

Mr Julian Knight MP
Chair
Digital, Culture and Media and Sport Select Committee
House of Commons
London
SW1A 0AA

Our reference: ICO/O/ED/L/RTL/0181

02 October 2020

Dear Mr Knight,

RE: ICO investigation into use of personal information and political influence.

1. Thank you for your letter of 10 September 2020 asking for an update on my office's investigation into the use of personal data for political purposes that was launched in 2017. This follows my last evidence to the predecessor Committee's sub-committee on Disinformation in April 2019.
2. Throughout this investigation, we have sought to keep the Committee informed of key developments and findings, having produced three written reports, the last being in November 2018. The investigation has been one of the largest and most complex ever carried out by a data protection authority and it is therefore right that Parliament is able to properly scrutinise the evidence we have uncovered and the actions we have taken as a result. The investigation has provided new understanding about the use of personal data in the modern political context and has transformed the way data protection authorities around the world regulate data use for political purposes. Where there was evidence of breaches of the law, we have acted. And where we have found no evidence of illegalities, we have shared this openly. This further work confirms my earlier conclusion that there are systemic vulnerabilities in our democratic systems.
3. Since my last appearance before the Committee in April 2019 my office has continued its investigative work, completing the remaining lines of enquiry as far as the evidence took us. This included analysis of materials obtained during the investigation and those seized under warrant. This has, overall, confirmed and reinforced the findings of my previous reports. I have therefore concluded that there is little in the vast volumes of evidence we have now worked through that has changed our initial understanding or identified new lines of enquiry that suggest they could drive new insight.

4. The investigation is therefore concluding, and the following letter and Annexes acts as our final written account to Parliament. It provides a summary of the conclusions we have drawn from our analysis of the evidence in the final stages of our investigation, the additional actions we have taken and why, and broader learning we and other data protection authorities can draw on to inform future investigations and regulatory work in the digital era. In addition, Annex 1 provides the Committee with detailed answers to the specific questions asked by the Committee. Annex 2 provides a deep dive into how SCL Elections / Cambridge Analytica used the personal data it held, whether these methods could be used in the future, and the associated risks to citizens.

Findings since April 2019

Outstanding areas relating to processing of data by SCL Elections Ltd and Cambridge Analytica (SCL/CA)

5. Detail of the data processing practices undertaken by SCL/CA is set out at Annex 2, but, in summary, we concluded that SCL/CA were purchasing significant volumes of commercially available personal data (at one estimate over 130 billion data points), in the main about millions of US voters, to combine it with the Facebook derived insight information they had obtained from an academic at Cambridge University, Dr Aleksandr Kogan, and elsewhere. In the main their models were also built from 'off the shelf' analytical tools and there was evidence that their own staff were concerned about some of the public statements the leadership of the company were making about their impact and influence.
6. I have also confirmed my previous understanding about the poor data practices at the company, which, had they sought to continue trading, would likely have attracted further regulatory action against them by my office. I found excerpts of what appears to be examples of the data obtained by Dr Kogan and his company Global Science Research (GSR) from the Facebook platform at various stages of its processing.
7. From my review of the materials recovered by the investigation I have found no further evidence to change my earlier view that SCL/CA were not involved in the EU referendum campaign in the UK - beyond some initial enquiries made by SCL/CA in relation to UKIP data in the early stages of the referendum process. This strand of work does not appear to have then been taken forward by SCL/CA.

Investigation into the data practices of organisations on both sides of the EU referendum campaign

8. I have concluded my wider investigations of several organisations on both the remain and the leave side of the UK's referendum about membership of the EU. I identified no significant breaches of the privacy and electronic marketing regulations and data protection legislation that met the threshold for formal regulatory action. Where the organisation continued in operation, I have provided advice and guidance to support better future compliance with the rules.

Evidence of Russian involvement

9. During the investigation concerns about possible Russian interference in elections globally came to the fore. As I explained to the sub-committee in April 2019, I referred details of reported possible Russia-located activity to access data linked to the investigation to the National Crime Agency. These matters fall outside the remit of the ICO. We did not find any additional evidence of Russian involvement in our analysis of material contained in the SCL / CA servers we obtained.

Securing the data obtained by Dr Kogan and GSR

10. There was concern that data and derived data from Facebook had been shared outside of GSR and SCL/CA. My investigation found data in a variety of locations, with little thought for effective security measures, which appeared to have come from GSR and SCL/CA. We found that individuals of interest to the investigation held data on various Gmail accounts. Data was also found in servers and appeared to have been shared with a range of parties, for example there was evidence that data had been shared with staff at SCL/CA, Eunoia Technologies Inc, the University of Cambridge and the University of Toronto.
11. Some of the individuals who worked for these organisations used their personal email accounts for work purposes. However, the data itself was sometimes shared using secure drop/file sharing sites. It was not always possible to identify if all this data was from GSR/Dr Kogan and derived from the app he built to gain access to Facebook data which he called *thisisyourdigitallife*. We also identified evidence that in its latter stages SCL /CA was drawing up plans to relocate its data offshore to avoid regulatory scrutiny by ICO. We have followed up their complex company structure with overseas counterparts and have concluded that while plans were drawn up, the company was unable to put them into effect before it ceased trading. We

have required those we contacted during the investigation to certify deletion of the data they held.

Action taken and follow up since April 2019

12. In our written update to Parliament in November 2018 and our oral evidence session in April 2019 we reported several actions we had taken against organisations for breaches of the law.
13. The following organisations have now paid the penalty notices levied on them:
 - Facebook (£500,000) paid 04 November 2019
 - Vote Leave (£40,000) paid 29 April 2019
 - Leave.EU (£15,000) paid 15 May 2019
 - Emma's Diary (£140,000) paid 29 August 2018
14. In addition, we successfully prosecuted SCL Elections for their failure to comply with my Enforcement Notice. We fined them £18,000.
15. My office also made a referral to the Insolvency Service about various conduct issues within the SCL and its group of companies. We worked together and shared relevant information and intelligence with the Insolvency Service arising from our investigation. Mr Alexander Nix, a Director of SCL Elections Ltd, is now disqualified from acting as a director for a period of seven years.

Appeals of my notices to the First Tier Tribunal

16. As the Committee will be aware, my actions are subject to judicial oversight by the First Tier Tribunal (General Regulatory Chamber). Appeals were made against my decision to issue the Liberal Democrats with an Assessment Notice (a formal notice allowing my office to audit an organisation's compliance with data protection legislation). UKIP similarly appealed an Information Notice (a formal notice requiring provision of information to my office) I had served upon them. Eldon Insurance (trading as GoSkippy) and Leave.EU also appealed their Assessment Notices, and some of the Monetary Penalty Notices. The First Tier Tribunal has dismissed all these appeals. I have therefore been able to advance the audits of the Liberal Democrats' and UKIP's compliance. Eldon Insurance and Leave.EU have further appealed to the Upper Tribunal but subject to the outcome of the appeal and COVID-19 restrictions, it remains my intention to complete the audits as soon as is practicable. Facebook also appealed the Monetary Penalty Notice served on

them. However, their appeal was withdrawn based on a settlement agreement. Facebook paid the full monetary penalty.

Audits of organisations involved in supply and use of personal data for political purposes.

17. My audit teams have also concluded audits of data protection compliance at 14 organisations associated with the original investigation, including: the main political parties, the main credit reference agencies and major data brokers, as well as Cambridge University's Psychometrics Centre. We have made significant recommendations for changes to comply with data protection legislation.

Closing the investigation and follow up

18. In accordance with the terms of the search warrants, I have started the return of materials to SCL's administrators. Where necessary, my team have ensured that any data, models and derivatives are safely destroyed. Several items obtained have been subsequently disowned and we are taking measures via our forensic technology provider to destroy these safely ourselves.
19. A small number of follow up enquiries remain, and these will be taken forward as business as usual over the coming months. Subsequent complaints or issues about political use of personal information in other political campaigns are being triaged and investigated in line with my Regulatory Action Policy.
20. It should also be noted that we will shortly be publishing the reports of our findings of our audits of the main political parties, the main credit reference agencies and major data brokers, as well as Cambridge University Psychometrics Centre. We will write separately to the Committee on those issues.

Wider impact of the investigation and conclusion.

21. This has been a complex and wide-ranging data protection investigation, touching on some of the most contentious and widely debated issues of recent times. At all times we have sought to follow the data and being transparent in our methodology and findings and acting only where there was a public interest to do so. We are continuing to work to address the systemic vulnerabilities we identified, working alongside other agencies.

22. What is clear is that the use of digital campaign techniques are a permanent fixture of our elections and the wider democratic process and will only continue to grow in the future. The COVID-19 pandemic is only likely to accelerate this process as political parties and campaigns seek to engage with voters in a safe and socially distanced way.
23. I have always been clear that these are positive developments. New technologies enable political parties and others to engage with a broad range of communities and hard to reach groups in a way that cannot be done through traditional campaigning methods alone. But for this to be successful, citizens need to have trust in how their data is being used to engage with them.
24. I believe that the findings of my investigation and the work we have done with the political parties through the audits has led to improvements to data handling across the political parties in the UK (which will be detailed in my audit report).
25. Much of the learning from this investigation was applied in the recent UK election, in which my office scrutinised political campaigning groups, tactical voting apps and the actions of individuals or political parties. The investigation led to extensive cooperation from a variety of social media platforms and collaboration with the Electoral Commission. This resulted in advice being provided to five data controllers to improve their compliance with the legislation during the election.
26. A final version of the updated political parties guidance that was published in draft before the general election, will be issued in the near future and will support political parties to use data protection legislation as an enabler to the transparent and lawful use of personal data in political campaigns as new techniques continue to come on board.
27. The impact of this investigation has also had international reach. I have been asked to brief parliaments and governments across the world and I have shared the learning from this investigation with election oversight and privacy regulators internationally. The prominence of the use of personal data in political influence has grown significantly, and several international counterparts have since undertaken similar work, as is appropriate to safeguard their national democratic structures.
28. A number of parallel international investigations of these issues have also concluded, including those in Canada, at which point the deletion of UK data held by AggregateIQ (AIQ - a company associated with SCL/CA) and covered by my Enforcement Notice on the company has been confirmed to

us. Facebook have also been investigated by several other international data protection authorities including those in Australia, Canada, the United States and others. These agencies have all reached conclusions consistent with the ICO's in their findings. My office was able to cooperate with these authorities to support their own investigations.

29. The scale of the investigation I conducted was unprecedented for a data protection authority. It highlighted the whole ecosystem of personal data in political campaigns. I believe that citizens are better informed as a result and policymakers are alive to the risks of data misuse. It has led to improvements in oversight arrangements and changes in operating practices of the major technology platforms.
30. In the UK, the major political parties have engaged positively in programmes of improvement to their data protection practices. The investigative and operational learnings together with the evidential insight we obtained have been shared with my international counterparts. This had led to a greater oversight of their respective election processes and mechanisms. This investigation showed, the value of international cooperation between authorities facing common threats. This is particularly relevant in the context of the UK's position post transition period 31 December 2020.
31. The investigation has also helped improve the investigative approach of my office and I have established a high priority investigations team as a result. I hope this will mean my office will have the standing capacity and capability to progress such complex investigations more easily in future.

Yours sincerely,



Elizabeth Denham CBE
UK Information Commissioner

Annex 1: Update to questions from the sub-committee on disinformation hearing on my work on 23 April 2019.

1. During the April 2019 hearing there were several questions which required further detail to be checked against the evidence in the investigation and I said we would report back to you about these. Below, you will find the responses to all the outstanding questions from these previous hearing, which I hope is helpful to the Committee.
2. The outstanding questions are bullet pointed below, complete with the ICO's response at the time and our update. All references are from Hansard April 2019.
3. The sub-committee have previously asked;
 - Is there any evidence that you are aware of that pre-presented datasets were used by AIQ in delivering advertisements through Facebook? [Q12]

Our Response: *We confirmed that we would need to check on this point.*

Update: To confirm; whilst there was evidence in some cases of using pre-presented datasets, this was dependent on the request of the client and type of campaign.

For example, one of the website custom audiences was named "Vote Leave Instapage Submissions". This was created based on visitors to www.voteleavetakecontrol.org.

AIQ used different methods of targeting for different campaigns. Some campaigns used Facebook's standard targeting tools to target users by age, location, gender and interests while others used datasets provided by the campaigns themselves to create lookalike audiences using Facebook's standard functionality at the time.

- Is it right, for example, that Vote Leave would present data to AIQ and they would then use Facebook as a method of dispersing messages through that dataset? Is that how it worked? [Q13]

Our Response: *We confirmed that we would need to check on this point.*

Update: To confirm my investigation found that Vote Leave provided personal data to AIQ. This data was used by AIQ to create lookalike

audiences on Facebook, using the standard Facebook processes available at the time.

- Did you find any evidence of datasets from one organisation being used by AIQ on behalf of another organisation to disseminate information through Facebook? [Q14]

Our response: *We looked at the sharing of those datasets and I do not think we found that kind of sharing, but I will double-check the file.*

Update: Further to our initial response, No. We investigated whether AIQ had used the same datasets to target adverts on behalf of Vote Leave, BeLeave, the DUP and Veterans for Britain. Initial information provided by Facebook had suggested that there were three audiences that were used for targeting by both Vote Leave and BeLeave. However, AIQ subsequently clarified that this was an admin error made by a junior member of staff while creating the BeLeave account. The error was corrected the following day and no information from those campaigns was disseminated through Facebook in the form of targeted ads.

- How was the information disseminated through Facebook? Was it only through datasets that were presented by one organisation? For example, would Vote Leave disseminate information only through a dataset that they provided? [Q15]

Our response: *Potentially, yes.*

Update: Further to this response, our investigation found that AIQ's own internal firewall policy prohibited the sharing of data between campaigns. We have not found any evidence to suggest that any personal data was shared between Vote Leave, BeLeave, the DUP or Veterans for Britain beyond the error by a staff member identified above. Therefore, our earlier answer is correct.

- If there was dissemination through a dataset presented, for example, by the DUP, that would be a data breach. Is that right? [Q16]

Our response: *Potentially, depending on the circumstances of the dataset.*

Update: Further to this response, the answer provided to you at the time is unchanged.

- And that is the evidence that you do not think you have now? [Q17]

Our response: *Yes, but I will double-check.*

Update: To confirm - we have not discovered any evidence to support that such data sharing occurred.

- Can you explain what would be the benefit of using a single company such as AIQ for different organisations seeking to disseminate information through Facebook? Why were all these businesses using AIQ? [Q18]

Our response: *In our inquiry we have not looked at the motivation behind that. Obviously, if somebody were particularly good at the work they did, that might be an incentive for them to be marketing their services to different parties, but the motivation behind why people placed particular contracts was not the focus of our inquiry—it was the basis on which that information was consented to be passed on.*

Update: Our position on this question remains unchanged. No further evidence that speaks to motives was uncovered during the investigation. However, we understand that the Facebook criteria for audience targeting varied from project to project and will have been informed by AIQ who placed the social media adverts. For example, voters were split into categories of persuadability and targeted on this basis (rather than necessarily by a discrete characteristic or criteria on Facebook).

4. I hope that these final points of clarification are helpful.
5. Additionally, I also refer to your question (Q20/21) over whether the ICO has sufficient powers to be able to establish what is going on in, for example, a closed Facebook group. We continually review the value and effect of our powers, particularly in the face of new and emerging technology. For now, the ICO can investigate and enforce whenever personal data is put at risk or misused.

Annex 2: Reporting back on the activity undertaken by SCL Elections and Cambridge Analytica

1. At the sub-committee hearings and in my earlier reports I explained that we were working through a considerable amount of electronic materials seized in searches and uncovered by the investigation to understand how data was handled by the parties involved. This included information received from other regulators and provided voluntarily by a number of parties including materials provided by Cambridge University, ex-Cambridge Analytica staff and their associates, materials from GSR and others connected with Dr Kogan and his studies at Cambridge University, as well as that provided by some of those directly involved in these matters when interviewed. Several senior figures have continued to maintain their silence and have declined to be interviewed.

Our approach and context

2. Since the last hearing the ICO has conducted a reverse engineering exercise to try to identify and confirm as far as possible, how SCL/CA processed the personal data they held. The primary aim of this exercise was to understand how personal data was processed and to determine whether the method used could be repeated and if so, the risks posed to data subjects. Whilst there was a technical aspect to this work my findings were also informed and corroborated based on accounts obtained from witness interviews and the contents of statements taken during the investigation.
3. During my investigation a large amount of material and equipment was reviewed including;
 - 42 laptops and computers,
 - 700 TB of data,
 - 31 servers,
 - over 300,000 documents, and
 - a wide range of material in paper form and from cloud storage devices.
4. Several the devices seized were encrypted or had been damaged or contained anonymised or pseudonymised data. The structure and pattern of material recovered confirmed the situation we have previously reported on at the time of the initial reports; there were a number of poor information governance practices within SCL/CA that meant personal data was not always organised or well-structured, or accurate records of processing kept.
5. In addition, SCL/CA Staff seemed to work interchangeably across several different email accounts. This seemed to be the company's ordinary

operating model and ordinary course of business rather than anything designed or intended to throw the ICO off the trail; evidence from email accounts attested to this; showing staff trying to establish if they had deleted the Facebook data and its derivatives and deal with the publicity the company was under at the time. The director of SCL Elections at that time was Mr Alexander Nix. Cambridge Analytica LLC was a subsidiary of SCL, with "Cambridge Analytica" serving as the brand under which the SCL group of companies predominantly operated. We have referred to SCL/CA in this document, save where it makes a material difference.

6. The sheer volume of material seized meant that we were presented with a digital 'haystack' of information in various states and locations and this has prolonged the work involved in reviewing and assessing the material to help us understand what happened. However, by piecing together the timeline of events we were able to get a thorough evidential insight into what was likely to have taken place.
7. We have used the material we could recover and access, to try and work backwards, over a timeline of many years, to understand the way data was gathered, stored, processed, combined and then used. We have focussed (given the volumes involved and not withstanding SCL/ CA's work for commercial clients) on political uses of data linked to Dr Kogan's work and GSR. As we have gone about this we have tried to match the digital work to other known records, statements and accounts already reported on by ourselves and others, including examples of data which have been presented to us, as examples of the data from GSR, at various stages of its development within the approach taken by SCL/CA.
8. We have examined emails and contracts between the key parties, financial information, data sharing agreements and invoices, publicity brochures, research papers, models, data sets and examples of code. By tracking the development of some of these sources of information we have gained insight of how SCL/CA's approach developed over time, and some pointers to how it was proposed to develop further.
9. The conclusion of this work demonstrated that SCL were aggregating datasets from several commercial sources to make predictions on personal data for political alliance purposes. For example, we recovered data which included Voter files (the US version of the Electoral Register), Consumer Data Sets, Social Media and Intelligence Data Sets that appeared to come from the following companies: Labels & Lists, InfoGroup, Aristotle, Magellan, Acxiom and Experian. Some data has the appearance of similar US voter data that has been subject to known cyber breaches and has been available on-line.

10. SCL's own marketing material claimed they had "Over 5,000 data points per individual on 230 million adult Americans." However, based on what we found it appears that this may have been an exaggeration.
11. Although we do not have a list of all the datasets, during the document review we discovered evidence that some of the data sets as at September 2015 included:
 - Nationwide voter files from L2 (meaning "Labels and Lists") and DataTrust (~50 data points for 160M individuals)
 - Nationwide consumer data from Acxiom and Infogroup (~500 data points for 160M individuals)
 - Election return results from Magellan (~20 data points for national census tracks)
 - Nationwide consumer data from DataTrust (3000 data points for 100M individuals)
 - Psychographic inventories (10 data points for 30M individuals)
 - Facebook social network (graph database containing 30M individuals)
 - Facebook likes (570 data points for 30M individuals)
 - In-depth Republican Primary focused surveys (80k)
 - ForAmerica member data (14.6M post comments, 240M post likes across 31 M users)
 - Emails from Infogroup (30M)
 - Emails from DataTrust (26M)
12. In short, the number of data points varied considerably, both from individual to individual and from one project to the next.
13. It appears that the company also had a variety of sources of data that were commercially acquired, on mainly what appeared to be US citizens.

Dr Kogan's app and SCL

14. In respect of Dr Kogan's application, which he called *thisisyourdigitallife* (the App), the material obtained in the evidence review corroborated our understanding as set out in our previous reports that it obtained data from individuals who authorised it to access their Facebook data. However, the App functioned in a way which meant that it was also able to obtain the data of that user's Facebook 'friends' (who had not themselves restricted such sharing through their own Facebook 'privacy controls'). In conjunction with the personality quiz function of the App, along with a record of each user's 'likes' information, Dr Kogan was able to model personality traits for users of the App, and for their Facebook 'friends'. This approach seeming built on

earlier work by Dr Kogan involving Facebook 'likes' and personality scores. Dr Kogan set up a new company, GSR, this was established and funded for the primary purpose of acting as a vehicle for the provision of the services anticipated under the contract between GSR and SCL / CA.

15. As we have explained in our earlier reports, in April 2014, Facebook introduced changes to their platform which reduced the ability of apps to access information about users, and about the Facebook friends of those users.
16. On 6 May 2014, Dr Kogan applied for extended permissions to access Facebook user data for research purposes beyond May 2015. Facebook rejected this application on the basis that the request would be in breach of Facebook's terms of service. Facebook did not at this time remove the App's access to the Facebook Platform, and therefore the App operated throughout the grace period. Dr Kogan and/or GSR continued to utilise the App through the Facebook Platform to harvest data of Facebook users for commercial purposes.
17. On or around 4 June 2014, GSR and SCL Elections Limited signed a contract pursuant to which data harvested by GSR through the App (or modelled data derived therefrom) would be sold to SCL/CA.
18. Dr Kogan/GSR subsequently shared subsets of the data harvested by the App (or at least modelled data) with Eunoia Technologies Inc, University of Cambridge, University of Toronto and SCL/CA. The data shared with SCL and Eunoia Technologies related eventually to approximately 30 million US registered voters, albeit it started with 4 'waves' of data covering some 2.1 million voters in autumn 2014. At least some of the shared data (or modelled data) is understood to have subsequently been used in connection with political campaigning, including (it is suspected) the 2016 US presidential election. For example, it is understood SCL (through contracts with firms including AIQ) deployed advertising on the Facebook Platform which was targeted to specific voter demographics informed by the profiling that had been undertaken by SCL/CA and GSR.
19. It was suggested that some of the data was utilised for political campaigning associated with the Brexit Referendum. However, our view on review of the evidence is that the data from GSR could not have been used in the Brexit Referendum as the data shared with SCL/Cambridge Analytica by Dr Kogan related to US registered voters. There was evidence of considerable focus in the data collection and data matching processes between GSR and SCL on US voters, as this was what was to be paid for under the contract(s) between them. Cambridge Analytica did appear to do a limited amount of

work for Leave.EU but this involved the analysis of UKIP membership data rather than data obtained from Facebook or GSR. Some evidence was recovered however that suggested an intention by SCL / GSR to target UK voters in 2014 through the same process. This work does not appear however to have been taken forward.

20. The App was however used by some 300,000 Facebook users worldwide. Since the App was able to collect data about the Facebook friends of its users, the total number of individuals about whom the App collected personal data has been estimated by Facebook as being up to 87 million worldwide. The number of UK Facebook users who used the App has been stated by Facebook to be 1,040 (though Facebook have also stated that 1,765 individuals in the UK used the App). The total number of UK Facebook users about whom the App collected personal data has been estimated by Facebook as at least 1 million.

Deletion of data

21. On or around 3 April 2017, Alexander Nix provided a signed certificate ("Deletion Certificate") to Facebook on behalf of SCL stating that *"all Facebook data gathered by the "thisisyourdigitallife" Facebook Application ...received from or on behalf of GSR or Dr. Aleksandr Kogan, including but not limited to Facebook user data and Facebook user friend data has been accounted for and permanently deleted and destroyed from both active and redundant storage ..."*.
22. Our review of internal email traffic and interviews with former SCL employees suggest that keyword searches were conducted on the servers in early 2016 to locate and delete the data received from GSR. We established that in April 2017, around the time Alexander Nix signed the deletion certificate to Facebook, SCL/CA employees used specific scripts to delete additional data in linked databases and backup files. This included the 'kogan_import' database and data stored in AWS. There was evidence recovered however that as the company came under increasing scrutiny there was confusion about the quality and effectiveness of the deletion process within the SCL/CA staff group.

AIQ

23. In early 2014, SCL/CA commissioned Aggregate IQ ("AIQ"), a Canadian based company, to build a Customer Relationship Management (CRM) tool for use during the American 2014 midterm elections. SCL called the tool RIPON. It was designed to help political campaigns with typical campaign activity such as door to door, telephone and email canvassing. In October 2014, AIQ also placed online advertisements (including on the Facebook Platform) for SCL on behalf of its clients.
24. AIQ worked with SCL on a similar software development, during the US presidential primaries between 2015 and 2016. AIQ have also confirmed it was directly approached by Mr Wylie when he was employed at SCL. AIQ has advised that all its work was conducted with SCL and not CA.
25. We understand from witness evidence that AIQ played a significant role in the deployment of targeted advertisement, leveraging their expertise in this digital marketing in order to assist SCL. There was a range of evidence that demonstrated a very close relationship between AIQ and SCL (such as evidence that described AIQ as the Canadian branch of SCL and evidence that Facebook invoices to AIQ for advertising were paid directly by SCL). However, AIQ has consistently denied having a closer relationship beyond that between a software developer and their client. Mr Silvester (a director/owner of AIQ) has stated that in 2014 SCL 'asked us to create SCL Canada but we declined'.

Methods utilised by SCL/CA

26. On examination, the methods that SCL were using were, in the main, well recognised processes using commonly available technology. For example, open source data science libraries such as 'scikit' were downloaded by SCL – containing well established, widely used algorithms for data visualisation, analysis and predictive modelling. It was these third-party libraries which formed the majority of SCL's data science activities which were observed by the ICO. Using these libraries, SCL tested multiple different machine learning model architectures, activation functions and optimisers (all of which come pre-developed within the third-party libraries) to determine which combinations produced the most accurate predictions on any given dataset. We understand this procedure is well established within the wider data science community, and in our view does not show any proprietary technology, or processes, within SCL's work.
27. However, it is important to stress that the output was only a prediction; and while the models showed some success in correctly predicting attributes on

individuals whose data was used in the training of the model, the real-world accuracy of these predictions – when used on new individuals whose data had not been used in the generating of the models – was likely much lower. Through the ICO's analysis of internal company communications, the investigation identified there was a degree of scepticism within SCL as to the accuracy or reliability of the processing being undertaken. There appeared to be concern internally about the external messaging when set against the reality of their processing.

28. My investigation found that the data transferred to SCL by GSR was incorporated into the pre-existing larger database already held by SCL which held voter file, demographic and consumer data for US individuals.
29. The data points collected by GSR with respect to survey users and their Facebook 'friends' was specifically selected to enable a 'matching' process against pre-existing SCL databases. Matching took place using file sharing platforms and by reference to name, date of birth and location – with SCL's existing datafiles being 'enriched' and supplemented by GSR's data about those same individuals – and this matched information being passed back into SCL systems. This resulted for example information including scores for voting frequency, whether likely republican or democrat, voting consistency, and a profile which predicted personality traits matched to information such as voter ID, name, address, age, and other commercial data.
30. Through such processes the relevant US voter GSR data (about approx. 30 million individuals) was then further analysed using machine learning algorithms to create additional "predicted" scores relating to partisanship and other criteria which were then applied to all the individuals in the database. Some of these focussed on likes as wide ranging as "gay rights", "Obama the worst president in US history", "Re-elect President Obama in 2012", "the Bible" and "National Rifle Association". These scores were used to identify clusters of similar individuals who could be potentially targeted with advertising relating to political campaigns. This targeted advertising was ultimately likely the final purpose of the data gathering but whether or which specific data from GSR was then used in any specific part of campaign has not been possible to determine from the digital evidence reviewed. There is however evidence recovered that suggests that similar approaches and models based on the predicted personality traits and other measures were used with Republican National Committee (RNC) data.

Further development of the approach

31. Although not a primary focus of this work the evidence review identified evidence that suggested that SCL were keen to further develop their capacities. This included seeking as much detail from GSR about the 30 million voters so they could supplement the material with their own data scraping exercise. There was also evidence of discussions into 2015 to replicate the survey-based work undertaken by the App and therefore to obtain the data used to train the models themselves so SCL could build their own arrangement.