our darndest to come up with a reasonable answer in the study we've made in the consent report and the recommendations and measures we are taking in relation to it. You augmented that significantly, I think, with your report as a committee in February. So it's not one thing. It's a series of things.

Mr. Frank Baylis: On a philosophical level, for example, should we allow...? For example, I'm asking if we should say, “Look, I'm giving you this data, but you may only use it for this. I'm making a donation. I'm in your database. I do not allow you to share it.” Or I might allow you to.

On a philosophical level, do I own what I give? Do I not?

Mr. Daniel Therrien: On that level, I think the answer is that the personal information of individuals is something they need to be able to be in control of. You put it in terms of ownership, and that is something that is sometimes said. I would rather say that it is a human right, that privacy is a human right, to control your privacy, and therefore what information you allow to be known by others, and to what end, is because you choose to do that, because you think it provides a benefit for you, as opposed to giving consent for an extremely broad purpose, which is then open season for others to interpret as they see fit.

Mr. Frank Baylis: So you would come down to the formulation of consent, from a philosophical point of view, as I give you this data, but I put parameters around what you can and cannot do with that data.

Mr. Daniel Therrien: Yes, with legal rules—that's the role of government—to ensure that this philosophical concept is actually respected.

Mr. Frank Baylis: Yes, we will build the technicalities. Once we decide what we want to do, then we can drill down and make, for example, your right to investigate and your right to fine, etc.

I agree with you that you apply it even to political parties. That's the technicality, once we decide what we want to do.

Mr. Vickery, how would you see this? You're in this world constantly.

Mr. Chris Vickery: I believe the incentives to spread around data by people who are profiting from it are great. Giving in a little bit is not only a slippery slope; it is a foregone conclusion that it will happen to a wide degree, and it's just a matter of time before political parties and commercial list builders and consumer surveillance groups all come together and offer each other large sums of money for the data.

A suggestion I would have for looking into a potential kind of compromise would be to decide that everybody has a right to own their own data. If you want to give a charity the right or permission to share your data with another charity of similar mind, I don't think it's unreasonable to expect that the first charity you gave the information to should send you an email saying they're planning to share your information with such and such a group and ask whether that is okay—"Opt in here to share it"—or at least send you a notification that they are sharing it. Nothing then is done in the darkness; nothing is done under the table; everything is known, and there's a paper trail and there is consent.

● (0950)

Mr. Frank Baylis: Thank you.

The Chair: Next up, for another five minutes, is Mr. Kent.

Hon. Peter Kent: Thank you, Chair.

Thank you, Commissioner, for noting this committee's unanimous report and recommendations to the government in February. We hope that the government has consumed it as you did.

One recommendation in that report, one that you have made in a variety of rather tangential ways, is to work with the European Union privacy regulators. In just a couple of weeks the new EU GDPR, the general data protection regulation, comes into effect. It protects virtually every data element of citizens across the EU, from their basic information—social insurance number, in the Canadian context—to all of their social media activity, all of their personal information, the computers they own, their telephone numbers, and so forth.

Has this Facebook scandal, the Cambridge Analytica scandal, AIQ, all of the things we're talking about today, and the fact that artificial intelligence, which has generated magnificent benefits to society, to mankind, while at the same time there's been a rush to develop new programs without any consideration for protections and precautions...? Is it time for Canada to consider something like the GDPR regulations to protect privacy, from the most minimum basic level up to the most complicated, when it gets to algorithms and stereotyping and exploitation?

Mr. Daniel Therrien: It is more than time that Canada legislates. I have made that point many times. The GDPR, the European regulation, is certainly a good standard to compare ourselves with, but I think it's important for each country to develop its own legislation. There might be cultural or constitutional reasons that certain rules would be different, but certainly the European model is a good model. I've made a number of recommendations inspired by that model.

The main point is that it is high time—it is past time—to legislate.

Hon. Peter Kent: Thank you.

Mr. Vickery, you touched on this in a number of your previous answers, but is there a technical way to ensure, if social media users don't opt in, that the multiplication or pluralistic development of their individual data through their friends, their lists, their contacts, their friends' faces, can be prevented? Or does this come down to a matter of trust that the social media companies that users place their trust in will be true to whatever commitments they may or may not make now or are regulated to make in the future?

Mr. Chris Vickery: I have both a positive answer to that and a negative answer to that.

I'll start with the negative. There is no way to guarantee that any bit of data, any string of characters you submit or that are identified with you, will not be propagated down the line to another company. Data multiplies. I see it all the time. There is just no way to prevent it. It's too prolific.

I do have a suggestion on how to work towards the goal of containing the amount of data that is multiplying out there for whatever purpose. That suggestion is to have on the books laws that have teeth. These companies will not deal with large databases of information if they know that it is a huge liability and a potential threat to their bottom line. It's not until the regulators can issue fines that affect profits and stock value that these companies will respect what the regulations say.

● (0955)

Hon. Peter Kent: So this would be along the lines of, for example, the GDPR, which has provisions for up to 20 million euros in penalties, in fines, or the equivalent of 4% of the revenues of that particular company? Some of these companies are multi-billion dollar revenue generators.

Mr. Chris Vickery: I can't speak specifically to the numbers and calculations, but I believe that is in the same vein as what I'm talking about, that, yes, it takes something with teeth attached to it to really get executives' attention. GDPR has gotten a lot of executives' attention.

Hon. Peter Kent: Thank you.

The Chair: Thank you, Mr. Kent.

Next up is Madame Fortier.

[*Translation*]

Mrs. Mona Fortier (Ottawa—Vanier, Lib.): Thank you very much.

Gentlemen, thank you for being here today.

Mr. Therrien, you have become a regular. It's like you're a favourite on *Tout le monde en parle* and can come whenever you like. I will begin with you, because I really want to understand the exercise we are doing now and, most importantly, the one you are doing on your side.

As you know, the committee also unanimously decided to investigate the apparent breach of Facebook data by Cambridge Analytica, but without compromising your own investigation.

I am curious to know how you characterize the breach of privacy in this case. If I have understood the comments you have made recently, it is your belief that the regulations in force have left too much leeway for Facebook in collecting personal data and that this has created the right conditions for Cambridge Analytica to use that information in an illegal or unethical manner.

Could you characterize the breach of privacy that you are currently studying?

Mr. Daniel Therrien: Because of the ongoing investigation and our legal obligations, the most important of which is not to draw any conclusions before completing this investigation, I would like to qualify your remarks slightly.

The conclusions you attribute to me would be more a function of what we generally see, as representatives of a regulatory agency, with the behaviour of all companies and the legislation that applies to them. Every day, we see that privacy policies are very permissive in that they allow for a very broad use of information, which is not always consistent with informed consent.

Can we say that Facebook violated privacy based on the facts alleged? We will certainly be looking into it. Our investigation is ongoing and we cannot draw conclusions yet. I can tell you what issues we will be looking into, but we are not going to talk about any conclusions in this case.

In general, we will be asking ourselves whether the two companies we are investigating, Facebook and Aggregate IQ, have violated the federal privacy legislation and, in the case of British Columbia, the provincial legislation.

More specifically, we will be examining whether Facebook's privacy policies actually were too permissive and whether they played a role in the subsequent use of the information by analytics firms to give advice that may or may not have been useful to political parties, among other things.

We will also be trying to determine, as I said earlier, whether the recommendations made by the Office before I arrived in 2009 are still applicable in 2018.

Finally, we will be looking at the role played by Aggregate IQ in all this and how the company collected the information. Was it done in accordance with the legislation? We will mainly consider the type of data analysis that was done. Did the final product, as communicated to the political parties, comply with privacy protection laws?

All those questions are relevant, and we will examine them. Obviously, I cannot draw any conclusions right now.

● (1000)

Mrs. Mona Fortier: I understand, thank you very much.

You are conducting your investigation on your side, but this committee will be receiving Facebook representatives later this week. In your opinion, are there any particular questions we should ask them? Do you have any suggestions for the committee?

Mr. Daniel Therrien: Yes.

Factually, how does Facebook ensure that a third party, the people conducting the research, obtains the personal information of its users in a manner consistent with the consent given by the users and with the privacy requirements?

In addition, how does Facebook protect the data of its users against anyone who might want to use them for inappropriate or unauthorized purposes? I am thinking here of malicious hackers, the so-called bad hackers.

Finally, last week, Mr. Zuckerberg said that the time has come for Facebook to have appropriate regulations. So what does this mean for Facebook, especially in terms of our recommendations—in the Office of the Privacy Commissioner of Canada—of this committee's recommendations, and the European regulations?

Mrs. Mona Fortier: Thank you very much, Mr. Therrien.

[*English*]

The Chair: Thank you, Madame Fortier.

Next up, for three minutes, is Mr. Angus.

Mr. Charlie Angus: Thank you.

Mr. Vickery, we started out dealing with a breach of 85 million Facebook accounts that may have upended the most important election in Europe in this generation. Then you come this morning and just casually mention that 48 million other people may have had their information breached, including very personal information.

I know this is something you're probably still investigating, but was this a Facebook breach?

Mr. Chris Vickery: No, this is something entirely separate, as far as I am currently aware.

Part of the work I do tries to highlight the prevalence of this type of data breach. It happens a lot more often than people realize. It is something I come across all the time. It is very hard to surprise me these days. People don't seem to have any grasp of how often these very large spills of information are occurring.

Mr. Charlie Angus: I'll be following up in my next round with much more focus on the issue of the database breach and AIQ's role in it. But you said that the database was open, that you could just go there. I figure that if you're looking for this stuff, other people are looking for this stuff. I mean, we have Russian troll armies, we have cyber-threats, we have criminal gangs. Was that database open to exploitation from others? You mentioned the danger of other players. Would you elaborate on the potential danger of other players getting access to that data?

Mr. Chris Vickery: Yes. Every breach I come across—then see that it is secured after I've come across it and then eventually talk about it in public reports—is completely open to anyone and everyone with an Internet connection. There's no username, password, or other protection involved.

To get at what I believe you're asking about, the answer is, yes, if I'm coming across all of these and I'm one guy—now working with a team these days, but relatively one guy coming across all this—it would be extremely surprising if adversarial nations were not devoting large sums of resources to do the same for malicious intent.

Mr. Charlie Angus: We're going to get to this in my next round, but it's not just that the information that came from the app that was used from Facebook was taken from Facebook users; it is the ability of political operatives such as AIQ or Cambridge Analytica to use Facebook—the platform—to then either distort news or influence voters.

Can you talk about how the use of Facebook is not just the taking of the information but the ability to plant information?

● (1005)

Mr. Chris Vickery: Yes. As a starter to answer that, I want to make it clear that Facebook app usage and potential exploitation, going over into the grey area of what you can do with it, is a prolific problem. I discovered over the weekend that one of the Facebook apps tied to AggregateIQ—it actually has their name on it as a scraper—was classified under the games category of Facebook apps. I don't believe it's one that anybody has mentioned yet. There are probably many—

Mr. Charlie Angus: Is that app still usable now?

Mr. Chris Vickery: They have been suspended from the Facebook platform, so I believe it is no longer functional, but the identifier for the app still exists within the code that I found.

That's the first part of your question. Can you remind me what the second part of the question was?

The Chair: We're at four minutes now.

Mr. Charlie Angus: Four minutes?

The Chair: Yes, believe it or not.

Mr. Charlie Angus: Really? I think you're cheating me, Mr. Chair.

Voices: Oh, oh!

The Chair: It just goes by fast.

Mr. Chris Vickery: I'm sorry. I'll try to be more concise.

The Chair: That's okay. It's valuable testimony.

We'll start a whole new round. We'll try to grab 10 minutes for committee business at the end of this, so I'm going to try to keep you a little tighter this time.

We'll start off with Ms. Vandenbeld for seven minutes.

Ms. Anita Vandenbeld (Ottawa West—Nepean, Lib.): Thank you.

I'd like to thank both of you for being here. There's a lot of information here. What we really want to get down to is what the remedies are, what things we can do as legislators, because I think a lot of this is very alarming to our constituents and to Canadians.

I want to be sure that I'm understanding exactly what the problem is. You mentioned, Mr. Vickery, that data multiplies, so you can't actually prevent it, but you can contain it, and that these spills are happening all the time. This to me is a very bad combination. On the one hand, you have the issue of legitimate use of data. Let's say a political party is going door to door, and they run into somebody who says,“I really like your child care platform. I'm going to vote for you because of that.” They make a note of that so that the next time they do something on child care, they can let them know. Even if the person gives consent and says, “Yes, please keep me updated about that”, you've got that. Then that goes into a database. The issue to me is not so much whether the candidate goes back to the person and says, “Hey, look at this great policy we have”, but whether it is then shared, either accidentally or maliciously, with, say, Toys“R”Us, who says, “Ah, they're concerned about child care, therefore let's sell them toys.”

Is that where we're looking? Is that the problem, or is it vice versa, that Toys“R”Us might be accessing somehow this political data...or you're looking at what Toys“R”Us is selling to kids, or somebody on Facebook who shows they have kids, and inferring it? So it's the cross-purposes of data: is that where the issue resides?

Mr. Chris Vickery: It's happening both ways. Political data is making its way into commercial ventures and marketing, and the consumer behaviour data gathered by Toys“R”Us or whoever is making its way into political campaigning purposes.

I don't believe there's much of a problem with going door to door and knocking on somebody's door and gathering that data. You can't scale that out exponentially, because you're limited by time and space.

I don't know what gun laws are like in Canada, but in America we have the concept that you have certain guns that are okay to own, but we don't allow civilians to own machine guns. It's the same type of thing. You can have knocking on doors and gathering the phone number of one person at a time or whatever, but when that turns into more of a machine gun situation, whereby you are sending out thousands of surveys and emails and Facebook advertisements and everything and harvesting en masse the private details—or personal details, at least—of many, many thousands of times the people you could normally reach, that gets into the machine gun category, and that is dangerous.

Ms. Anita Vandenbeld: What worries me is what you were talking about; I think you said there are spills happening all the time. Even if you wanted to maintain it in a small, local campaign—just things that you're keeping or that a commercial entity is keeping track of, purchases of people in their store, or for advertising—when this is happening accidentally....

We're not talking about legislation that says you must notify when you share the information with X, Y, or Z, because many of these entities won't even know that they're sharing this information. It's out there in a place where then somebody else can access it. Really, it seems to me the problem is at the aggregate level, at which you have other entities seeking information, combining information, and then selling that information. Really, that's where we need to be focused, on the selling of that large mass of information that's perhaps collected from multiple different sources.

● (1010)

Mr. Chris Vickery: That's true to a degree, but I think it's also worth taking a whack at the source of the data as well. Many of the leaks that occur are happening because companies are willingly ignorant of their security posturing. They have no monitoring going on. They know that what they're doing is profitable, so they don't want to look for problems. The problems exist, and they are being taken advantage of, and companies don't want to know about it. There's no incentive to find a data breach, because then it's just your problem.

We need to incentivize looking for the problems and punish those who are not willing to up their game.

Ms. Anita Vandenbeld: Just going on what you said about the cryptocurrency, you have another whole level when you talk about maybe underground organizations, or nation-states, or criminals who might be trying to do this in other jurisdictions where we can't legislate, where they may be collecting this data and using it for very nefarious purposes. We almost can't legislate that.

Are there ways to be able to prevent that, or is it really, as you said, going back to where the data is collected and housed and making sure that there are incentives for security at that level?

Mr. Chris Vickery: The good news is that for this data to be optimally useful for the malicious purposes, the bad guys have a need for it to be continually updated and accurate, just as for legitimate usage. If we can stop the flow of accurate, up-to-date information, in time what they already have will become irrelevant, to a large degree.

Ms. Anita Vandenbeld: As well, in terms of the solution here being that we need to make sure companies are taking it more seriously in terms of privacy, is that part of it? The other question I have is about what you mean by a legitimate interest exception. That was the other piece you mentioned that I wasn't quite clear on.

Mr. Daniel Therrien: I would agree with what Mr. Vickery has said as to the sum of the solutions. Given the mandate you have, perhaps you want to look at what might be some legitimate uses of the information to communicate legitimately with electors. I agree that there's a concern both ways—information collected for political purposes being used commercially, or vice versa—and that needs to be looked at, but what I'm perhaps adding on the table is that this flow of information is certainly worth looking at, but it may not all be inappropriate. If you take your example of the family that buys toys and you say that political parties need to communicate with electors, to convince them properly, knowing who the electors are, is it necessarily a bad thing that the commercial habits of the family are part of what is assessed?

I'm not an expert in elections and in what goes against the integrity of an election or does not, but I'm looking at it conceptually. As there is a need for parties to communicate with electors intelligently, knowing who the electors are, some of the data analysis may be okay, but certainly not all of it, and the allegations in the case of Facebook and Cambridge Analytica certainly suggest an inappropriate use of information for political purposes. I'm just saying that there might be some legitimate uses.

As for legitimate interests, we're not in the world now of Facebook and Cambridge Analytica. We're more in the world in which, if the privacy laws are strengthened, there is a legitimate concern that the rules, or some would say restrictions, should not inhibit legitimate, responsible innovation. In answer to Mr. Baylis, I said that the value at stake for the most part is consent—control by individuals over their personal information. In the modern world, however, information may be used for several purposes, and it may not always be possible to inform the holder of that data of all the purposes to which the information will be put. The information is properly put to use in certain artificial intelligence initiatives, for instance.

Part of the challenge is to have strong rules that generally ensure that consent is respected, but in the world of big data and artificial intelligence, it may be that there's a need for an exception to consent. The Europeans use this exception of legitimate business interest as a way to ground lawful processing of data without consent. I think a balanced piece of legislation would enhance consent, on one hand, but also needs to consider what we do as a country with proper business or social concerns—it may be in the health sector—that need to have information without necessarily the consent of the individual, for a true benefit for society.

● (1015)

The Chair: Thank you, Mr. Therrien.

Next up for seven minutes is Mr. Gourde.

[*Translation*]

Mr. Jacques Gourde: Thank you, Mr. Chair.

My question is for Mr. Vickery.

In the last U.S. presidential election, which was relatively tight, the candidate who won the popular vote lost, while the candidate who, in some minds, should have come second in the voting managed to win by collecting a majority in the electoral college, perhaps through more targeted advertising.

Last week, the founder of Facebook explained that his company's raison d'être, its business model, is to sell advertising. And Facebook does it very well, being particularly able to target regions, even streets or buildings: if someone lives in building X, they will receive advertising Y.

As an example, I own a Mazda, and, as if by chance, Facebook sends me a Mazda advertisement every day on my Facebook feed. So we see that Facebook targets ads in an extremely effective way. It is likely that American political parties use Facebook to advertise in certain sectors, states, or parts of states where voters are more likely to be supportive and therefore to vote for them.

Do you think that American political parties, both Democrats and Republicans, have done any electoral profiling or used the services of companies that have analyzed the best way to target advertisements or influence Americans in certain states? Would it be possible to conclude that the person or party who was most effective in his Facebook advertising campaign won the U.S. election?

[*English*]

Mr. Chris Vickery: I want to be clear that I'm not here as a proponent of either party, but I can state for a fact that every voter data breach that I have found regarding U.S. politics, and every system of influence and so forth that I have looked at just recently in this past find, has been a Republican-based operation. I have not come across, to my knowledge, a Democratic system of selectively micro-targeting and influencing. Although it may exist, I have not come across it, so I can't speak to the Democratic side. I think it's clear that, yes, at the very least, the Republican side did, to great effect, utilize micro-targeting, compilations of disparate and wide-ranging databases, all-encompassing databases, from places you might not even expect, and, yes, brought it all together in a very effective way that sought out people who were influential and influenceable, and targeted those types of people with messaging to get a desired outcome.

● (1020)

[*Translation*]

Mr. Jacques Gourde: Thank you.

Mr. Therrien, it seems that Cambridge Analytica had access to data from 650,000 Canadian Facebook users. If those Canadians had been undecided voters, could they have been influenced and, given our Canadian electoral system, could that have determined the outcome of a relatively tight election?

Mr. Daniel Therrien: I think it's about 620,000 users. I am not an expert on electoral matters, but that number is obviously significant. So I think the answer to your question is yes.

Mr. Jacques Gourde: In your investigation, will you be able to determine whether these 620,000 users were scattered across Canada or whether certain ridings were targeted? This is because 620,000 people targeted in 90 ridings have a lot more influence than if they are spread across 338 ridings. Maybe we could discover something that nobody has seen yet.

Mr. Daniel Therrien: I will certainly take note of that question.

Our point of departure is more the matter of purpose. Our investigation will focus on the use of Facebook user information—a network that essentially exists to communicate with friends—for analytical purposes in support of political goals.

You are suggesting that we push our work to a level of detail that would probably not be necessary for our purposes, but that could be useful. We will consider it, but I feel that the matter would be more the responsibility of Mr. Perrault at Elections Canada.

Mr. Jacques Gourde: If Facebook offered a political party the opportunity to advertise on Facebook and, a week before the election, it provided a list of 620,000 Canadians who had watched the advertisement about its leader, it would not mean that the vote of those Canadians would be assured or that they intended to vote. On the other hand, if, coincidentally, a second advertisement from the party was broadcast encouraging people to vote according to the values of the party, and it appeared four or five times a day on their Facebook pages in the week before the election, could it have an effect, especially if those people had said that they had seen the previous advertising? That would be more than profiling. That would be inappropriate.

Mr. Daniel Therrien: It can certainly have consequences on the election. So I encourage you to consider those questions. As you describe this situation to me, my main thought is that Facebook users provided the company with data mainly in order to communicate with a certain number of people, certainly not in order to receive advertising on the eve of the election urging them to vote for one reason or another.

In terms of the principles of privacy, in the scenario you are describing to us, the interpretation of consent seems to have been excessive. As to whether it would have electoral consequences, I would say that would probably be the case, even if these issues are not in my area.

[*English*]

The Chair: You are out of time, Mr. Gourde. It goes by fast.

Next up, for seven minutes, is Mr. Angus.

Mr. Charlie Angus: Mr. Vickery, let's try to play the game of follow the data. We have SCL, we have GSR, we have Cambridge Analytica. Now, GSR sets up this Facebook app for scientific research, gathers around 86 million profiles, and then for a pittance —probably the price of two Cokes—sells all that data to Cambridge Analytica. Facebook becomes aware of it and asks them to delete the information, and they say they do. Then this company, AggregateIQ, which is a completely unknown company from Victoria that doesn't even have its own website, suddenly gets 40% of the Brexit leave budget to run the leave campaign.

Would you be able to tell us if that Facebook information that was scraped is what's in the AggregateIQ's database?

● (1025)

Mr. Chris Vickery: I'd like to dispel some misunderstandings that may exist and further this along. The Facebook data that was scraped through Facebook apps as well as surveys that were conducted over Mechanical Turk, which is an Amazon offering that people can do—there were a lot of different ways that GSR was gathering data and tying it all back into Facebook profiles—wouldn't necessarily be needed anymore after the modelling, the analysis, the behavioural tools had been developed using that data. Once you understand the interactions and the way to make people do certain things with certain messaging, the raw data from Facebook can be purged. You don't need it anymore. You can then take that framework and use it with more election-based voting data that you start to build up to get the desired outcomes, because you've already used those frameworks on the social media data.

So no, I have not come across what is obviously the Facebook data in question within this repository. That's not to say it never existed, but yes, that's the—

Mr. Charlie Angus: You said the information could also have come from Amazon?

Mr. Chris Vickery: I want to be clear that Amazon's Mechanical Turk system is a way to have people fill out surveys and pay them for it. That is one of the methods that GSR employed as they were gathering this type of data. They tied it back into the Facebook profile, but Amazon's Mechanical Turk was a vector used.

Mr. Charlie Angus: Okay. So what is the connection, then? Cambridge Analytica and AggregateIQ claim to be completely separate. Again, AggregateIQ gets 40% of the leave budget. Christopher Wylie says they were basically used as an electioneering money-laundering scheme through Cambridge Analytica, as a franchise. What in the database connects the two?

Mr. Chris Vickery: One of the earliest connections between Cambridge Analytica and Aggregate IQ is the fact, reflected in the code and the commentary from the employees, that they got the original software for the Ripon platform from a server owned under the name of Alexander Nix, who was the recently departed CEO over there at Cambridge Analytica SCL Group. So there is that direct tie-in.

One of the apps that was developed by AggregateIQ is a phone, community outreach, and voter-influencing messaging platform. It operated on a domain known as dclisten.com, which is also registered under the name of Alexander Nix. There are plenty of examples of resources and assets flowing between these two groups.

Mr. Charlie Angus: For Alexander Nix, who is with SCL and on the board of Cambridge, it's his website that AggregateIQ has listed as the only website they had when they were given 40% of the Brexit Vote Leave. Wasn't it an SCL-AggregateIQ website?

Mr. Chris Vickery: The only website they had up at the time...? I don't know the history of that website now. I just don't know.

Mr. Charlie Angus: Okay.

There's been talk about AggregateIQ's involvement in many elections, and there have been some very disturbing allegations from Christopher Wylie about a culture of illegality. One of those allegations was the illegal collection of raw user data from ISPs in Trinidad and Tobago. Were you able to confirm any of that?

Mr. Chris Vickery: The Trinidad and Tobago allegations—

Mr. Charlie Angus: Yes, of ISPs. Did you not mention that on Twitter?

Mr. Chris Vickery: I believe the evidence to show that those allegations are factual exists here. It's not in a final conclusion phase, but yes, there is a Trinidad and Tobago project in here with personally identifiable information of a great deal of people.

● (1030)

Mr. Charlie Angus: You can say that AIQ, that database, had collected raw data from ISPs in Trinidad and Tobago?

Mr. Chris Vickery: Whether they got it from ISPs or other means —

Mr. Charlie Angus: Other means, yes.

Mr. Chris Vickery: —is not definitive, but yes, there is personally identifiable—

Mr. Charlie Angus: Okay. Let me rephrase that. There was data that could have affected the elections in Trinidad and Tobago that was in the AIQ database.

Mr. Chris Vickery: Yes, absolutely.

Mr. Charlie Angus: Okay. Thank you.

I'm concerned about the cryptocurrency issue you mentioned and these ad networks they were setting up, because one of the things that Christopher Wylie stated was that the real money wasn't in elections. The real money was after.... He talked about the ability to influence governments if you've got the right government in. With AggregateIQ, does it look like this database is being used for other commercial purposes that would have furthered their interests?

Mr. Chris Vickery: The potential is there.

I want to emphasize that I see AggregateIQ more as a division of a larger entity. You could think of them being more like a development department within a larger corporation.

Mr. Charlie Angus: Who is that larger corporation?

Mr. Chris Vickery: I would say that corporation is likely to be thought of as SCL. I see their goals and end points aligning in parallel.

Mr. Charlie Angus: Thank you.

The Chair: Thank you, Mr. Angus.

We'll go to Monsieur Picard for seven minutes.

[*Translation*]

Mr. Michel Picard (Montarville, Lib.): I would like to come back to the theme of our study, that is, the information that we describe as personal and what we do with it. The different possible scenarios aside, I believe that the use of this information by a company is only an ancillary dimension of the essential problem that we have to study.

I have two questions, which I will illustrate with two scenarios. Based on those scenarios, I would like your comments on my understanding of the problem.

My questions are as follows. Is the government's role to define in detail what constitutes personal information? Or would the role of the government be to ban any transaction that contains this personal information?

Here are my two scenarios.

In the first, I do business with a book supplier: Amazon, as it happens. I find it normal and expected that, when I purchase my first book or on a subsequent visit, Amazon will suggest a number of other books based on the preferences of other readers, or buyers, or just simply based on my own history of buying books from Amazon. In establishing my relationship with the company, I provided it with a certain amount of personal information, so that it can provide me with a service based on its expertise in this area.

Here is my other scenario. I am naive enough to announce that, in a month, I will be going on a cruise for a week. It would not be surprising if a user who reads my Facebook feed and works in a travel agency contacts me to let me know about some cruise-related deals. Nor should I be surprised at the risk of my house being broken into during the one-week absence I announced. Both the criminal and the travel agent used my personal information, but I was the one who made it public. This is personal information that I shared on Facebook with my friends and followers, which is the service that the social network offers. So I made that information public.

Let me go back to my questions. Both scenarios describe realistic situations. Who is responsible for defining the granularity of personal information? Each type of company requires different categories of information. In addition, to the extent that a transaction depends on the expertise of the company—such as Amazon—of which I am a customer, I do not expect that company to sell my personal information to another company for purposes, including commercial solicitation, other than those established in my relationship with Amazon, that is, buying books.

Which role do you think is better, or should we consider a mix of both?

Perhaps Mr. Therrien could answer first.

Mr. Daniel Therrien: I'm going to expand on your questions. If I misinterpret them, please tell me.

Basically, individuals give certain information in order to get a service. One of the consequences is that information is communicated at the time the service is provided. In the case of Amazon, for example, the company uses the information of people who are like you or who share your interests, that is, people who have liked a particular book. I would say that too is personal information to an extent, within the meaning of the definition of the term.

The conclusion that Amazon takes from your interests, for example, that you like detective novels, is actually the result of your personal information, but that conclusion itself becomes your personal information too: your actual or potential interest in detective novels is personal information about you.

The role of the state is to define what personal information is. In that respect, I think that the legislation is doing a good job, because it provides a very broad brush that provides me with the interpretation that I am giving to you.

Is it the role of the government to prohibit the use of personal information? No. The use should be regulated, but it should not be prohibited.

Have I answered your question?

● (1035)

Mr. Michel Picard: Yes. Thank you.

Mr. Vickery, do you have anything to add?

[*English*]

Mr. Chris Vickery: With regard to the usage of your personal information, just for starters, I agree that we need to define what personal information is. Everyone needs to understand what the rules are and to not have any ambiguity there. So yes, I would say that there need to be clear definitions that everyone plays by. The usage by Amazon is not one of “malintent”; it is to offer a better experience for you, and clearly to not be taken advantage of. Now, that does go up to some interpretation....

I think your other example of posting something on Facebook is a little different, in that you chose to post that. Facebook did not share that on your behalf. You put it up on your Facebook wall. It's clear that you opted in to show this to the world. Getting a few calls from a travel salesman might be the appropriate consequence of oversharing that data, and maybe you won't do it in the future, but that was totally your decision. There wasn't another company making the decision for you.

Mr. Michel Picard: If a company looks at my line for the last two years and from analysis decides that my behaviour is such-and-such, and therefore another company starts to get in touch with me based on what is published publicly, it means that there's no bad intention anywhere. You cannot look at the end user as doing something wrong, but rather as using what is available to the company and proceeding.

Mr. Chris Vickery: I believe that's a very risky area to get into if you're saying that something you chose to put out in the public should be off limits. If you put something publicly on the public Internet, it's fair game for anybody in the public to view. Maybe the actions that companies take in response to that data being available could be seen as not so great; they'll deal with the reputational consequences of that when they decide to use that publicly available information.

Mr. Michel Picard: Thank you.

The Chair: Thank you, Monsieur Picard.

Mr. Daniel Therrien: Mr. Chair, if I may...?

The Chair: We're out of time, Mr. Therrien. My apologies. Do you have a brief comment?

Mr. Daniel Therrien: Mr. Chair, I'll try to be brief on the notion of what is publicly available.

A number of things could be said, but one of the elements to bear in mind is this: did the person who put up the information as being publicly available realize that this was what they were doing? This goes to what information should be given to individuals so that they make the appropriate consent decisions. Many people have no clue what they're doing. That's one point.

Another point, to be brief, is that there are regulations currently in Canada that define publicly available information. They are outdated. I would encourage you to look at them.

● (1040)

The Chair: Thank you. I know that it's a difficult subject to be brief with.

Mr. Erskine-Smith, you have literally 20 seconds, and if you can use less, that would be great.

Mr. Nathaniel Erskine-Smith: Mr. Vickery, you've mentioned that you downloaded the information you accessed on GitHub. You've said today that you think there is a potential illegality with respect to using information for a commercial purpose for which it was not collected. Have you shared this information with the proper authorities, including our Privacy Commissioner? If not, would you be willing to do so?

The Chair: A quick answer, please.

Mr. Chris Vickery: I was very quickly in contact with federal authorities in my country, and I am fully willing to co-operate with investigations that are relevant to Canada.

The Chair: Thank you, Mr. Vickery.

We're going to suspend and go into committee business. If all the guests who are not part of committee could exit as soon as possible, it would be appreciated.

Once again, thank you to Mr. Vickery and Mr. Therrien for testifying today.

[*Proceedings continue in camera*]