HOUSE OF COMMONS

Digital, Culture, Media and Sport Committee

Oral evidence: Fake News, HC 363

Wednesday 2 May 2018

Ordered by the House of Commons to be published on 2 May 2018.

Watch the meeting

Members present: Damian Collins (Chair); Simon Hart; Ian C. Lucas; Christian Matheson; Brendan O'Hara; Rebecca Pow; Jo Stevens; Giles Watling.

Questions 2501-2616

Witness

I: Chris Vickery, Director, Cyber Risk Research, UpGuard

Examination of witness

Witness: Chris Vickery

Q2501 **Chair:** Good afternoon. I would like to welcome Chris Vickery to this further evidence session of the Digital, Culture, Media and Sport Select Committee in our inquiry into disinformation and fake news. We are delighted that you have been able to join us, Chris, and you have come all the way from the west coast to be here in London today. We have been following a lot of the work you have done during our inquiry. It has been a great benefit to us and I know to other Parliaments around the world who are following similar issues. It is great to have a chance to discuss this with you today.

I thought maybe for the benefit of the record and the Committee you could explain a little bit about the nature of your work and how you got into it.

Chris Vickery: Currently I am the Director of Cyber Risk Research for a company named UpGuard. UpGuard is in the business of managing cyber risk more than just being a simple cybersecurity company. It is a holistic approach to understanding what is in your environment, how it is configured, how your configuration relates to third-party vendors and how their configuration relates to known data breaches and security issues. It is an "encompassing view" type of business.

How I got into it was that I was training to be a paralegal at a law firm in Austin, Texas. I had my foot in the door as a runner and I was going to school to be a paralegal. There was an opening in the tech department. I said to them, "I have always been tech-savvy. Why don't I take a shot at it?" I did that for a few years and started finding data breaches for clients of the law firm. I got involved in finding data breaches outside of work, just for fun, as a hobby. I started finding bigger and bigger ones and got more and more attention in the media. I started finding millions of people in data breaches involving social security numbers, drivers' licences and insurance information. There were plenty of Government-related ones. Then I started getting offers to do it professionally and I have been doing it ever since, first for one company and now with UpGuard as of April of last year.

Q2502 **Chair:** Your role is really advising your clients on how to protect their data online, to make it less vulnerable to data breaches. Would that be correct?

Chris Vickery: Yes, that is definitely an aspect of it.

Q2503 **Chair:** When you are finding evidence of data breaches, are you looking at systems that companies have and looking at their vulnerabilities or finding data that has perhaps been taken and stored in another place?

Chris Vickery: Specific to my work individually, I do not look as much at vulnerabilities as finding data that is simply exposed to the public internet. It is out there and available to everybody. Most of the time it turns out it was not supposed to be out there and available to the public internet but I know where to look to find data that is not necessarily hidden but kind of in the shadows.

Q2504 **Chair:** How is data exposed in this way? How is it that data ends up being publicly accessible or in places you would not expect it to be?

Chris Vickery: There are a few reasons or a few causes for this type of data exposure. Legally I would call it a data breach but it is easier to think of it as an exposure. It can come down most of the time to people cutting corners, taking the easy way out and not wanting to put authentication measures into absolutely everything that they code because it is easier to simply make it available to everybody. If it is available to the world, all of your other scripts are going to be able to connect to it. You just hope that nobody finds it who you did not intend to find it. That happens a lot more often than people realise. It is to the level of an epidemic, I would say. There is so much data out there on which people have just not bothered to batten down the hatches.

Q2505 **Chair:** Perhaps you could give us an idea. When you say "find this data", where is it to be found? What level of sophistication would you need to be able to find this information? Would anyone with a basic level of computer coding and training be able to do it?

Chris Vickery: You do not need to build a code at all whatsoever. You can use a web browser to find the vast majority of what I come across. You just need to spend some time reading about a protocol or a type of storage system. It does not involve special advanced anything. You just have to be lucky about where you click and know how to ingest a large amount of potentials and you will come across the gems.

Q2506 **Chair:** Within this information you are finding datasets, passwords and usernames that could then give access to further sets of data and information. Is that correct?

Chris Vickery: Yes, definitely. I do not take advantage of or use the passwords that I come across, but if I were a malicious actor I could definitely go to the next level and use those passwords to gain access to all sorts of advanced systems further down the line.

Q2507 **Chair:** As you say, if you can find this information then other people can find it too.

Chris Vickery: Yes. No question about that.

Q2508 **Chair:** We should cover, for the record, the legality of the work you do. If this is openly available, is it totally accessible and able to be used by anyone who discovers it, apart from the using of private passwords and usernames?

Chris Vickery: That is my contention, yes. It is not 100% figured out in American law. I know that we are in a different territory here but I believe the work that I do is completely within the confines of the law.

Q2509 **Giles Watling:** You say that you fell into it. It was almost an accidental thing; you found the data and you followed more and more. Are there armies of people like you doing this? Clearly this data can be easily accessed. As you say, you do not need to know how to code a computer or whatever. Are there armies of people out there doing this?

Chris Vickery: Figuratively and literally, yes. There are nation-state army-type groups doing this, as well as plenty of hobbyists and professionals doing it.

Q2510 **Giles Watling:** As you say, the data you harvest is open. Does it have commercial value? Can you harvest it, package it up and sell it in different ways? With statistics, you can get data and use it to back up any political point of view.

Chris Vickery: If I were less scrupulous I could gather this data, turn around and sell it for a massive profit, yes.

Q2511 **Giles Watling:** You once said that cryptocurrencies are an inevitability. We know that they are sometimes used by bad actors in nefarious ways. How can we regulate that? How do you see that being regulated?

Chris Vickery: That is a very difficult question. I do not believe there is a good understanding in general, by the people that would be enacting the regulations, of what a cryptocurrency is or can be, and I do not think we have advanced far enough in blockchain technology to understand what it is we need to understand. I do not know of a good way to regulate it. It is definitely something that needs to happen. There needs to be some regulation and some control and there are some anti-money laundering issues that need to be worked out, but I do not have a solution right now. I do not have a financial background; I am more on the cybersecurity side of things. However, if the right minds come together I think it can be done.

Q2512 **Giles Watling:** What makes you think that they are going to be an inevitability? Forgive me, I do not understand that.

Chris Vickery: I believe the specific tweet you are talking about had to do with the idea of bounties being put out. Was that the concept?

Giles Watling: Yes.

Chris Vickery: What I meant as an inevitability was the bounty aspect of it, not necessarily the cryptocurrency aspect of it. What I was tweeting about was my personal opinion that somewhere down the line somebody is going to figure out that, "I can put up a token sale or cryptocurrency offering and get people to buy into it and have it done in an anonymous way". A smart contract can be created that pays out somebody when a fact changes. Smart contracts are all about verifying data and a human does not have to interact to get the contract finished. It is a computer algorithm, basically. If you did that based on whether or not somebody is deceased or living, you could very simply anonymously have a hit put out on somebody. That is what I believe is the inevitability. Somebody will figure that out, put all the tools together, and that type of thing will become a reality.

Q2513 **Giles Watling:** You feel that we need to move in and start regulating?

Chris Vickery: We need to figure out a way to prevent that type of reality from happening. I do not know the best way to do it. If it is regulation, which is probably the easiest way to do it, we need to get some controls in place.

Q2514 **Christian Matheson:** Mr Vickery, good afternoon. I want to follow on from Mr Watling's questions. You have painted a picture of the business that you are in. I am guessing a business will come to you and say, "Chris, we need you to check and see if our data is secure or if any of our data is publicly available that should not be". Is that right? Will they come to you and ask that?

Chris Vickery: One of the platform offerings that UpGuard has is known as BreachSight. That is essentially a concept of us shepherding an enterprise-level company's web assets, keeping an eye out, seeing if they have any easily publicly exposed data or not so easily publicly exposed data, letting them know and giving them the heads-up so that they can get secure. That is part of what we do.

Q2515 **Christian Matheson:** You also talked about this being a hobby for you.

Chris Vickery: It was a hobby before I started doing it professionally, yes.

Q2516 **Christian Matheson:** What drew you to the SCL, Cambridge Analytica, AggregateIQ network of companies?

Chris Vickery: That was a situation where it was about 9.30 pm California time, around 20 March, I believe. I have a TV that sits on the wall behind me in my home office and I was listening to a news show. They had mentioned SCL. I did not know who AggregateIQ was. I was prompted in my regular searches through the internet to look and see if SCL had any open source code. There is a website called GitHub where developers can put open source code and collaborate. A lot of big companies do contribute open source code to the public realm and I was curious if SCL had, because that gives you a view into their processes and various techniques sometimes. They were being talked about in the news. I was at my computer, I was on GitHub, so I figured, "Hey, why the heck not? Just take a look at it".

When I searched for SCL Group there was a result for a repository apparently owned by a person named Ali Yassine, who worked for or maybe still works for SCL. Within that there was a reference in his open source code repository to ab.aggregateiq.com, in one of the comments. I had never heard of it. I went to the website and I was met with the phrase "peer data intelligence". It seemed to me when I scrolled further down that they did voter intelligence, voter canvassing, matching up data to donors and everything. It seemed weird to me that SCL and this group would be working in the same kind of area. Cambridge Analytica is basically part of SCL. Why would another company be doubling up efforts? That is what it seemed to me, just in my very initial understanding of what they were doing.

It did not make sense to me so I started looking into the company a little bit more, enumerated their subdomains, and I saw that one of the subdomains was gitlab.aggregateiq.com. I went to it. I knew what GitLab was, but then I noticed the registration page was active. It had never been turned off. They were, in essence, inviting the world to join their GitLab repository.

Q2517 **Christian Matheson:** Sorry to interrupt. Is this the same night that you saw it on TV?

Chris Vickery: Yes, this is the same night.

Q2518 **Christian Matheson:** This is quite a quick process?

Chris Vickery: I tend to spend a large amount of time on my computer so I accomplish a lot of tasks in one sitting, very much to my wife's dismay. Yes, this was in the same night, definitely. It was within a half-hour of me running the search.

Christian Matheson: Sorry, I interrupted your flow. Carry on, please.

Chris Vickery: Was I explaining in too much detail?

Christian Matheson: No, it was great.

Chris Vickery: They were inviting the public to register on their GitLab repository. It occurred to me that this may be a misconfiguration, or maybe they really are collaborating with the world on developing the software that is inside of here. I did not know what software was inside of there or if any software was inside of there. I created an account, logged in, and it hit me that, "Oh my God, this is the repository of the tools that AggregateIQ and, it looks like, Cambridge Analytica and SCL are using. This is the source code behind these scripts". I immediately understood the gravity of the situation and thought, "I need to document this and download what is here because a lot of people are going to be very interested in seeing this".

Q2519 **Christian Matheson:** It was the underlying programme as opposed to any data storage itself?

Chris Vickery: The line there is a little fuzzy, but yes. The best way to understand it, the way I have been explaining it to people who have been asking, is that in a kitchen you have the cooking pot and the utensils, the spoon, the ladle and all that stuff. Then in the refrigerator you have the meat, potatoes and all of the foodstuffs. When the whole process is going you pull out the stuff from the refrigerator, you pull out the utensils, you mix it all together and the magic happens. I found the cooking pot, the ladle and all that, and the codes to enter the refrigerator. I did not actually enter the refrigerator and get out the raw data. I could have. It is my contention that if I were a malicious actor I could have got into everything that AggregateIQ has their hands in.

Q2520 **Christian Matheson:** You make it sound, for somebody like you who knows what they are doing, quite simple and quite easy. Were you surprised at how straightforward it was?

Chris Vickery: I have been doing this so long that I have seen a lot of crazy stuff. As far as surprise goes, I am not necessarily surprised. I am impressed by the gravity of the situation and the scope of what was there. It is not a common find but it is not out of the realm of possibility that somebody involved in so much would make a mistake along the way and have a little chink in their armour, or whatever you want to think of it as. I am sorry, did I answer your question fully there?

Q2521 **Christian Matheson:** Yes. If I think about the online communities you are talking about that will work with this open source code, do some of them ever come with discussion boards where they discuss things like this? If so, has there been any chatter on any of these about what you have found?

Chris Vickery: Do you have a specific chat board or forum in mind?

Christian Matheson: No. Forgive me, I know nothing about this whatsoever.

Chris Vickery: There has been mention, definitely, on plenty of sites. There are more socially angled sites like Reddit. There are sites like GitHub. I was explaining GitHub. It does not really have a forum for people to congregate and talk about this stuff. It is more like there is a project and people can comment within the project's commenting page in various ways. These days there is not so much an open message board community somewhere where people talk about specific projects like this. There are so many of them, they are so spread out and they are so varied it is not easy to answer that question in that way.

Q2522 **Rebecca Pow:** I am very interested in the fridge. Why did you not go and find out what was in the fridge and what did you suspect was in the fridge?

Chris Vickery: I believe that accessing the fridge or the raw databases holding the information would have put myself and potentially my employer in legal danger. I avoid using passwords and usernames when I go looking for data because I do not want to be prosecuted, although it is still questionable what level of going into it would be a criminal offence in America. We have the Computer Fraud and Abuse Act and it is extremely vague. They never define what unauthenticated or authenticated access is. It is up to the prosecutor's determination. I do not want to get so close to that edge that a prosecutor could determine that I have committed a criminal act. That is why I do not.

Q2523 **Rebecca Pow:** You suggested that for people like yourself it was easy to find this way into the fridge. Why do you think that was? Do you think that was on purpose?

Chris Vickery: I have been finding large voter databases since December 2015, about one to two a year, and I have had that same question every single time I have come across one of them. People ask me, "Do you think that people were co-ordinating by leaving this open on purpose?" The answer I have come to to that question, which I believe is the same question you are asking here, is that I cannot say one way or the other. It would be a smart way to do it. If you wanted to collaborate with somebody but have plausible deniability, this would be a good way to do it. You can say, "Oh my God, it was open to everybody. I can't believe I made that mistake", while under the table you are grinning, nodding and saying, "I got away with it". That is definitely a possibility. I cannot say that is why this happened, I cannot say that is why any of the large databases that I find have been exposed, but it is a possibility.

Q2524 **Rebecca Pow:** There was no protection on it? They did not protect it?

Chris Vickery: None. Not at all. Anybody in the entire world with an internet connection could have got into everything.

Q2525 **Chair:** Why do you think this information was left in such an exposed state?

Chris Vickery: I believe the most likely explanation is that when the developers or staff members, whoever worked for AggregateIQ, were interacting with GitLab, the repository and all the code and scripts, they were likely accessing it through other tools accessing the GitLab server. For example there is GitKraken, which allows you to have a good graphical user interface on the desktop but plugs into a Git server. The Git server has to exist. They probably could have turned off the webpage serving the registration of the login page and still been able to access the Git server; they just never turned it off. They were using third-party tools, never going to that login page and noticing that the registration link was active. That is why I believe they had the oversight of leaving that up.

Q2526 **Chair:** Would the consequence of them working in this way, if it was a mistake that the data was left there, be to allow anyone working at AIQ, SCL or Cambridge Analytica to effectively access, using your analogy, the same tools and the same fridge?

Chris Vickery: There is no limitation on who could have accessed it: anyone in the entire world; anyone in any country who has an internet connection.

Q2527 **Chair:** Certainly within those companies it would facilitate a completely open way of working all the time. I suppose you also have to know it is there. You found it by chance.

Chris Vickery: I followed a trail to it.

Q2528 **Chair:** Yes, but there are other people who would know it was there and could easily access it because they knew it was there. Do you think there was a particular virtue for them in making it so easily available?

Chris Vickery: I really do not know their intention behind it. It would just be speculation on my part to guess at why they did it. It would be one thing if I had comments within the GitLab repository saying, "We are leaving this open for x, y and z", but I have not come across anything that said why they left it open. I do not want to guess and put any thoughts into anybody's head about why.

Q2529 **Chair:** It would also be quite a convenient excuse. If someone was asked the question, "Did you give this data to this person?" you could say no, but you could have let them know where to find it.

Chris Vickery: Very much so, yes. Very easily.

Q2530 **Chair:** We know that obviously AIQ developed the Ripon platform that was used by SCL. Could tools like that have been connected to this dataset?

Chris Vickery: Let me clarify the SCL, Cambridge Analytica, AggregateIQ thing. I believe the flow was that Cambridge Analytica pitched Ripon to Senator Ted Cruz's campaign originally and Ted Cruz's campaign was paying for the development of Ripon when they thought Ripon already existed. Cambridge Analytica had subcontracted it out to AggregateIQ without the Cruz campaign knowing. I believe that is the situation. Cruz's campaign thought this tool existed but they were actually developing it. Cambridge Analytica was not developing it, AIQ was. I think their relationship is a little bit closer than they admit. Ripon does plug into lots and lots of voter databases—it certainly does. Is that the answer to your question?

Q2531 **Chair:** Yes. We have had lots of conflicting evidence during this inquiry so far about the value of different datasets. Indeed, last week Aleksandr Kogan said that he did not feel that the results of his surveys were of much value as a predictive tool. Then again, none of these datasets were designed to be used in isolation but to be combined with other things.

In this case, building on the analogy you have used, it strikes me that what people are doing is combining multiple datasets that help to create a richer profile of an individual and then linking that profile to their Facebook profile, which is the mechanism by which they can be contacted and communicated with. Once you have created your different profile groups based on the different opinions, different beliefs and different fears that people have, and created your targeting groups from that data, all you then need to know is who those people are on Facebook and where to find them. They are your custom audience. Then you do not need the original datasets anymore. You have created the hybrid product that you are going to use for your campaign. Do you believe that is what was being done? Do you believe that is why people are saying that you could destroy that original data or give that original data back, because you no longer need it?

Chris Vickery: My understanding is that saying it has no value—just to start at the beginning of what you were asking—is a little bit deceptive. You have this graph in front of you, I believe, talking about various sources—

Chair: If I could just say this for the record, Christopher has given us this chart. It has been published by the Committee and will be available on the Committee's website as well. Go ahead.

Chris Vickery: It is important to realise that this is something I put together that has to do with my current understanding of a data flow, not necessarily who has a business relationship or who is the same company as somebody else. This is just the way data is flowing. It is different from an organisational graph.

The dataset that is involved in the Deep Root Analytics box down there cost $33 million to develop, I believe. Back when I found the Deep Root

Analytics bucket that had all that data in it, I looked into it. I believe it was $30 million-plus that was paid by various campaigns and so on to develop that one. Anything that comes out of $33 million is not going to be worthless. Large datasets have value. I am not saying that Deep Root Analytics has directly given this to AggregateIQ or anybody but it is in the same flow of data and it does have extreme value.

Another example is related to the first large voter database I found in 2015. I pretended to be a potential buyer of a large database like that and I called up a company that sells these voter databases. I said, "If I wanted to buy this, how much would it cost me?" They quoted me $260,000 for a list of names, addresses and phone numbers for the entire United States. Just a simple list like that they were trying to sell for quarter of a million dollars. I am sorry, what was the rest of the question?

Q2532 **Chair:** People are collecting these datasets. They are doing that, presumably, to create target groups out of those datasets. Let us say you are in the Ted Cruz campaign or the Trump campaign and you are saying, "We want a dataset of people who believe there should be tighter restrictions on guns. That is a massively motivating factor in how they are going to vote in the election. We want people who have that belief or that view, and who live in New Hampshire. We are going to try to create a database of people like that. The way in which we will communicate with them will be using Facebook. Therefore, we need to link those datasets to an actual person, in order to build our custom audience. Then we could also ask Facebook to find people who are like that".

If that is what is being done, once you have created those target groups based on data and you have identified them—you have their email addresses, you have their Facebook profiles—the actual dataset that was used to create this hybrid set you do not particularly need anymore. Not for that campaign, anyway.

Chris Vickery: Ted Cruz's campaign does not need it anymore. That is what you are saying?

Chair: Yes.

Chris Vickery: I was thinking for a second that you were talking about the people who have this data that this is being pulled from. To Ted Cruz's campaign, yes, all they need to know is who is the right person to target. They do not necessarily need the background data because the number crunchers and the analytics companies are dealing with all that.

With the written documentation from AggregateIQ's GitLab, they describe this process. What is happening is this "database of truth" exists with all the data, and then they pull out escrow copies—they are called escrow databases in their documentation—of what is being asked for and is relevant to a certain campaign. That would be your idea of the audience of pro-gun people from New Hampshire. It would go out into an escrow copy that would be available to Ted Cruz's campaign, or if AggregateIQ determines that this group of people is relevant to this campaign they would pull out that group. Ted Cruz's campaign, or whoever's campaign, does not need access to the whole database of truth. They just need access to that list of people. Yes, I believe that is an accurate depiction.

Q2533 **Chair:** Is the database of truth—it is starting to sound a bit like a science fiction novel—a sort of constantly evolving organism that gets richer and richer depending on what is put into it? It gets, if anything, more effective?

Chris Vickery: Yes. That is also described in the documentation. The foundation of the database of truth, according to what is written in there, is the Republican database called the Data Trust voter vault. There is a company named Data Trust that is very close to the Republican party, the GOP, in America, to the point that I would say they could be considered the same entity. They argue differently. That is the foundation of the database of truth, at least as far as American politics is related to this.

Then they add state voter files to, I assume, corroborate what is in the RNC's database. Then they add consumer purchasable data such as from Experian. It is limitless what other companies would sell you, data similar to Experian. Then they add lists that the campaign itself gives to them. Senator Ted Cruz could have a list of people who have donated to his campaign. He would then want to integrate that into it if he was looking for people to communicate with.

Then with the escrow copies that come out, there is the ability to add the changes that are made to the escrow copy back to the database of truth if they are approved. It is not like the campaign has full access to the original dataset and can change things willy-nilly. No. They can make changes to their own dataset and if it meets a bar of acceptability or would be helpful to the original database of truth, it can be fed back into it. The different iterations are kept track of.

Interestingly enough, the RNC has argued in hearings in front of the FEC, the Federal Elections Commission, in America, that the RNC's Data Trust giving that full dataset to campaigns is not an illegal contribution or a contribution that is of any value of all because they have agreements where they get data fed back to them from the campaign. If you consider that as part of the equation, if the RNC has the same agreement with AIQ, any data that is gathered from an "in" campaign goes back to the database of truth. It would then go back to the RNC database and enhance everything along the way. If you ask me, giving out that dataset is extremely valuable and is worth more than the in-kind contribution that they get back, but that is up to interpretation. That whole data flow is definitely occurring.

Q2534 **Chair:** Going back to some of the things we know about SCL and some of the people from AIQ, the campaigns they were working on, we know that staff from both AIQ and SCL were working on what was done for the Bolton Super PAC in 2014, the mid-term races. Aleksandr Kogan's research work was also feeding directly into that process. Dr Kogan said he was not aware of the candidates they were working on but he was aware that this data was being asked for to work on election races.

Do you believe that the datasets developed for the state races that were being supported by Ambassador Bolton would all end up in the database of truth and that future campaigns could benefit from that acquired knowledge as well?

Chris Vickery: The raw data itself maybe not, but it is extremely likely the derivatives would be integrated into it. In this type of industry, data does not just go away. It does not just disappear. It is sticky. It gathers up. The good stuff stays. Even the bad stuff stays, but it is not used. It is held in an archive somewhere. Nothing disappears.

Q2535 **Chair:** Is the derivative in this case, for the Bolton work, going to be a psychological profile of somebody linked to their Facebook account? Then they are probably also linked to other datasets about them and their electoral roll location. That would be the derivative of those pieces of research?

Chris Vickery: Within Ripon there is an ability to search for people based on OCEAN psychographic scoring. Yes, derivatives would have been integrated if that is going to function at all. I do not know why that page would be there if it does not function that way.

Q2536 **Chair:** Was Ripon used in the Donald Trump campaign by Cambridge Analytica?

Chris Vickery: I cannot say that it was. I think the most likely situation is that things that they learned from developing Ripon were leaned upon but I have not come across anything that directly says, "This was used in the Donald Trump campaign. This was used here, there and there". If it ever existed it could have been moved or deleted. I do not know if it did or not. I have no idea.

Q2537 **Chair:** Do you think any of the derivatives from Dr Kogan's work could have been used to support Cambridge Analytica's work for the Trump campaign?

Chris Vickery: Yes, definitely. They do not need the original allegedly ill-gotten Facebook data if they have the derivative modelling, training data and everything from it. They do not need that original dataset anymore. I do not know why they would learn it and not use it.

Q2538 **Chair:** The same for the Cruz campaign. They did a lot of psychological profiling of voters for Ted Cruz.

Chris Vickery: No question about it, yes.

Q2539 **Chair:** Alexander Nix said that was not used for the Trump campaign. Does that add to the long list of clarifications we need from Mr Nix? You think there is no question that would have been used?

Chris Vickery: If I were on a jury in America taking this in, I would put forward the conclusion that yes, that is the most likely situation that must have occurred, just about beyond the shadow of a doubt.

Q2540 **Chair:** On that basis, psychological profiling was used to support the Trump campaign.

Chris Vickery: I would say that would be an accurate statement. People are going to argue against it but I think so.

Q2541 **Chair:** Whether or not the Trump campaign actually commissioned it, they benefited from previous work that had been done.

Chris Vickery: If Cambridge Analytica was involved in the Trump campaign, which is a yes, then yes, the Trump campaign would have benefited from psychological profiling. Yes, I believe so.

Q2542 **Ian C. Lucas:** Can I stick with the fridge? I find it extremely helpful. I am particularly interested in what I call "the ingredients"—what is actually in the fridge, which you did not explore. Is that right?

Chris Vickery: I did not use any usernames or passwords to log in to any databases that feed into this stuff.

Q2543 **Ian C. Lucas:** If the ingredients had been put in the pot and then mixed with the utensils, that creates the product that is very valuable. Do you know what the ingredients were? For example, was it individual data like names, addresses and passwords? Do you know that?

Chris Vickery: Yes, because the tools themselves have the field names that feed into them and everything ready to parse out that data. On this hard drive that I am delivering to the Committee you will find all of that, the code itself that deals with the ingredients, and you will be able to see for yourselves the field names and everything explained, at least everything that was present on the GitLab. You will be able to determine a bit of that.

Q2544 **Ian C. Lucas:** Is any of that information Facebook data?

Chris Vickery: What would you call Facebook data? Somebody's Facebook ID, the ability to target somebody on Facebook and things about somebody's Facebook account but not their Facebook postings? Would that still count as Facebook data? Yes. Then that is there.

Q2545 **Ian C. Lucas:** We know from the evidence of Mr Kogan that information was taken from Facebook and went through GSR to Cambridge Analytica. What I am very interested in is the chain between Facebook, GSR and Cambridge Analytica, to AIQ. Is it possible to trace the information from Facebook through to AIQ?

Chris Vickery: You are referring specifically, I assume, to the allegedly ill-gotten Facebook data?

Ian C. Lucas: Yes, the data that Facebook wanted back in 2015.

Chris Vickery: I do not know if you will be able to determine definitively that that chain of events happened, given what is on the hard drive. You will see, though, I believe, that the derivatives—the trainings and the models gathering what is relevant—are what counts and what matters. The original Facebook dataset, the raw data itself, is very inconsequential. Whether or not that flowed through here does not matter. It is like you have a steam engine ready to go. You do not need the prototype of the steam engine or the framework you built it on anymore. That can go back in the warehouse. You have the thing to toss everything with. You have it there to go. I cannot say one way or the other if you will be able to determine concretely that that flow happened.

Q2546 **Ian C. Lucas:** If the data was inconsequential from Facebook, why did Facebook want it back?

Chris Vickery: It was inconsequential to the modelling once the modelling existed. Cambridge Analytica, having that modelling, have different reasons and purposes for existing from Facebook's reason for existing. Facebook legally wants to be able to say, "That data does not exist anymore". For their purposes of not getting sued they want to be able to say, "That data does not exist anymore". They have totally different purposes for existing. That would be Facebook's motivation for getting it deleted, not to stop the data modelling, analytics and graphing that existed from it.

Q2547 **Ian C. Lucas:** Once the data had moved to Cambridge Analytica in 2015 via Mr Kogan, is it the case that the value of that could have been used by Cambridge Analytica in its relationship with AIQ? It seems to me that once the value is released from Cambridge Analytica to AIQ, it is gone. You cannot give it back unless you chase it down the chain to AIQ to get it back.

Chris Vickery: It is a little complicated. I do not believe Facebook ever claimed that they got rid of any derivative works and modelling that came out of the dataset. They just wanted the raw dataset gone. There was no way for them to know what was modelled from that dataset and they are not interested, as far as I know, in getting the modelling destroyed or, even if they could request it get destroyed, the derivatives of it. Does that make sense?

Q2548 **Ian C. Lucas:** Yes. When Cambridge Analytica got the data from Facebook they made the pudding. They made the dish in the pot, if we go back to the analogy, and then they passed it on.

Chris Vickery: They made the pot to put in more dishes in the future. They moulded it from the pudding, basically. Does that make sense? Once they have moulded the pot they can make more pudding in the future with different datasets. They do not need the original pudding anymore. That was what Facebook wanted to delete, the pudding, so they did not get sued.

Q2549 **Ian C. Lucas:** They had a recipe?

Chris Vickery: Yes, essentially. That is one way to think of it.

Q2550 **Ian C. Lucas:** They had the recipe so they could give back the ingredients and make the dish on the basis of the recipe, which they then passed on to AIQ.

Chris Vickery: It appears that yes, that would be a likely scenario.

Q2551 **Ian C. Lucas:** What also mystifies me about AIQ is that we know AIQ had a relationship with Cambridge Analytica and we know that Cambridge Analytica had a relationship with Leave.EU. I do not understand why AIQ then had a commercial relationship with a different organisation, Vote Leave. Do you have any knowledge?

Chris Vickery: Of why, or the relationship?

Ian C. Lucas: Do you know why that is? Do you know why Vote Leave had a relationship with AIQ?

Chris Vickery: The why of it, I do not know. That is something that is in their heads, why they became partners or whatever. What I can tell you definitively is that there were projects within this GitLab repository for Change Britain, Vote Leave, DUP and Veterans for Britain. All of it looked like it had been worked on by AIQ.

Q2552 **Ian C. Lucas:** You have presented a very interesting list that you have just referred to—DUP, Vote Leave. I know that separately and later the Countryside Alliance was also involved with AIQ.

Chris Vickery: For legal purposes, I have to be very clear that there was not evidence that Countryside Alliance was involved in any Brexit stuff. It was later on.

Q2553 **Ian C. Lucas:** Yes. I want to make that clear. I agree with that entirely. We know that AIQ had a clear commercial relationship with Vote Leave because there were payments made to AIQ by Vote Leave.

Chris Vickery: The results of those payments, the projects, are right here.

Q2554 **Ian C. Lucas:** Okay. We are going to have to ask other people, probably, about that relationship in due course.

Chris Vickery: The why.

Ian C. Lucas: The why, yes, but there is clearly a chain here, if I can mix metaphors dreadfully.

Q2555 **Chair:** Is the information you have on that portable drive still available? Is that still on GitLab or was that removed?

Chris Vickery: That was taken down 11 minutes after an e-mail was sent notifying AIQ that this was perhaps exposed more than it should be.

Q2556 **Chair:** Was that sent by you? Did you notify them?

Chris Vickery: That was sent by a journalist. Normally it would come from me but—and I have not delved into this in any interviews of anything because it is a legally sensitive topic—the reason I did not send it is that there were prima facie cases of potential unlawful activity going on, and it seems weird for me to notify people that I suspect may be committing criminal acts that their criminal acts are being exposed to the world by something that they have put out. I allowed a journalist that was going to cover the situation to send the notification e-mail just to be one step removed from that.

Q2557 **Chair:** That is clear. Was the journalist who sent the notification a journalist whom you had notified about the existence of this data?

Chris Vickery: The way that happened was when I originally found a little breadcrumb on GitHub, which is a little bit different from GitLab, the public GitHub; I was in communication with this journalist who has become kind of a friend over the years. I said to him, "There is this interesting little thing involving SCL. I know they are in the news and maybe it is something you want to look into". I did not think it would lead anywhere, and then this whole rabbit hole opened up, and since he was already aware of the first breadcrumb he was aware of the further investigatory issues that came up.

Q2558 **Chair:** There are lots of cooking analogies today. The breadcrumb that you refer to, is that just the reference to AIQ that you mentioned at the beginning, and the other information that you were looking at with SCL was something else?

Chris Vickery: Just ahead of that was the fact that there was this apparent SCL employee, Ali Yassine, who had the GitHub repository; it was just the existence of some SCL-related code. That was the initial, initial part and then that led to the comment about aggregateiq.com.

Q2559 **Chair:** I am sure you will be familiar with the testimony given by AIQ to the Canadian House of Commons.

Chris Vickery: Yes, I did watch that.

Q2560 **Chair:** They were asked about, "Have you got data?" to which they say, "No". But do you think that they may not have data but they have access to data? Do you think that is a real distinction that we should take an interest in?

Chris Vickery: Yes. However, they are also not being quite fully straightforward about having data. They may not have the large oceans but they have the residual ponds from them accessing it, such as I was able to locate my own information in a spreadsheet called Online Activists, which had been pulled and was available to AIQ. They had my home address from when I lived in Austin, Texas. I don't know when the data was pulled, but I found myself in there, my own personal details.

Q2561 **Chair:** I suppose what I was interested in is: is it a convenient excuse for people to say, "We don't keep datasets. We don't keep databases. We don't hold this data"? They continually say things like that, which may be true but it could be it exists in a cloud system or in a remote location that they have easy and ready access to?

Chris Vickery: Yes. It is a weasel way of saying it. It is a technicality, kind of kindergarten, cross your fingers behind your back when you are giving testimony type of answer.

Q2562 **Chair:** Who owns these? Who do you think owns the database of truth?

Chris Vickery: From looking at the code, different scripts on here, one of the companies that seem to be referred to as one of the holders is WPA Intel, which is a political intelligence analysis company, one of the head people being Chris Wilson. It may be that if you were to ask WPA, I wouldn't be surprised—and they have not been asked this as far as I know—if they would try to give you the same answer and say, "It is being held somewhere else". It brings up the idea of, okay, if something is just ephemeral and existing between our organisations, each side can point somewhere else and say, "No, I am accessing it from there" or, "No, I am just accessing it from there".

With technology these days and peer-to-peer stuff and whatever, you could arguably make a file system that has no real owner, that is just in existence in cyberspace, so you may run into those weasel answers, left and right, and never find somebody who will admit to having the data itself. It is purely possible.

Q2563 **Chair:** That would support the portrait that Chris Wylie painted to the Committee of here you have a sort of franchise model of work. Effectively, you have a group of people who work to a common interest but they are legally separate from each other, and the dataset that they work from is not owned by any one of them but is just held in common. That would support that model as well.

Chris Vickery: Yes, and that would be a very convenient way of avoiding lots of potential problems that they might run into. Yes, that would be beneficial.

Q2564 **Chair:** Why do you think the EU referendum files were uploaded to the AIQ GitLab?

Chris Vickery: What do I think of it?

Chair: Why do you think they were? A lot of this data would have been based on American elections. Was it just there for convenience because AIQ had worked on it and worked on other things too?

Chris Vickery: Yes. They had projects from lots of things that they had worked on. I don't see any reason why there would have been a separation. If they have their repository of projects they worked on, those are projects they worked on.

Q2565 **Chair:** Yes, so it is just the deep store of all the stuff that they have done?

Chris Vickery: Yes.

Q2566 **Brendan O'Hara:** Going back to the chink in the armour that you found and this idea that data could have been left for people to find, when you found that, was there any way of telling if you were the first person to have found that information or is there a way to find out who had been there before you?

Chris Vickery: Hypothetically, there are ways that you could have logging in place to see who accessed it but the person in control of that logging would also be able to modify that logging, so I don't know if you could rely on it to a forensic level but, theoretically, they could have been keeping logs of access. I cannot say there is a disinterested third party that would know, but logging could have been enabled.

Q2567 **Brendan O'Hara:** That was not something that you could have found out when you discovered it? This idea that perhaps it had been deliberately left there, as you said earlier on. There is no way that you could find out who had accessed it before you?

Chris Vickery: Not unless I was willing to cross that boundary into trying to exploit the system, so there is no way that I could have—within my understanding of the law—come across a log of who was accessing it. I didn't come across any sort of logging of who may have accessed it.

Q2568 **Brendan O'Hara:** Would AIQ have been able to know who had accessed it?

Chris Vickery: Theoretically, yes. I don't know if they were keeping those logs or not but, yes, they would have been able to. Yes, they could.

Q2569 **Brendan O'Hara:** Could I move on to Facebook for a moment? Could you describe the issues that you believe are still occurring between Facebook app usage and potential exploitation of data? Does it still exist?

Chris Vickery: What do you define as exploitation? Like, mass harvesting of people's private information?

Brendan O'Hara: Is that still possible, given what has happened?

Chris Vickery: They made a lot of changes within the past two weeks or so, and I am not sure what app developers are still allowed to gather at this very moment. My feelings in general are that Facebook has been way too open with information and allowed app developers way too much access, as well as allowing too much data to be gathered even through their simple search bar, as well as the password recovery function, which journalists have used for a while as well as companies seeking to exploit the "data".

What I mean there is there was a way to say, "I have lost my password. Here is my account", and it would give you a little thing that pops up and says, "Okay, let me text you with this number and give you a code to recover your password". You are not trying to recover the password. You are trying to get that little phone number, so that then you can match it up to somebody else; just a little trick of the trade that people have been using for a while.

There are lots of examples where Facebook has not been as great as it could have been in protecting people's information. But are you meaning within the past two weeks, those changes?

Q2570 **Brendan O'Hara:** It is a two-parter, isn't it? It is up until the last two weeks. It is what was in it for Facebook to build a platform that made data abuse so easy to achieve.

Chris Vickery: It was profitable.

Q2571 **Brendan O'Hara:** In what way?

Chris Vickery: If you make your platform very open and easily spread and you give app developers incentives to use your platform and easily develop apps and gain data themselves, they are going to have every incentive to develop apps and gain data. Then when people use those apps they have to log in to Facebook, so they have to have a Facebook account. You have a spiralling graph of you have a platform and a bunch of happy app developers pushing a lot of apps out to the public. The public has to have a Facebook account originally to use those apps, so they will sign up to Facebook. It is a flowing lifecycle and then people see ads that Facebook puts up because they created an account and Facebook makes money.

Q2572 **Brendan O'Hara:** Do you think it would be fair then to say that Facebook deliberately created this system that was open to exploitation because it was profitable and this idea, which we have heard on numerous occasions in this investigation, that they were naive and did not quite know what they were creating and they have been as surprised as everyone else has been by these data breaches, is in fact nonsense?

Chris Vickery: It would be harsh to think of it as they created something exploitable because they thought it would make money. I think it is more accurate to think of it as, "This was working, so we had incentive to keep doing this and looking for more ways to make it work and it kind of evolved." They perhaps should have put some constraints around that evolution, but they did it in a way that was largely profitable, and they had incentives to do that and they did not put as much control on it as they should. I will just put it that way.

Whether or not it was foreseeable? Probably. Whether or not they should have foreseen it? Probably. They definitely had the minds and the talent to foresee it. I don't know if you could easily say they did it because it was profitable and they knew it was exploitable but, yes, kind of.

Q2573 **Brendan O'Hara:** How early in this process do you think Facebook would have been aware of the fact that developers would have been interested in the data for purposes that went way beyond simply adding social functionality to their applications?

Chris Vickery: I think Facebook was very much aware that developers were interested in this, and I think that was part of the selling part to developers. Not that they were actually selling to developers but the selling point to get developers to use their platform. I believe they very much understood that developers could gather as much data as Facebook was allowing them to gather.

Q2574 **Brendan O'Hara:** That is from a very early stage or was this a revelation that would have appeared in 2014 when they then took action to prevent it, or did they know well in advance of that?

Chris Vickery: When they were designing the system they would have known what developers would be able to access. They would have just known. They would not have necessarily known, "And this is the key to our success" or, "We have to do this otherwise we won't get developers". They would just have known that developers can have access to this. Whether or not you have the cognitive dissonance to not realise what you are enabling is a question for somebody to interpret, but they definitely knew what developers could access.

Q2575 **Brendan O'Hara:** Do you know whether any data derived from Cambridge Analytica's models got loaded on to Facebook using their custom audience feature?

Chris Vickery: I don't know the answer to that. If it was loaded on or not, I don't know.

Q2576 **Brendan O'Hara:** When you testified to the Canadian Committee you discovered that a Facebook app linked to AIQ was classified under the games category of Facebook apps. Could you tell me a bit more about that app and how you discovered that?

Chris Vickery: I wish I knew more about it because I find that very confusing. The way I discovered that—and I have discovered one or two aggregate AIQ-related Facebook apps since then—was looking through this co-depository, finding reference to an app ID, looking it up through the Facebook API explorer and seeing the name BBAGGIQ Reach, or something like that. A similar app to that was involved in the Breitbart's scraping that was going on. When the app ID is in the code and it makes a reference to AggregateIQ in the name of the app, I consider that pretty good evidence that it was AggregateIQ using that app in some fashion, and one of them was classified as a game. I don't know how they were using it or if they were pushing it to people for any use. I don't know. I wish I knew more about it. I wish they would come forward and tell us why various apps on Facebook operated the way they did.

Q2577 **Brendan O'Hara:** Do you know what the game was?

Chris Vickery: No, I can get you the app ID but that is as close as I can get. It has been taken down, I believe, at this point. I don't know what the game was or if it really was a game. That is just a category that somebody had classified it under.

Q2578 **Brendan O'Hara:** Do you know how long it remained active?

Chris Vickery: I do not. Facebook might be able to tell us.

Brendan O'Hara: Yes, we may speak to them again.

Q2579 **Rebecca Pow:** I want to pick on something that you said to the Canadian Committee on 17 April: "the use of Facebook is not only about the taking of information, but about the planting of information". Could you very briefly expand on that?

Chris Vickery: What was the exact context about planting data? Can you—

Rebecca Pow: You suggested that it was about planting information. You were actually giving a warning, in fact, that people should keep their eye open about "the very focused efforts of others who rely on Facebook as a pillar of their operations". Can you expand on that? It sounds somewhat sinister.

Chris Vickery: Facebook and any platform that people interface with can be used as a manipulation tool to influence opinions, and actions and behaviours. An example that I gave on Twitter is of a way that you could use advertisements not only in their original intended way to bring a message to somebody but to suppress a message.

If you know that a certain percentage of people are adverse to, let's say, a certain colour—this group of people hates purple—if you don't want people to see that your opponent is advertising a rally or whatever, you can make sure there are purple ads shown when people click on an article about that rally. If you understand what makes people hate something, you can make sure that ads appear that bring about that hate. You can influence people to not like something almost as easily as you can influence somebody to like something. You just need to understand the modelling, the graphs and all that stuff. Basically, all the work that has been done would enable that type of action to be taken.

Q2580 **Rebecca Pow:** So, harmful basically? There is a harmful—

Chris Vickery: You could consider it that way. If you don't want your opponent to win, you might not see it as harmful. You would think of it as suppressing or dissuading, but whether or not that is harmful is up for interpretation. I personally see it as negative, bad, not good for society, but that is just my personal interpretation.

Q2581 **Rebecca Pow:** There was one other thing I wanted to follow on from that. In that same interview you mentioned that there was a separate Facebook-related incident that has not been reported yet, involving—

Chris Vickery: It has been reported; by now it has been.

Rebecca Pow: Yes. Is there anything else you wish to say about that or add to that?

Chris Vickery: That was the LocalBlox situation that you may have read UpGuard's report on. I have no reason to believe that it is related to the topic that we are discussing today, other than the fact that it involved Facebook data. It involved a bunch besides Facebook, LinkedIn and Twitter and Zillow. It was a large aggregation of data, but it was a good example of how somebody or a company can kind of abuse the Facebook searching, or at least before the changes that have been made recently.

Q2582 **Giles Watling:** You were talking a little while ago about companies claiming not to store data. Would it be fair to say that they can claim to have deleted data, not stored it, but in actual fact—the weasel words that you mentioned—they know how to access it any time?

Chris Vickery: It has been my experience within the past few years of data breach hunting that companies have very little incentive to lose access to data. Generally, they do not lose access to data. That is the rule of thumb. Data is not really deleted, destroyed, got rid of. It is either somebody's access to it is turned off but the administrator still has access to it, or it is archived on an offsite, or it is given to a partner. Shadow copies always seem to exist. It is the way it just works out.

Q2583 **Giles Watling:** It just lurks there somewhere in a cloud. I said a while back that it seems to me that data is the one thing that you can steal and leave where you found it, so it goes on that way. Then all that data, that really sensitive data, the phone numbers, social security numbers, credit card details I believe you can find out there, it is kind of still there. You find it and then it is left there. Do you find any evidence that somebody, some actor somewhere, is following you up and cleaning up, perhaps deleting and going down there, or is it actually still there?

Chris Vickery: The cycle of how things go in the research that I do at UpGuard is we find the data, we analyse the data to make sure that the data breach is not a false situation, or fake dummy data, and once we are satisfied that this is a situation that needs to be acted upon, we notify the company and make sure that they secure the data. There would not be somebody needing to come by after I have gone through that whole thing, because we make sure the company locks it down. We notify them and we keep reminding them or we elevate it to somebody above them, if necessary, to get it secured. We don't—

Q2584 **Giles Watling:** You are satisfied that that actually happens all the way down the line?

Chris Vickery: The one that we find gets closed up. Maybe they opened up other ones somewhere else that we were unaware of, but we make sure that the exposure that we find gets secured.

Q2585 **Giles Watling:** One of my earlier questions to you was that you are not alone in this. There are probably many hundreds, if not thousands, of other people doing the same kind of thing. When companies like AIQ and Cambridge Analytica access data, would they be able to see who else has accessed that data? Is there any way that that can happen?

Chris Vickery: Can we put it in the sense of: if company A is holding the data and company B accesses the data, can company B tell that others have accessed the data?

Giles Watling: Yes.

Chris Vickery: Not usually. There usually is not a public-facing log of who else is accessing it. Theoretically you could make something that shows that data, but I do not commonly come across the ability to see who else is accessing data, if you are the outside party accessing it.

Q2586 **Giles Watling:** Right. There are no breadcrumbs leading to the microwave?

Chris Vickery: You could theoretically do that. It would be possible, but I haven't come across that in an obvious sense. If you are alluding to ways of collaboration that are hidden or not obvious, there are ways you could do it, definitely, but there is not a log in front of you as soon as you find any data saying, "This person is also accessing it", just thrown at you.

Q2587 **Giles Watling:** You do not leave a shadow?

Chris Vickery: It is just not obvious in front of you but there are ways you could do that.

Q2588 **Ian C. Lucas:** It is pretty clear that the information that AIQ was collecting was being used for political campaigns. Would you agree with that?

Chris Vickery: Yes.

Q2589 **Ian C. Lucas:** In a referendum in the UK, different campaigns are forbidden from co-ordinating unless they declare their spending jointly. From your investigations, would you say the campaigns in favour of exiting the EU during the referendum were in fact co-ordinating through the work of AIQ?

Chris Vickery: Again, if I was on a jury in America and asked that question, I would conclude that, beyond a shadow of a doubt, there was some sort of collaboration going on. With all the evidence that I have seen, I do not think there is any ability for a reasonable, rational person to deny that there was some level of co-ordination or collaboration going on between the pro-Brexit campaigns.

Q2590 **Ian C. Lucas:** Why do you think that the EU referendum files were being uploaded on to the AIQ GitLab?

Chris Vickery: What do you mean by the referendum files—like the projects that were being worked on?

Ian C. Lucas: Also, by the different organisations, by Vote Leave, by DUP, and so on. Why do you think?

Chris Vickery: They were being uploaded to which area?

Ian C. Lucas: To the AIQ GitLab. Is that what happened?

Chris Vickery: I think there may be a small confusion here. I don't believe that, for example, Vote Leave and Change Britain had direct access to the GitLab repository. There was some relationship between the campaign and AggregateIQ, and AggregateIQ's developers were working on things for them but I do not think the campaign had direct access to this particular repository. They may have had access to other things, but this was a developer thing.

Q2591 **Chair:** Would it be fair to say the campaigns—going back to your earlier evidence—stocked the fridge but then it is up to the developers to use the ingredients?

Chris Vickery: For example, you could e-mail a list of people that have donated on your website, or whatever, when you start doing work with AIQ, and you could say, "I want to include this in our escrow database" or whatever and contribute data all sorts of different ways. There is no limit on the various ways you could contribute data, send files and synchronise and everything. AIQ could even provide tools to you, theoretically, that would allow you to synchronise your data totally from your end to their servers, and none of this needs the campaign to have direct access to the actual database itself. It is a little bit of a different concept than I think you were describing.

Q2592 **Ian C. Lucas:** I understand that process where, for example, a political party might have its own canvass returns, and employ a commercial organisation to target those people by supplying them with the information, but what I find difficult is that once they have done that and go away after the election or after the referendum or whatever, what happens to the information that has been passed over to the commercial organisation? It is very sensitive information that is owned by the political party. How do you ensure that that information is not then held or moved on to somebody else? What often happens—and I think there is evidence in your articles—is that information that was in the referendum, for example, becomes useful for subsequent political campaigns, such as the Conservative leadership campaign.

Chris Vickery: That is part of the problem. There is no way to make sure that goes away, and it has been my experience that it does not go away. That is exactly one of the big problems in this type of industry.

Q2593 **Ian C. Lucas:** It seems to me that at the moment our regulatory system cannot deal with that issue properly, because once the data is passed on it is out there and impossible to pull back.

Chris Vickery: That is very much the reality of the situation. Yes, data is sticky and flows out and does not oftentimes disappear. I think that is a reality we need to realise and figure out something to fix it.

Q2594 **Chair:** Facebook yesterday talked about this nuclear history button they want to introduce, but in some ways it is not worth very much, is it, because that old data, your old data trail, has probably been used by the people that want to use it already? They have derived tools from it. They have added you to their customer audience, and the fact that you can just delete that original data from Facebook really makes very little difference to anyone.

Chris Vickery: I want to start with saying that in the past one to two days I have been travelling and stuff, so I have not caught up on anything that Facebook said yesterday, for example, but from what you describe they have announced, yes, that does not seem like it would provide much real security. It would make you feel a little better perhaps if you were not a security or data professional and knew how things really work. But I can almost guarantee that data is going to exist in some way, shape or form continually. It is just the way this type of thing seems to work. It may not appear in your Facebook profile anymore, or be able to be looked up by an end user but, like you are describing, the modelling and the targeting and all that stuff has already been squeezed out of it.

Q2595 **Chair:** Yes. How do you think companies that work in this way can be fully compliant with the European GDPR rules?

Chris Vickery: I want to be clear that I am not an expert in GDPR stuff. I do know a bit about it, and UpGuard is stepping into those waters, but I think there is going to be a very hard-fought war between companies that deal with data and GDPR enforcers. There is not an easy fit. It is going to be very bloody.

Q2596 **Chair:** But one of the principles of GDPR is you can get your data back.

Chris Vickery: That is preposterous.

Q2597 **Chair:** Also, under existing European law, pre-GDPR, there are restrictions on who can hold data that contains information about your political beliefs or your religious interests or affiliation. From the process you have described—and we have heard a lot about during this inquiry—all this information has been blended together and is held by multiple private organisations, which can already be considered to be in breach of the law.

Chris Vickery: You would be astounded, shocked and awed by the amount of non-compliance with that requirement. This industry of big data does not care about laws where you can’t hold somebody’s religious preference. They don’t care about laws that restrict that. They are going to hold the data that they want to hold.

Q2598 **Chair:** Yes, and they believe there is not very much anyone can do about it.

Chris Vickery: That is the impression that I get, yes.

Q2599 **Chair:** The dataset that you have been referring to, and the copy of which you have with you, is that mainly based on elections work that has been done, or does it include information about other projects that these companies were involved with?

Chris Vickery: There are some non-political frameworks or information in there. There is reference to at least two different commercial advertising networks in there. I have found that one of the AIQ employees apparently was working with a commercial online advertising company at the same time that is headquartered in Switzerland and has sub-headquarters in Russia and the United States, so there are plenty of signs, little pointers and breadcrumbs, leading to commercial ventures.

Q2600 **Chair:** Can you say which two advertising networks you have in there?

Chris Vickery: I know we are having a private session after this. Could I tell you their exact names in the private session? The reason I say that is I have obtained a domain name that was at one time used by that group. I am gaining valuable information on it, as far as who is involved in this stuff, because there are autonomous systems still trying to connect to it, and if I say the name right now a lot of people are going to visit it and the data that I gather will be ruined.

Chair: I completely understand that reason. We can discuss it further in the private session.

Q2601 **Jo Stevens:** We have talked about the Trump campaign and we have talked about the referendum and elections elsewhere at the moment that have happened. We have had evidence from Christopher Wylie and from Brittany Kaiser about the work of SCL in foreign elections. Specifically Chris Wylie told us, and gave us some documentation, about a project that AIQ had done for the Minister of National Security of Trinidad. There was an email and there were some contract documents, and part of the project was to go out and find a way of accessing raw internet service provider data for the entire country and to monitor what people were browsing in that country.

You have written about the data you retrieved from the Trinidad and Tobago election. Could you describe the connections that you found between AIQ, SCL and Cambridge Analytica?

Chris Vickery: I do not think myself, or as part of UpGuard, we have done a large-scale report on the Trinidad-related files that are present. I did confirm on Twitter that there is a fairly extensive Trinidad and Tobago project on here that you will be very interested in reviewing, and there is private, what I would call PII—personally identifiable information—for a large number of people on this hard drive. Some questions will be answered once a real analysis of that particular project is done.

Q2602 **Jo Stevens:** When you say "large number of people", are you able to say roughly how much of the population of the country?

Chris Vickery: I wish I had spent more time looking into that one. My initial feelings on looking into that project was that there is so much other stuff here. To me, it is hard to get people to care about foreign countries that are not America and are not Brexit-related things. Usually when I deal with large finds like this I will put the foreign countries off to the back burner and get to them later on, and I have not done a real good, hardcore analysis of the Trinidad files. I have looked through it but I could not tell you exact numbers right now. That is not because I don't care about other countries. It is just because it is hard to get the public to care about that stuff.

Q2603 **Jo Stevens:** One other question on international influence campaigns involving these companies. Have you done any investigation into AIQ's role in the current Irish referendum on whether or not the eighth amendment will be repealed?

Chris Vickery: There is a project in here for the campaign, Save the 8th, so AIQ has been involved to some degree. I am not sure to what degree but it looks like at one point there was a website developed or they were working on developing a website for a campaign called Save the 8th, so there is some involvement there.

Q2604 **Jo Stevens:** That is as far as you have gone with it and what you know about it?

Chris Vickery: It is not an extensive project. It is the skeleton or maybe the frame of a website. I have no idea what kind of relationship they have with them or why they were related to that campaign or who is setting up AIQ with these clients. I don't know.

Q2605 **Chair:** On some of the other work, Brittany Kaiser told us that she was asked about whether she would be interested in working on some elections in other countries, which she declined to do. One was working with Marine Le Pen or instructions with a view to working for Marine Le Pen's party in France and also with AfD in Germany. Have you seen any information on the site that links AIQ or the other companies to those parties?

Chris Vickery: No, I haven't. To be honest, I am not an expert in European political situations, so there may be reference to those things on here that I did not catch that just went over my head but no, I haven't. To my knowledge, there is no French or other European countries' projects related to them on here. When I say "European", I mean what we have mentioned already, other than what we have covered.

Q2606 **Chair:** We touched earlier on cryptocurrencies. Could you say a bit more about the connections that you found between AIQ, Cambridge Analytica and SCL and cryptocurrencies and, in particular, the Midas token?

Chris Vickery: The Midas token is a little interesting. There is a project that was present on the AIQ GitLab that had the Midas token website, the initial coin offering website for it. The initial coin offering they like to think of it like an IPO, like a stock offering but it is for a cryptocurrency instead. It was a site where you could buy tokens for the Midas token and then, once it launches, you could have shares essentially of this Midas thing. It was interesting because on the site it had a $10,000 minimum buy-in, so we are talking about large-scale investors here. The public-facing live site disappeared shortly after I and UpGuard reported on the whole AIQ situation.

The Midas token site had launched. I was confused by the AIQ response when they were asked about it. I believe in the Canadian Parliament they said that it had not actually launched and it was a client of theirs that was looking into something. It had launched. The timer had gone down to zero, zero, zero, and I saw this on the live site and there is actually a live, at this very moment, Amazon S3 bucket holding a script that will query the Midas token status.

I believe I have a screenshot to show you—not right now but in the private session—that shows there have been tokens sold and ethereum transferred and cryptocurrencies are going through AIQ, at least the Midas token project. I don’t know how far this extends or what other cryptocurrencies they may be involved in. I believe Cambridge Analytica has admitted to being somewhat involved with something called Dragon Coin, which had to do with gambling and gaming on the other side of the planet. Does that answer your question?

There was a comment in the GitLab in one of the areas where employees can comment about making a specific badge for the Slack channel that was talking about Midas token. I haven’t seen the Slack channel but I know that they had a channel on Slack that was talking about it.

Q2607 **Chair:** Yes, I believe they have been asked—if it still exists—to provide chats from Slack to the Canadian Parliament.

Chris Vickery: Well, make sure that they provide all relevant Slack channels. They may provide the AggregateIQ one, but this Midas one was separate from the AggregateIQ one.

Q2608 **Chair:** Yes. What do you think the interest is in cryptocurrency? Do you think that is linked to their election work in any way?

Chris Vickery: What their interest is in cryptocurrency?

Chair: Yes.

Chris Vickery: I don’t want to speculate on their intent but I will go into what might be nice about cryptocurrency for groups like Cambridge Analytica, SCL and AggregateIQ. There is the ability—and I am not saying they did this—to transfer large amounts of funds very quickly in ways that are completely unregulated. That is obviously a very attractive way of operating with funds, if you are dealing with large amounts and you don't want to pay taxes or transfer fees or anything like that. You can also transfer data very easily, quickly and optionally in a permanent fashion. That way I believe Cambridge Analytica has said they were looking to create a data information-based cryptocurrency at some point down the road. There are a lot of different angles to the blockchain technology that would be very attractive, the money part being the most attractive probably.

Chair: Yes, indeed. We could all speculate as to why someone may want to do that.

Q2609 **Christian Matheson:** Chris, I want to ask you a couple of questions about Breitbart, if I may. I understand there is a reference within the GitLab files to Breitbart's scanning process and a reporting project. Does this ring any bells with you?

Chris Vickery: Yes, there is a project in there that has to do with Breitbart and Breitbart's advertising, Breitbart's users, scanning and scraping Breitbart and I believe that will probably lead down a few rabbit holes in the future as more minds can look into this. It seems that Breitbart was part of some campaign of either gathering messaging for advertisements or pulling messaging from Breitbart or giving it to Breitbart. I didn't see references to Breitbart UK but there is plenty of interaction with the Breitbart US-based site.

Q2610 **Christian Matheson:** You mention rabbit holes. What kind of applications might we see?

Chris Vickery: If you wanted to automate the process or create a process of feeding imagery and phrases and advertising messages to the conservative alt-right movement in America or anywhere else, Breitbart would be a wonderful provider of information, imagery, messaging that resonates with people. It would also give you the ability to pull commentary and sentiment and further refine your messaging, see what works and what doesn't. If you are working directly with Breitbart—Cambridge Analytica I believe has claimed at one point they were a distributor of Breitbart information—you have access to all the user analytics, the visitation analytics, the interaction, the users themselves. You have a wealth of information to pull from that resonates very quickly and easily with the alt-right movement.

Q2611 **Christian Matheson:** Yes. There was an e-mail that Oczkowski, Head of Product at Cambridge Analytica, sent in January stating that they were the exclusive seller of Breitbart data.

Chris Vickery: Well, it may very well be the case that you will see the skeleton or basic framework of how that data is being gathered, or interact with their process in the projects that are on this hard drive. I would not be surprised if it turns out to be something that developed into what they are using today.

Q2612 **Christian Matheson:** There was a reference in that email to DMP tags. Do you have any idea what they are, what that means?

Chris Vickery: Data management platform tags probably. That is probably what it is, and tagging can be used in all sorts of ways but generally through track users, track topics, track key words, see who is interacting with what and what sentiment and spreading it and what goes where. It is a very good way of figuring out what works and what does not and who you should be targeting.

Q2613 **Giles Watling:** One last little thing, and I know you have been travelling for a couple of days and you have my sympathy. There is this article on Recode that Facebook will allow users to opt out of letting Facebook collect their browsing history. It says it's "arguably the company's biggest update since the Cambridge Analytica scandal broke". It is sort of reactive, isn't it? If there is a history button you can say, "I just want to delete my history". In your opinion, is there any way that you can make that proactive so that you can opt out before, so as you post or whatever you do on Facebook, it is deleted straightaway, so there is no way that that data could be stored? Is that possible?

Chris Vickery: Facebook could not gather it in the first place because that is not something that, just by default, they have to have in order to operate just technically speaking. Then there is the option of you could have browsers themselves somehow not hand that data over. By browsing history, do you mean specifically browsing through Facebook or just browsing, any history?

Q2614 **Giles Watling:** I am using Facebook at this moment but browsing generally, yes.

Chris Vickery: In general, okay. Theoretically you could do it. There is no incentive for companies to do it, though, which is the biggest stumbling block because they gain great analytical understanding of user habits, behaviours, opinions and learn how to hone their platform and be more profitable off that data. I think the biggest block you will run into is there being no incentive for companies to do that.

Q2615 **Giles Watling:** Would it be fair to say you think that this much vaunted history delete button that Facebook is introducing, as of yesterday, is just show casing, it has no value?

Chris Vickery: It is kind of theatre in the way that you are describing it. I would say it would give people peace of mind and help some less tech savvy people sleep better at night maybe, but as far as actual removing any sort of risk of personal information from getting further out there—

Giles Watling: Someone like you coming along with—

Chris Vickery: I don't know if somebody like me would be able to take advantage of it but advertisers and everybody. I still think there will be ways that the data that advertisers want to pay Facebook for will continue to exist in one way or another.

Q2616 **Jo Stevens:** There was a huge data breach in 2016 from Uber. When you have looked at this repository that you have described, have you ever seen any Uber data in there?

Chris Vickery: The only time I saw the word "Uber" exist in there—and I am not saying this is related to Uber itself—is reference to a script or system called "Uber auth". I don't know if they were experimenting with Uber auth—I am assuming "auth" means authentication—for any particular reason and I don't know if that is tied in any way to Uber the company. That is just the only time I saw the word "Uber" in any of the scripts, so that is the only little connection there.

Chair: That is great. Thank you very much for your public evidence. We are now going to move into a private session of the Committee, so I thank the people who have joined us but we now need to go into private session.