scripted

Volume 14, Issue 2, December 2017

Law Enforcement in the Age of Big Data and Surveillance Intermediaries: Transparency Challenges

*Teresa Scassa**

DOI: 10.2966/scrip.140217.239

Abstract

In October 2016 Geofeedia made the news when it was reported that police services in North America had contracted with it for data analytics based on georeferenced information posted to social media websites such as Twitter and Facebook. Geofeedia is not the only data analytics company to mine social media data and to market its services to government authorities. These activities raise important issues around the transparency of state surveillance activities, as well as the targeting of protesters exercising their constitutional rights to free speech. This paper examines how the public sector reliance on purchased georeferenced data and analytics changes the dynamics of transparency of government action and calls for new measures and approaches.

Keywords

social media; surveillance; georeferenced data; transparency; data analytics

* Canada Research Chair in Information Law and Policy, Faculty of Law, University of Ottawa, Ottawa, Canada, tscassa@uottawa.ca

1 Introduction

People around the world share an enormous volume of information on a daily basis across multiple social media platforms. Some of this information is of interest to law enforcement officials, and police have learned to use social media to gather information about suspects, witnesses or events, in order to solve crimes and to prevent planned criminal activity.[1] Although police are able to access non-public social media data through the use of search warrants, in many cases the information they seek is posted in public forums, where no search warrant is required for access.

As we enter the big data era, the nature and scale of police use of social media is changing. Aided by start-ups offering data analytics tailored to law enforcement needs, police services have begun experimenting with social media analytics for the purposes of surveillance, profiling and predictive analytics. These activities raise new privacy and social justice issues. They also raise important issues of transparency and accountability. In doing so, they highlight the growing overlap between privacy and other "disciplines" of law. This paper examines these issues in the context of disclosures in the fall of 2016 that revealed the widespread use of social media data analytics by police services in the United States. It identifies the privacy and other social justice issues raised by these activities, and examines both normative and transparency frameworks to govern the use of such technologies.

2 The Geofeedia example

In October 2016, Geofeedia, an internet company that described itself as a

[1] See Alexandra Mateescu et al., "Social Media Surveillance and Law Enforcement" (Data and Society, 27 October 2015), available at http://www.datacivilrights.org/pubs/2015-1027/Social_Media_Surveillance_and_Law_Enforcement.pdf (accessed 12 June 2017).

"location-based analytics platform," drew significant media attention after the American Civil Liberties Union of California (ACLU) issued a news release.[2] It documented how Geofeedia entered into contracts with police services across the US to provide data analytics based on georeferenced information posted to social media websites such as Twitter, Instagram and Facebook. The ACLU based its report on the results obtained from a series of freedom of information requests that targeted sixty-three police services across California. The responses they received included email exchanges between representatives of Geofeedia and different police services. These emails painted a picture of a company that marketed its service to law enforcement officials as a means to, among other things, monitor activists and protestors.

The ACLU report had a swift and significant impact. Social media companies, whose business models are entirely dependent upon user-generated content and the tracking of user activities, were extremely unwilling to be seen providing ready access to user data for broad-based law enforcement surveillance activities. Following the release of the ACLU report, Instagram and Facebook terminated Geofeedia's access to their APIs for public user posts,[3] and Twitter announced that it was cutting off Geofeedia's access to its data immediately.[4]

---

2 Matt Cagle, "Facebook, Instagram and Twitter Provided Data Access for a Surveillance Product Marketed to Target Activists of Color" (American Civil Liberties Union of Northern California, 11 October 2016), available at https://www.aclunc.org/blog/facebook-instagram-and-twitter-provided-data-access-surveillance-product-marketed-target (accessed 12 May 2017).

3 Access to Instagram's feeds was terminated on 19 September 2016. Facebook also terminated access to its feed that provided ranked public posts that mentioned a specified topic, which could be a hashtag, an event, a place, etc. Access to this feed was terminated on 19 September 2016. See *Ibid*.

4 At the time it released its report, the ACLU noted that it had earlier made the social media companies aware of Geofeedia's uses of their Data. The ACLU reported that Twitter had "taken some recent steps to rein in Geofeedia though it has not ended the data relationship." (See *Ibid*.) Following the public release of the report, Twitter announced it had ended the relationship. See Ally Marotti, "Twitter Cuts off Chicago Startup Geofeedia after ACLU

Although Geofeedia was not the only company providing social media data analytics services to law enforcement, the backlash against Geofeedia may have been particularly strong because of its services' descriptions, which were revealed in emails obtained by the ACLU. For example, the ACLU posted an email from Geofeedia to one police force that explained why its service should be chosen over its competitors. Reasons offered were that Geofeedia used a combination of keywords, hashtags and geolocation to gather its data; it drew from eight social media sources; it was able to access social media data "in perpetuity", and it had access to richer and more complete data because it paid for premium levels of access. The email correspondent specifically noted that Geofeedia paid for Twitter's Firehose service.[5] Geofeedia also claimed to have a partnership with Instagram. It offered an "alerts" functionality as well as mobile apps. In an email to another police service, Geofeedia boasted of a confidential and binding agreement with Facebook, which they claimed would eventually lead to an even greater volume of data becoming available.[6]

The backlash against Geofeedia was also heightened because the company used two recent sets of protests as examples of its usefulness to police services. These protests followed the death in police custody of Freddie Gray, a young African-American man who had been arrested in Baltimore,[7] and the shooting death by police of Michael Brown,[8] an eighteen-year-old African-American man

reports police surveillance" (Chicago Tribune, 11 October 2016), available at: http://www.chicagotribune.com/bluesky/originals/ct-twitter-suspends-geofeedia-access-bsi-20161011-story.html (accessed 12 June 2017).

5 For the embedded email of 20 October 2015, see Cagle, *supra* n. 2. Twitter's Firehose service is a fairly exclusive, paid premium access to the full flow of Twitter data. See *infra* n. 82.

6 For a reproduction of the Email of 11 May 2016, see Cagle, *supra* n. 2.

7 David Graham, "The Mysterious Death of Freddie Gray" (The Atlantic, 22 April 2015), available at https://www.theatlantic.com/politics/archive/2015/04/the-mysterious-death-of-freddie-gray/391119/ (accessed 13 June 2017).

8 Larry Buchanan et al, "Q & A: What happened in Ferguson?" (New York Times, 10 August 2015) available at https://www.nytimes.com/interactive/2014/08/13/us/ferguson-missouri-town-under-siege-after-police-shooting.html (accessed 13 June 2017).

in Ferguson, Missouri. By offering services to monitor those who protested police violence against African Americans, Geofeedia aggravated a climate of mistrust, racial division, and a sense that the authorities were using surveillance and profiling to target minority communities.[9]

The public pillory of Geofeedia, combined with the swift response of social media companies to terminate its access to their data feeds, led to a marked drop in Geofeedia's business. Yet, while many police services subsequently terminated their contracts with Geofeedia, it is unclear whether they did so because Geofeedia no longer had access to key data, or because they were mindful of the public outcry against the use of such services.[10] In November 2016 Geofeedia reportedly laid off almost half of its workforce.[11] In early 2017, the company's site was still accessible on the internet; yet there were no news stories on its media page since before the release of the ACLU report, and its events and

---

9 For just a few examples of media reports that emphasised the issue of targeting of minority communities, see Benjamin Powers, "How Police Use Social Media to Track and Target Activists of Colour" (Complex Life, 17 November 2016), available at http://ca.complex.com/life/2016/11/police-surveillance-activists-people-of-color (accessed 13 June 207); Tanasia Kenney, "ACLU Blasts Facebook, Twitter and Instagram for Helping Police Track Black Activists Using Social Media Surveillance Product"(Atlantic Black Star, 12 October 2016), available at http://atlantablackstar.com/2016/10/12/aclu-blasts-facebook-twitter-and-instagram-for-helping-police-track-black-activists-using-social-media-surveillance-product/ (accessed 13 June 2017); Janine Jackson, "Our Identity Is Often What's Triggering Surveillance"(FAIR 1 November 2016), available at http://fair.org/home/our-identity-is-often-whats-triggering-surveillance/ (accessed 13 June 2017).

10 Not all police services terminated access right away. See Tim Lockette, "Alabama Police Maintain Geofeedia Subscription Despite Dwindling Social Feeds" (The Anniston Star, 9 January 2017), available at http://www.govtech.com/social/Alabama-Police-Maintain-Geofeedia-Subscription-Despite-Dwindling-Social-Feeds.html (accessed 13 June 2017). In some cases, police services indicated that they were reviewing their contracts with Geofeedia, although the reasons appeared to relate more to Geofeedia's loss of key data. See Kevin Rector and Alison Knezevich, "Social media companies rescind access to Geofeedia, which fed information to police during 2015 unrest" (Baltimore Sun, 11 October 2016), available at http://www.baltimoresun.com/news/maryland/crime/bs-md-geofeedia-update-20161011-story.html (accessed 13 June 2017).

11 Amina Elahi, "Geofeedia cuts half of staff after losing access to Twitter, Facebook" (Chicago Tribune, 21 November 2016), available at http://www.chicagotribune.com/bluesky/originals/ct-geofeedia-cuts-jobs-surveillance-bsi-20161121-story.html (accessed 13 June 2017).

webinars page showed no new activity since the fall of 2016. Links for demos or other information were no longer functional. It is difficult to determine if the company is still in business, although a news report in January 2017 suggested that it was still operating at that time, although in a limited capacity.[12]

3 Other intermediaries

Geofeedia is not the only company to build itself upon the mining of social media data feeds, nor is it the only one to make use of the geospatial data available in these feeds. In the emails accessed by the ACLU, Geofeedia compared itself with its competitors for law enforcement business. It identified Snaptrends as its closest competitor.[13] Snaptrends was also caught up in the backlash caused by the ACLU disclosures. It was reported that the company lost its own access to certain social media data streams because of its law enforcement clientele.[14] Although an early report indicated that it had ceased operations following the ACLU disclosure, the CEO of Snaptrends described the company as "pivoting", abandoning its law enforcement clientele to focus on business intelligence.[15] Today, Snaptrends describes itself as providing software that "empowers organizations to visualize social conversations by analyzing social media content in any specified geographic location."[16] Interestingly enough, its description of its services could include surveillance services based on geolocation data –

12 Lockette, *supra* n. 10.

13 For a reproduction of the Email of 20 October 2015, see Cagle, *supra* n. 2. Geofeedia also named HootSuite and TweetDeck as competitors of sorts, but noted that these two companies performed their analytics using only keywords and hashtags, whereas Geofeedia also used geolocation information.

14 Lania Rosales, "Snaptrends quietly lays off entire staff, ceases operations" (The American Genius, 21 October 2016), available at https://theamericangenius.com/business-news/snaptrends-quietly-lays-off-entire-staff-ceases-operations/ (accessed 13 June 2017).

15 Billy Utt, "Snaptrends CEO Responds: No more Govt Surveillance" (Austininno, 2 November 2016), available at http://austininno.streetwise.co/2016/11/02/snaptrends-ceo-responds-no-more-govt-surveillance/ (accessed 13 June 2017).

16 Snaptrends, available at http://snaptrends.com/ (accessed 13 June 2017).

although the company now distances itself from such activities. Snaptrends also states that it uses location, keywords and user profiles in its analytics. In its privacy policy relating to content harvested from social media sites, it indicates that it uses only publicly available data that is harvested using publicly available APIs.[17] As with Geofeedia, it claims to have access to data from multiple social media networks. It advertises its integration with Esri, a digital mapping company, to provide a new tool called Esri partnership and integration. According to Snaptrends, "[t]he integration between Snaptrends and Esri gives users of both platforms the ability to access a more comprehensive data set than ever before, broaden their understanding of particular locations, and use that information to make positive, impactful decisions."[18]

As of 30 March 2017, Snaptrends' website has not referenced any relationships with law enforcement or national security. It identifies client clusters in marketing and brand management, business insights, healthcare, sports and athletics, and energy and utilities.[19] However, it does identify "public safety" as an area for which its data is useful, noting:

> Public safety organizations routinely use Snaptrends technology to understand community sentiment, increase situational awareness, and improve community engagement. With the power of social media, public safety and emergency management organizations can gain valuable insights about their communities in order to better serve and protect.[20]

---

17 Snaptrends, "Social Media Content Privacy Policy" (16 July 2016), available at http://snaptrends.com/social-content-policy/ (accessed 13 June 2017).

18 Snaptrends, "Esri Partnership and Integration," (Snaptrends), available at http://snaptrends.com/esri/ (accessed 13 June 2017).

19 Snaptrends, "Where does Snaptrends work best?", available at http://snaptrends.com/social-media-for/ (accessed 13 June 2017).

20 *Ibid*. Note that Snaptrends uses the phrase "to serve and protect" which has been the motto of the Los Angeles Police Department since 1955, and which has been embraced by other police services since that time. See Mike Burg, "To Serve and Protect" (Police Patrol, 1

While there is no doubt that these types of data analytics may prove useful in disasters and other crises, there is considerable ambiguity in what "public safety" means, or what it means for public authorities to "gain valuable insights about their communities in order to better serve and protect."[21]

Another company linked to social media surveillance is Dataminr. The company provides tools for social media data analytics. On its website, it states that it "transforms the Twitter stream and other public datasets into actionable alerts, providing must-know information in real-time for clients in Finance, the Public Sector, News, Corporate Security and Crisis Management".[22] It provides instant analysis of "all public tweets and other publicly available data".[23] Regarding the public sector, Dataminr emphasises its capacity to enable rapid response by alerting first responders to breaking events.[24]

In December 2016, the ACLU reported it had discovered that Dataminr provided its tools to Fusion Centers. Fusion Centers are partnerships between different law enforcement agencies that are created under the auspices of the Department of Homeland Security in the US. Their role is to collect vast amounts of data to perform threat analysis. They contracted with Dataminr for the searching of real-time and historical tweets. In particular, ACLU was concerned with the fact that Dataminr offered a geolocation dimension to its analysis. According to the ACLU, Dataminr's Geospatial Analysis Application facilitated surveillance. It observed that:

December 1998), available at http://www.policemag.com/channel/patrol/articles/1998/12/to-serve-and-protect.aspx (accessed 13 June 2017).

21 *Ibid*.

22 Dataminr, available at https://www.dataminr.com/ (accessed 13 June 2017).

23 *Ibid*.

24 Dataminr, "Public Sector", available at https://www.dataminr.com/public-sector (accessed 13 June 2017).

> Settings in the Geospatial App even allowed the government to focus on monitoring journalists and organizations. Using Dataminr, fusion centers like JRIC could search billions of real-time and historical public tweets and then potentially share information with the federal government.[25]

Similar to Geofeedia, Dataminr's connections with the Fusion Centers were uncovered by the ACLU following a series of public record requests. Emails received as part of this set of requests revealed Dataminr's claims to have access to Twitter's "full firehose of 500 million tweets in real-time."[26] The company linked its access to the Firehose data to its ability to provide analytics that would assist in the surveillance of protest activities.[27] It also promoted its new Geospatial Analysis Application.[28] Following the revelations, Twitter announced that Dataminr (which is partly owned by Twitter)[29] had stopped providing Fusion Centers with access to the Twitter Firehose of data, and that it no longer provided social media surveillance tools to law enforcement, whether at the local, state or national levels.[30]

Other companies have been associated with social media data mining for police surveillance purposes. For example, the Brennan Center for Justice created an interactive map of police contracts with social media data mining companies,

---

25 Nicole Ozer, "Twitter Cuts of Fusion Spy Centers' Access to Social Media Surveillance Tool" (American Civil Liberties Union of Northern California, 15 December 2016), available at https://www.aclunc.org/blog/twitter-cuts-fusion-spy-centers-access-social-media-surveillance-tool (accessed 13 June 2017).

26 American Civil Liberties Union of Northern California, "Email to LAPD" (2015), available at http://www.aclunc.org/docs/20151130_dataminr_email_to_lapd.pdf (accessed 2 December 2017).

27 *Ibid*.

28 ACLU, "Email from Dataminr to JRIC," (15 March 2016), available at http://www.aclunc.org/docs/20160315_dataminr_email_to_jric.pdf (accessed 13 June 2017).

29 Twitter is reported to own 5% of Dataminr. See Jen Wieczner, "CIA Director is disappointed in Twitter and Dataminr" (Fortune, 20 June 2016), available at http://fortune.com/2016/06/20/twitter-dataminr-cia/ (accessed 13 June 2017). This article describes the two companies as having a "data partnership".

30 Ozer, *Supra* n. 25.

and published the related procurement invoices.[31] While Geofeedia, Snaptrends and Dataminr are among the contracting companies, others include Media Sonar,[32] Babel Street,[33] Beware (Intrado Systems),[34] Meltwater,[35] PATHAR,[36] and Digital Stakeout.[37] The Brennan Center for Justice was careful to note that it documented procurement of analytics services, but that it was not always possible to tell precisely what type of analytics services had been procured. Following the publicity surrounding its activities, Twitter was reported to have

---

31 Brennan Center for Justice, "Map: Social Media Monitoring by Police Departments, Cities and Counties" (16 November 2016), available at https://www.brennancenter.org/analysis/map-social-media-monitoring-police-departments-cities-and-counties (accessed 13 June 2017).

32 Media Sonar, available at https://www.mediasonar.com/ (accessed 13 June 2017).

33 Babel Street describes itself as a "multilingual, geo-enabled text analytics" company. See Babel Street, available at http://www.babelstreet.com/ (accessed 13 June 2017).

34 Beware, "Empower First Responders with Enhanced Situational Awareness," available at https://www.west.com/safety-services/public-safety/powerdata/beware/ (accessed 13 June 2017). Beware reportedly "scans commercial and public databases, as well as social media activity, in order to assign individuals a 'treat rating'." (See Mateescu et al, *supra* n. 1, p. 7.) They raise issues about the inability of officers and targets alike to assess or challenge the algorithms that produce this rating.

35 Meltwater, "Welcome to Outside Insight" available at https://www.meltwater.com/ (accessed 13 June 2017). Note that Meltwater is a service not specifically targeted to law enforcement. The adaptability of the software of many social media analytics companies to both business and policing needs makes it easier for them to "pivot" and it also makes it more difficult to gauge the nature and extent of their use by police services. As the Brennan Center for Justice noted in discussing their public records requests, in those jurisdictions where requisitioning is done centrally, it is difficult to tell whether analytics services have been acquired for policing or for other municipal purposes. See Rachel Cohn and Angie Liao, "Mapping Reveals Rising Use of Social Media Mapping Tools by Cities Nationwide" (Brennan Center for Justice, 16 November 2016), available at https://www.brennancenter.org/blog/mapping-reveals-rising-use-social-media-monitoring-tools-cities-nationwide (accessed 13 June 2017).

36 PATHAR is described as a company that "develops real-time data analysis software for intelligence, defense, and law enforcement communities. It also provides end-to-end analytical and training services, as well as produces analysis products, papers, presentations, and assessments on various topics." See Bloomberg, "Company Overview of PATHAR, Inc.", available at http://www.bloomberg.com/research/stocks/private/snapshot.asp?privcapId=247089363 (accessed 13 June 2017).

37 Digital Stakeout, "Stakeout Your Digital Risk", available at http://www.digitalstakeout.com/ (accessed 13 June 2017).

banned Media Sonar from access to its Firehose of data; the same company was also reportedly banned from access to Instagram data.[38]

Regardless of the changes that may have been wrought with respect to the activities of Geofeedia, Snaptrends, Dataminr and Media Sonar, the fact remains that the data streams that enabled these companies' business models continue to be available and that the algorithms for processing and analysing social media data continue to evolve. It is also the case that the social media data available through these feeds contains important geo-referenced information such as location data, and geographical references in hashtags and in posts. The capacity remains, as does presumably the interest, in using "public" georeferenced social media data in surveillance activities. In fact, although both Snaptrends and Geofeedia lost business after stringent media attention, Snaptrends survived the backlash. A report on Snaptrends in late 2016 noted that "SnapTrends is one of at least a dozen companies whose sole purpose is social media surveillance."[39] This remains a competitive service industry with ever-evolving analytics capacity.

Concerns over the use of social media data for surveillance and tracking purposes are not limited to the United States. In fact, such concerns are heightened in countries with repressive governments and draconian laws for cracking down on dissent. Bloomberg News reported that Snaptrends "promoted social media analytics tools to authorities in Azerbaijan, Bahrain, Malaysia, Saudi Arabia, Turkey, and other countries known to suppress online

---

38 Amanda Margison, "Twitter and Instagram ban London, Ont. company for helping police track protestors" (CBC News, 19 January 2017), available at http://www.cbc.ca/news/canada/toronto/twitter-bans-firm-police-protesters-1.3942093 (accessed 13 June 2017).

39 Dell Cameron, "Twitter cut ties with second firm police use to spy on social media" (The Daily Dot 20 October 2016), available at https://www.dailydot.com/layer8/twitter-snaptrends-geofeedia-social-media-monitoring-facebook/ (accessed 13 June 2017).

speech".[40] It also reported that: "The company often approached potential customers during moments of social unrest."[41] In June 2016 there was controversy over the fact that although Twitter had recently terminated Dataminr's relationship with the Central Intelligence Agency of the United States, it continued to have a commercial relationship with a state-owned Russian news outlet.[42]

Clearly, the Geofeedia story is just part of what was uncovered by the ACLU and by a companion study by the Brennan Center for Justice. There are a significant number of social media data analytics companies that are prepared to offer a broad range of analytics services to their clients, and that do not hesitate to include law enforcement agencies amongst their clientele. While there has been other documentation of the use of social media platforms by police services,[43] these cases are different because the reliance on intermediaries provides a) enhanced access to a greater volume of data and b) additional analytics tools.

4 Social media data mining and surveillance

Social media policing scholar Daniel Trottier defines surveillance as "the

---

40 Benjamin Elgin and Peter Robison, "How Despots Use Twitter to Hunt Dissidents" (Bloomberg Businessweek, 27 October 2016), available at https://www.bloomberg.com/news/articles/2016-10-27/twitter-s-firehose-of-tweets-is-incredibly-valuable-and-just-as-dangerous (accessed 13 June 2017).

41 *Ibid*.

42 The head of the CIA was reported to have said, at a US Senate Hearing, that: "It appears as though Dataminr was directed to not provide its service to the CIA Intelligence Community and so therefore, we need to be able to leverage other capabilities in order to make sure that we have the insight we need to protect this country." (Wieczner, *supra* n. 29.)

43 Use of social media by police services is now commonplace, to the extent that an initiative of the US Department of Justice led to the publication of a set of guidelines to aid police services in developing social media policies. See Global Advisory Committee, "Developing a Policy on the Use of Social Media in Intelligence and Investigative Activities: Guidance and Recommendations" (2013) <https://it.ojp.gov/documents/d/Developing%20a%20Policy%20on%20the%20Use%20of%20Social%20Media%20in%20Intelligence%20and%20Inves....pdf> accessed 13 June 2017.

sustained collection of personal information from a strategically advantageous position."[44] Social media sites offer unique potential for state surveillance activities. These sites have been described as:

> […] web-based platforms that integrate different media, information and communication technologies and that allow at least the generation of profiles that display information that describes the users, the display of connections (connection list), the establishment of connections between users that are displayed on their connection lists and the communication between users.[45]

As such, they offer rich stores of data about individuals and about their diverse and multiple relations and connections with one another. Because of the way that social media sites are "embedded in everyday life",[46] they create a situation where "[m]ediated everyday activity is more visible to policing and investigations."[47] Using social media surveillance tools, police are able to access private information with little effort, which is indicative of asymmetrical relations of visibility with the public on social media. They are able to know intimate and aggregate details about targeted individuals, who in turn are unaware that they are under watch, or collaborating in investigations.[48]

---

44 Daniel Trottier, "Policing Social Media" (2012) 49(4) *Canadian Review of Sociology* 411-425, p. 415.

45 Daniel Trottier and Christian Fuchs, "Theorising Social Media, Politics and the State: An Introduction" in Daniel Trottier and Christian Fuchs (eds.), *Social Media, Politics and the State: Protests, Revolutions, Riots, Crime and Policing in the Age of Facebook, Twitter and YouTube* (New York: Routledge, 2015) pp. 3-38, at 6.

46 Trottier, "Policing Social Media", *supra* n. 44, p. 413.

47 *Ibid*.

48 *Ibid.*, p. 415. A 2014 Survey of law enforcement professionals found that 81% used social media platforms in their investigations. 40% of respondents indicated they used social media to monitor events. See LexisNexis, "Social media Use in Law Enforcement: Crime prevention and investigative activities continue to drive usage" (2014), available at https://www.lexisnexis.com/risk/downloads/whitepaper/2014-social-media-use-in-law-enforcement.pdf (accessed 14 June 2017).

Given the nature of the data they host, it is not surprising that social media platforms lend themselves to surveillance in a number of ways. Trottier identifies six types of police searching that can take place via social media.[49] These are distinguished by search methodology from a policing perspective. Not all of these methods involve surveillance; some are investigative. First, there is manual searching of social media platforms. Police officers, like any other members of the public, can go online to search through the public-facing content of social media websites.[50] Police need no special judicial authorisation to engage in these activities.[51] The information is considered publicly available. Where a police officer requires further information that is not public, such as user account information or information stored on private pages, additional steps are required.

These additional steps may fall within Trottier's second category of search: Police contact social media companies to request information about a user or users of the site. While the law may generally permit companies to reveal some information to the police without need for a warrant, a warrant will be necessary where the request is considered a "search" for information in which there is a reasonable expectation of privacy. Social media companies make undertakings to their users in the form of privacy policies. These policies usually alert users to

---

49 Daniel Trottier, "Vigilantism and Power Users: Police and User-Led Investigations on Social Media" in Daniel Trottier and Christian Fuchs (eds.), *Social Media, Politics and the State: Protests, Revolutions, Riots, Crime and Policing in the Age of Facebook, Twitter and YouTube* (New York: Routledge, 2015), pp. 209-226, at 215ff. Trottier does not address impersonation in his six categories. This occurs where police create false identities to "friend" certain individuals or to otherwise extract information. See Mateescu et al, *supra* n. 1, p. 1.

50 That they do this is evident. See, for example, Trottier, "Policing Social Media", *supra* n. 44. See also the Global Advisory Committee guidance, where it states: "As a part of the agency's authorized law enforcement purpose, social media sites may be accessed to follow up on tips and leads, suspicious activity reports, investigative support, development of criminal intelligence, and the development of situational awareness reports." (Global Advisory Committee, *supra* n. 43, p. 11.)

51 Global Advisory Committee, *supra* n. 43, p. 12.

the fact that the disclosure of information to the police might take place and in what circumstances.[52] Where warrants are obtained for this type of information, there is a measure of transparency and accountability. This creates a paper trail and allows for judicial oversight of the search and its parameters. In other instances, voluntary transparency reporting may provide the public with some sense of the frequency of this type of information request, the rate of compliance of the company, and other relevant details.[53]

The third type of search identified by Trottier involves the police search of "open source intelligence". This combines manual and automated searching of social media information that is publicly available. It resembles the use of a search engine to find results from a large compilation of data. Because of the public nature of the data, it is unlikely that a warrant would be required for such searches.

A fourth type of searching involves the interception of communications over social media. Interception of communications is an area where search warrants are typically required as this type of searching is considered highly intrusive. A fifth category of surveillance involves the installation by law enforcement officials of some form of device on a targeted individual's computer in order to track and record their activities. This too is considered highly intrusive and would require a warrant.

---

52 For example, Twitter's Privacy Policy provides that: "Notwithstanding anything to the contrary in this Privacy Policy, we may preserve or disclose your information if we believe that it is reasonably necessary to comply with a law, regulation, legal process, or governmental request; to protect the safety of any person; to address fraud, security or technical issues; or to protect our or our users' rights or property." Twitter, "Privacy Policy", available at https://twitter.com/privacy?lang=en (accessed 14 June 2017).

53 Twitter and Facebook, for example, submit voluntary transparency reports. See Twitter "Transparency Report", available at https://transparency.twitter.com/en.html (accessed 14 June 2017); Facebook, "Facebook Government Requests Report", available at https://govtrequests.facebook.com/ (accessed 14 June 2017).

The sixth category is the focus of this paper. It involves data analytics performed on substantial collections of social media data. The analytics can be quite sophisticated and can include mapping, predictive policing, profiling, the use of facial recognition software, and so on. While some of these activities may relate to the investigation of specific incidents, others are more predictive in nature and raise issues of targeting and profiling. As far as surveillance activities are concerned, these methods are relatively low-cost.[54] This category is interesting and problematic from a transparency and oversight point of view. First, it is unclear whether the data that is searched is "public" or "private", and indeed the question is itself misleading.[55] While in bulk it is comprised of vast quantities of publicly posted data, it is not public in the sense that it is freely publicly accessible. Typically, access to this volume and richness of data would require paid premium access.[56] The volume of information and its mode of access can thus be argued both to privatise the data and to render it more obscure from a user perspective. Further, the algorithms used in searching are typically proprietary and are developed by third party companies that offer access to both the data streams (for which they contract) and the analytics tools. Once again, there is a privatised dimension that comes from the private ownership of the

---

54 For example, the Brennan Center for Justice's map of police services using social media analytics also includes data about how much each police service paid for access (See Brennan Center for Justice, *supra* n. 31). The amounts are relatively small in the broader context of surveillance activities. See Elizabeth Dwoskin, "Police Are Spending Millions of Dollars to Monitor the Social Media of Protestors and Suspects" (Washington Post, 18 November 2016), available at https://www.washingtonpost.com/news/the-switch/wp/2016/11/18/police-are-spending-millions-to-monitor-the-social-media-of-protesters-and-suspects/?utm_term=.34adc7183059 (accessed 14 June 2017).

55 Nissenbaum, for example, is critical of the use of the public/private dichotomy in shaping approaches to privacy; she argues for a more richly contextualized approach. See Helen Nissenbaum, *Privacy in Context: Technology, Policy and the Integrity of Social Life* (Stanford, CA: Stanford Law Books, 2010).

56 This is certainly the case in the examples provided here of companies such as Geofeedia, Snaptrends and Dataminr, among others, who had paid premium access to social media data.

tools to extract meaning from the data.[57] At the same time, the data is less public and more obscure because these new meanings can only be extracted through specialised intervention.[58] Nevertheless, some may still choose to argue that the public nature of the data at its origin allows it to retain its public dimensions.

In this context, third party data analytics companies act as intermediaries between the social media companies and the police. As will be seen below, this creates a buffer that can hinder both transparency and oversight. And, while there may be significant privacy implications for individuals who are profiled, it is not clear that profiling or data analytics carried out using these third party services constitutes "searches" for which warrants are required. In essence, the police service contract is for data analytics to be carried out on a stream of publicly available data, and data for which user consent has been given for it to be licensed to analytics companies.

Privacy is an important consideration because it triggers judicial oversight in the context of police activities. Yet it is not the only issue raised by social media surveillance, a fact that makes transparency and oversight more complex. Some of the concerns raised by the analytics services will relate to profiling carried out on prohibited grounds of discrimination, or will relate to the undermining of constitutional values such as the freedoms of expression or association. The extent to which this is the case may depend upon the nature of the searches carried out or the algorithms deployed by the intermediaries. Public disclosures of police service contracts for social media analytics do not reveal this information and indeed, third party algorithms are protected as confidential

---

[57] The complexity of the relationship between public and private in the context of data and algorithms is a growing issue. See, for example, Rebecca Wexler, "Life, Liberty, and Trade Secrets: Intellectual property in the Criminal Justice System", Stanford Law Review (forthcoming), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2920883 (accessed 27 November 2017).

[58] See, for example, Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown Publishing, 2016).

information. The ability of companies or those behind them to remake themselves, change corporate identities, hide behind dummy accounts, or operate through third parties makes controlling the use of social media data in surveillance a challenging task – and an ever-shifting target.[59]

5 Social media in crisis contexts

Social media is capable of important uses during emergencies and other crises. Its use during the major earthquake in Japan in 2011[60] is well documented, and it has been used to communicate about other natural disasters such as Hurricane Sandy,[61] and the Alberta wildfires in 2016.[62] Murthy observes that social media communications about such events often occur more rapidly than with conventional media,[63] and they may cover a range of different topics that are important and that may not be central in mainstream media reporting.[64] For

[59] Coen and Liao, *supra* n. 35.

[60] Dhiraj Murthy, *Twitter: Social Communication in the Digital Age* (Cambridge, UK: Polity Press, 2013), pp. 79-81.

[61] Yury Kryvasheyeu et al, "Rapid Assessment of Disaster Damage Using Social Media Activity" (2016) 2(3) Science Advances, available at http://advances.sciencemag.org/content/2/3/e1500779 (accessed 27 November 2017); Department of Homeland Security, "Lessons Learned: Social Media and Hurricane Sandy" (2013) available at https://www.dhs.gov/sites/default/files/publications/Lessons%20Learned%20Social%20Media%20and%20Hurricane%20Sandy.pdf (accessed 14 June 2017).

[62] Vincent McDermott, "As Wildfire Crisis Unfolds, the Displaced Turn to a Twitter Account" (FortMcMurrayToday.com, 8 May 2016), available at http://www.fortmcmurraytoday.com/2016/05/08/as-wildfire-crisis-unfolds-the-displaced-turn-to-a-twitter-account (accessed 14 June 2017); Lisa Grant, "Behind the Twitter account that helped evacuate Fort McMurray" (660 News, 9 May 2016), available at http://www.660news.com/2016/05/09/behind-the-twitter-account-that-helped-evacuate-fort-mcmurray/ (accessed 14 June 2017).

[63] Wexler, *supra* n. 57, pp. 80-81.

[64] *Ibid.,* pp. 71-72. For a discussion of the growing importance of social media data to public safety organisations, see Babak Akhgar et al, "Social Media in Crisis Events: Open Networks and Collaboration Supporting Disaster Response and Recover" (2013 IEEE International Conference on Technologies for Homeland Security (HST), Waltham, 12-14 November 2013).

example, social media can give the precise location of individuals in need of assistance; it can provide photo and video images from otherwise unreachable areas; it can also provide details of shortages of food or supplies, feelings of anxiety, and a range of other topics. The usefulness of social media in times of crisis is, of course, part of what makes it of interest for surveillance purposes. There is also a fine line between crises and policing contexts. While natural or other disasters are public emergencies, some crises have more political dimensions. Further, "crisis" may be a matter of perspective. From a participant's perspective, a march or protest may be a moment of political expression. From a police perspective, it might be an unstable public gathering that has the potential to deteriorate into violence and damage to property. Crisis moments may therefore come in the form of public protests, whether these are single instances or a part of more sustained campaigns such as the Occupy movement[65] or the Arab Spring.[66] Poell notes that social media platforms often have conflicted relationships with activists and protestors both encouraging the use of their platforms in crises and permitting state authorities to use their data resources for investigative purposes.[67] While on the one hand social media companies may wish to be associated with certain movements, not all such movements will fit their corporate images. Further, the sites remain essentially commercial enterprises, which, as Poell notes, "have a strong interest in limited access to the data that is shared and generated on their platforms, as this data allows them to

---

See also Carlos Castillo, *Big Crisis Data: Social Media in Disasters and Time-Critical Situations* (Cambridge: CUP, 2016).

65 Craig Kanalley, "Occupy Wall Street: Social Media's Role in Social Change" (The Huffington Post, 6 December 2011), available at http://www.huffingtonpost.com/2011/10/06/occupy-wall-street-social-media_n_999178.html (accessed 14 June 2017).

66 Nahed Eltanawy and Julie Wiest, "Social Media in the Egyptian Revolution: Reconsidering Resource Mobilization Theory" (2011) 5 International Journal of Communications 1207-1224.

67 Thomas Poell, "Social Media Activism and State Censorship" in Daniel Trottier and Christian Fuchs (eds.), *Social Media, Politics and the State: Protests, Revolutions, Riots, Crime and Policing in the Age of Facebook, Twitter and YouTube* (New York: Routledge, 2015), pp. 189-206, at 197.

target personalised advertising and services at their users."[68] The result is a lack of user control over the data they contribute, and, as Poell observes, activists "are left to guess what exactly happens with this data, which is evidently particularly problematic for activists facing dictatorial regimes."[69] The private ownership of social media databases "renders activists using them more vulnerable with regard to their privacy and contents, especially when private corporate interests align with the desire of state repression in times of mobilisation."[70]

6 Accessing/obtaining data

Social media companies treat the content that is made available over their sites as "publicly available data". In fact, this public sharing of information is at the heart of what it means to function as a social media platform. As Kitchin notes: "These sites are all reliant on active participation by a public willing to share information about their lives and undertake work such as writing, editing, extending, remixing, posting, sharing, tagging, communicating, and so on."[71] These activities generate significant volumes of data. Some of these data, particularly those contributed by users, are freely publicly accessible from the host site. Other data, collected through the tracking of user activities is not public, but may be commercialised by the social media company. User-contributed data are publicly accessible in the sense that members of the public may freely access these data by visiting and browsing the platform. Some social media platforms

---

68 *Ibid.,* p. 199.

69 *Ibid*.

70 Donatella della Porta and Alice Mattoni, "Social Networking Sites in Pro-democracy and Anti-Austerity Protests: Some Thoughts from a Social Movement Perspective" in Daniel Trottier and Christian Fuchs (eds.), *Social Media, Politics and the State: Protests, Revolutions, Riots, Crime and Policing in the Age of Facebook, Twitter and YouTube* (New York: Routledge, 2015), pp. 39-63, at 58.

71 Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* (London: Sage, 2014), p. 95.

such as Facebook may limit broad public access to the "public" portions of their pages. Other platforms, such as Twitter, make almost all user-contributed content publicly accessible. Publicly *accessible* (capable of being accessed by the public) information must be distinguished from publicly *available* information (available for the public to reuse). Most social media companies also make user-contributed data available to the public through APIs which permit the searching of historical data or allow access to limited bulk quantities of data.[72] Kitchin observes that because the social media platforms are private companies, "data are being traded into privately owned hands who then seek to produce new models of capital accumulation by extracting value from them".[73] Social media companies therefore typically provide premium access to larger quantities of their data for a fee. It is this data that is used by data analytics intermediaries.

Users generally consent to the public dissemination and reuse of their contributed data through the privacy policies of social media companies. For example, Twitter's privacy policy informs its membership that: "Twitter broadly and instantly disseminates your public information to a wide range of users, customers, and services, including search engines, developers, and publishers that integrate Twitter content into their services, and organizations such as universities, public health agencies, and market research firms that analyse the information for trends and insights."[74] Where a user also enables location functionality, their precise location will be associated with their tweets, and users are notified that this information will also be accessible via the APIs.[75]

---

[72] *Ibid*.

[73] *Ibid*., p. 95.

[74] Twitter, "Privacy Policy", *supra* n. 52.

[75] Twitter, "FAQs about adding location to your tweets" (2017), available at https://support.twitter.com/articles/78525# (accessed 14 June 2017). There is a link to these FAQ's from the Twitter Privacy Policy. See *ibid*.

Most users think of public social media data as being that which they freely contribute to public fora. Thus, for example, Twitter users would be naïve, if not oblivious, if they did not consider their public profiles and each tweet to be "public". However, what may be less widely understood is the fact that each tweet consists not just of the permitted number of characters, but also contains metadata. Thus, for example, as Bloomberg news explains, tweets contain:

> […] more than 30 other data fields mostly hidden from regular users. With this data, users are sortable by, among other things, the device they tweeted from, the names and locations they give in their profiles, and the number of times tweets have been retweeted. If the user has consented, the data can also include the exact location from where a tweet was sent, a practice known as geotagging.[76]

While Bloomberg News reported in 2016 that only 2% of tweets were geotagged, they also noted that data analytics companies used other tools to provide geo-referencing for tweets, including linking Twitter accounts with other social media accounts (which may include geolocation data), or using other geolocation references in tweets or hashtags.[77]

Twitter makes some of its data available through free public APIs, which enable public access to certain streams of data. These include "public streams",[78] "user streams",[79] and "site streams".[80] Public streams offer a level of access that

---

76 Elgin and Robison, *supra* n. 40.

77 *Ibid*.

78 The public stream API allows a company or individual to follow tweets according to certain parameters. These parameters include "up to 400 track keywords, 5,000 follow user ids and 25 0.1-360 degree location boxes". Twitter, "Twitter Development Documentation" (2017), available at https://dev.twitter.com/streaming/reference/post/statuses/filter (accessed 14 June 2017). Those requiring more extensive access are directed towards Gnip, which provides commercial API solutions.

79 The user stream API allows a company or individual to follow tweets about a particular individual. See *ibid*.

80 The site stream API allows a user to follow tweets about a particular website. See *ibid*.

is suitable for many non-commercial uses. For those who require more extensive access to Twitter data, Twitter has established Gnip as its affiliate company that manages paid access to its commercial APIs. Through Gnip it is possible to contract for different levels or types of access to data. These include the Historical Power track (providing full historical access to tweets); an API that enables searching of tweets in the last 30 days; an API that permits a full archive search; or an API that provides aggregate data about audiences defined by the client.[81] There is also the full Firehose API, which provides live streamed access to the full volume of tweets (currently estimated at between 500 million and 1 billion a day).[82] Twitter also provides a service that allows a client to "simultaneously collect social data from multiple public APIs, allowing you to simplify and save on valuable engineering resources."[83] Bloomberg news reports that although pricing for full Firehose access is not public, data extracted from a lawsuit involving Twitter suggests that full Firehose access may cost more than one million dollars per year.[84] The same report suggests that Twitter's ability to monetise its data in this way may be essential to its viability as a business.

The high cost of Firehose access is borne by intermediaries who then provide a combination of analytics software and data access to their clients, including police services. This access is relatively inexpensive, compared to other forms of surveillance.[85]

---

81 Twitter, "Gnip APIs" (2017), available at http://support.gnip.com/apis/ (accessed 14 June 2017).

82 Twitter describes its Firehose Access in these terms: "Gnip offers complete access to real time streams of public data from the top social networks, giving you access to every social activity in the blink of an eye." Twitter, "All the Social Media Data You Need Under One Roof" (2017), available at https://gnip.com/sources/ (accessed 14 June 2017). The full Firehose of data contains, as noted above, not just the tweets, but additional metadata.

83 *Ibid*. The social media sites included in this service include: Facebook, YouTube, Instagram, and Flickr.

84 Elgin and Robison, *supra* n. 40.

85 Mateescu et al, *supra* n. 1, p. 1.

7 Privacy, civil liberties and social justice

The right to privacy – and to be free from unauthorised surveillance – has been the traditional lens for oversight when it comes to the activities of law enforcement and national security officials. Indeed, intrusiveness on privacy rights has become the measure of determining what constitutes an "unreasonable" search.[86] The system that has been built around privacy values provides an important means of achieving transparency, as authorities are required to seek judicial authorisation where their surveillance activities intrude upon a reasonable expectation of privacy. The process of seeking judicial authorisation provides a measure of transparency as it requires disclosure of activities, at least to a judge, who provides oversight and who sets boundaries on permissible activities. Any resulting court process also provides both transparency (through the open courts principle) and oversight, as it is a forum in which any objections to the state's conduct of the investigation can be raised.

The public nature of social media data undermines traditional oversight and transparency paradigms, particularly those based on privacy norms. This is in large part because social media data is considered to be public. Its visibility is generally due to conscious decisions on the part of the individual who tweets or posts, to make the information publicly available. There is generally no reasonable expectation of privacy in information that one makes public about oneself. This does not mean that social media surveillance through data analytics does not raise new privacy issues. Privacy scholars have argued that evolving technologies may require new privacy paradigms.[87]

---

[86] The constitutions of both the United States and Canada, for example, protect against "unreasonable" searches and seizures; these provisions have come to be understood as constitutional privacy rights. See United States Constitution, Amendment 4; The Constitution Act, 1982, s. 8.

[87] See Julia Lane et al (eds.), *Privacy, Big Data, and the Public Good: Frameworks for Engagement* (Cambridge: CUP, 2014); Daniel Solove, *Understanding Privacy* (Cambridge, MA: Harvard

Transparency and oversight in this context therefore require a deeper understanding of the technology, as well as more complex arguments about reasonable expectations. Thus, for example, it becomes important to consider what data is being analysed (not just user-contributed data, for example, but also metadata harvested by the system and of which the user may be unaware), whether the information is truly public, or public only to paying customers such as private sector data analytics companies, and even if individual data is provided voluntarily by users, there should still be a reasonable expectation of privacy when it comes to automated analytics carried out upon a vast collection of data linked across different platforms.[88] Where data is derived from multiple sources, code-based analytics may have a privacy impact that goes beyond whatever consent has been provided to individual pieces of data or whatever voluntary sharing has taken place. In addition, data analytics can create information that is of a heightened personal nature, raising reasonable expectations of privacy.[89] In a sense, then, data analytics can generate new personal information – at least new in the sense of not being previously disclosed in that form.

Transparency and oversight in relation to big data analytics also require lenses other than privacy. Big data analytics used in surveillance may involve constitutional values such as the freedoms of speech and association, as well as anti-discrimination values. This is particularly the case where the analytics are used to look for connections or associations between individuals, or for

University Press, 2008); Helen Nissenbaum, *Privacy in Context: Technology, Policy and the Integrity of Social Life* (Stanford, CA: Stanford Law Books, 2010).

88 For a discussion of some of the legal challenges in this area, see Katherine Strandburg, "Monitoring Datification, and Consent: Legal Approaches to Privacy in the Big Data Context," in Julia Lane et al (eds.), *Privacy, Big Data, and the Public Good: Frameworks for Engagement* (Cambridge: CUP, 2014), pp. 5-43.

89 See Bruce Schneier, *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World* (New York: Norton & Co., 2015), ch. 3.

connections between individuals and particular political expressions.[90] Thus, for example, a search algorithm that includes the hashtag "BlackLivesMatter" identifies individuals on the basis of both political speech and association with a group or movement. Further, because the movement challenges systemic racism, surveillance that is based on association with the movement also raises issues of racial discrimination.

There is a growing body of literature that challenges both the quality of the algorithms used in data analytics[91] and the assumptions and values that may shape them.[92] Such issues also support calls for greater transparency[93] since quite apart from issues of privacy and surveillance, there may be problems of bias and error, as well as systemic discrimination that affect the data generated by the algorithms and relied upon by authorities.

8 The normative approach

In the public outcry that followed the ACLU disclosures about Geofeedia, social media companies were quick to distance themselves both from Geofeedia and

---

90 For example, see Brennan Center for Justice, *supra* n. 31. It is noted that "[t]he Oregon Department of Justice and police in Oakland, CA monitored prominent figures of the Black Lives Matter movement by tracking hashtags on social media." It also found that "[i]n correspondence with the Fresno, CA, police department, a representative from software company Media Sonar proposed a list of keywords to scan in order to 'identify illegal activity and threats to public safety,' including 'dissent,' 'BlackLivesMatter,' and 'WeWantJustice.'"

91 See, e.g., O'Neil, *supra* n. 58; Elizabeth Joh, "Feeding the Machine: Policing, Crime Data, & Algorithms", William & Mary Bill of Rights Journal (forthcoming), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3020259 (accessed 27 November 2017).

92 See, e.g, Latonya Sweeney, "Discrimination in Online Ad Delivery" (2013) Data Privacy Lab White Paper 1071-1, available at http://ssrn.com/abstract=2208240 (accessed 27 November 2017); O'Neil, *supra* n. 58; Jennifer Winter, "Algorithmic Discrimination: Big Data Analytics and the Future of the Internet" in Jenifer Winter and Ryota Ono (eds.), The Future Internet: Alternative Visions (Cham: Springer, 2015), pp. 125-140; Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2015).

93 See, e.g., Danielle Citron and Frank Pasquale, "The Scored Society: Due Process for Automated Predictions" (2014) 89(1) Washington Law Review 1-33; Pasquale, *supra* n. 92.

from the use of their public data streams in surveillance activities. Most quickly severed ties with Geofeedia. The fact that so many social media companies cut off Geofeedia's access to their data may explain why that company was forced to lay off roughly half of its employees within six weeks of the ACLU story, and why it is unclear whether Geofeedia continues to operate today.

Perhaps because they collect such an enormous volume of often quite personal user information, social media companies seem very reluctant to be linked to police surveillance activities. In the wake of the ACLU disclosures, social media companies took steps to distance themselves from these activities. For example, as noted earlier, most social media companies announced they were cutting off data access to Geofeedia, and access was also reportedly cut to Snaptrends.[94] Social media companies took steps to make clear public statements disapproving of the use of their data in this way and asserting their intention to put a stop to it.[95]

In the aftermath of the Geofeedia revelations, social media companies were also quick to point out that their terms of use disallowed the kinds of uses made of their data by Geofeedia. While some of these companies maintained that it had never been legitimate to use their data for surveillance purposes, in the wake of the ACLU report, language in Terms of Service became more explicit on this point. For example, on 13 March 2017, Facebook and Instagram changed their developer policies to make it clearer that the use of data for surveillance was

---

94 See Lani Rosales, "Snaptrends Quietly Laws Off Entire Staff, Ceases Operations" (The American Genius, 31 October 2016), available at https://theamericangenius.com/business-news/snaptrends-quietly-lays-off-entire-staff-ceases-operations/ (accessed 14 June 2017); Dell Cameron, "Twitter Cuts Ties with Second Firm Police Use to Spy on Social Media" (The Daily Dot, 20 October 2016), available at https://www.dailydot.com/layer8/twitter-snaptrends-geofeedia-social-media-monitoring-facebook/ (accessed 14 June 2017).

95 For example, Twitter quickly tweeted its decision to cut off Geofeedia's access in light of the ACLU report. See https://twitter.com/policy/status/785861128589025281 (accessed 14 June 2017).

not permitted.[96] Facebook's Platform Policy for developers now provides that developers should not "use data obtained from us to provide tools that are used for surveillance."[97] Instagram, which is a company related to Facebook, contains a similar proviso in its Platform Policy.[98] Facebook also reserves the right to audit third party uses of their data, and developers are asked to disclose, in requests for access to Facebook data, how they intend to make use of the data.[99] Twitter now explicitly provides that data may not be used for surveillance: "Using Twitter's Public APIs or data products to track or profile protesters and activists is absolutely unacceptable and prohibited."[100] The company announced that "if developers violate our policies, we will take appropriate action, which can include suspension and termination of access to Twitter's Public APIs and data products."[101]

Although Twitter maintained that its policy had always been to prohibit the use of its data to track or profile protestors and activists, this did not prevent the rather widespread use of its data in this manner via companies such as Geofeedia, Snaptrends, Dataminr and others. Part of the reason for this is that there is little or no oversight to ensure compliance with the Terms of Service. The Terms of Service between a social media company and a data analytics firm is a contract between commercially sophisticated companies. At best, compliance is most likely presumed; at worst, insistence upon compliance is a choice guided by business interests.

---

96 Facebook, "Facebook U.S. Public Policy" (2017), available at https://www.facebook.com/uspublicpolicy/posts/1617594498258356 (accessed 14 June 2017).

97 Facebook, "Facebook Platform Policy", s. 3(1), available at https://developers.facebook.com/policy (accessed 14 June 2017).

98 Instagram, "Platform Policy", available at https://www.instagram.com/about/legal/terms/api/ (accessed 14 June 2017).

99 Facebook, "Facebook Platform Policy", *supra* n. 97, s. 6(8).

100 Chris Moody, "Developer Policies to Protect People's Voices on Twitter" (2016), available at https://blog.twitter.com/2016/developer-policies-to-protect-people-s-voices-on-twitter (accessed 14 June 2017).

101 *Ibid*.

Further, it is also a matter of interpretation as to what conduct falls within or outside norms set by a social media company. For example, in a story in the Washington Post, Dwoskin wrote that although Facebook no longer allowed the use of its data for surveillance and tracking of individuals, it still permitted law enforcement access to its feeds for the purposes of dealing with natural disasters and other emergencies.[102] The problem of course, is that there may be different interpretations of what constitutes an "emergency". Similarly, bans on the use of data for surveillance purposes presumes that there is a shared understanding of what constitutes surveillance. For some, data analytics may not be a form of surveillance.

The presence of an intermediary between the social media company and the end user of the analytics services can also create oversight challenges. Prior to its public disgrace, Geofeedia's Terms of Service offered an interpretation of the terms to which it was bound by the social media companies that provided access to their public data. Clause 7.4 of Geofeedia's terms of service to its customers provided:

> 7.4 Access and Compliance (Law Enforcement Customers).
>
> If You are utilizing the Services in a law enforcement capacity, You acknowledge that third party licensors of Social Media Content (a) only permit use of Social Media Content, in a law enforcement capacity, to serve Public Safety Purposes; and (b) prohibit the segmentation of social media user data (and derived profiles of social media users) in a law enforcement capacity by (i) alleged or actual commission of a crime; (ii) health status (having a disease or condition); (iii) negative financial status or condition; (iv) political affiliation or beliefs; (v) racial or ethnic origin; (vi) religious or

[102] Elizabeth Dwoskin, "Facebook Says Police Can't Use Its Data for 'Surveillance'" (Washington Post, 13 March 2017), available at https://www.washingtonpost.com/news/the-switch/wp/2017/03/13/facebook-says-police-cant-use-its-data-for-surveillance/?utm_term=.863df70cb1d8 (accessed 14 June 2017).

> philosophical affiliation or beliefs; (viii) sex life (including pregnancy); or (viii) trade union membership.[103]

Through their Terms of Service, Geofeedia passed responsibility for compliance with social media company policies to its end users. The Terms of Service may thus become a responsibility "hot potato" that begins with the social media companies, passes to the analytics company, down through to the client for the analytics service.

As noted earlier, police services increasingly use social media for a variety of purposes. There has been a move to encourage police services to develop social media policies to govern appropriate use of social media in policing.[104] These policies may address a range of issues including the use of social media by police for communicating information to the public, the use of personal social media by police officers, and the use of social media to gather information about specific individuals. However, because of the more recent emergence of social media data analytics, not all police social media policies address the use of data analytics. This leaves a normative gap.

In its document titled Developing a Policy on the Use of Social Media, the Global Advisory Committee states that: "The purpose of a social media policy is to define and articulate acceptable law enforcement practices related to using information obtained from social media sites."[105] The document addresses a number of key issues including identifying situations in which it is permissible to use social media data, and defining the level of authorisation needed in order

---

[103] Geofeedia Service Agreement, available at https://geofeedia.com/legal/service-agreement/geofeedia_service_agreement__july_2016_website_version.pdf (accessed 29 November 2017). .

[104] Global Advisory Committee, *supra* n. 43; Community Oriented Policing Services, "Social Media and Tactical Considerations for Law Enforcement" (Washington, D.C.: U.S. Department of Justice and Police Executive Research Forum, 2013), available at https://ric-zai-inc.com/Publications/cops-p261-pub.pdf (accessed 14 June 2017).

[105] Global Advisory Committee, *supra* n. 43, p. 9.

to use social media sites. The level of authorisation required may vary depending on the nature of the usage. The policy also calls for the evaluation of the quality and reliability of information obtained from social media sources. In addition, the Global Advisory Committee recommends that a policy address the storage and retention of data and intelligence products as well as the sharing or dissemination of information. Although these recommendations are not specific to the use of social media data analytics, they are all guidelines which can and should be adapted to that context.

The Global Advisory Committee also reminds police services that the use of social media tools must be situated within the overarching norms that govern police activity. Among these is the principle that:

> Law enforcement agencies should not collect or maintain information about the political, religious, or social views, associations, or activities of any individual or any group, association, corporation, business, partnership or other organization unless there is a legitimate public safety purpose, such as the information directly relates to criminal conduct or activity.[106]

A normative approach would thus also require police services to develop policies, or to revise existing social media policies, to expressly address the use of data analytics and to set parameters for legitimate use.

9 Transparency

While normative approaches are important in articulating the boundaries of acceptable use, greater transparency can help to address the challenges posed by the use of big data analytics for social media surveillance.[107] The disclosures by

[106] *Ibid.,* p. 10.

[107] For example, see Schneier, *supra* n. 89, pp. 170-171. Schneier advocates for greater transparency as a means of addressing the privacy threats of big data surveillance.

the ACLU and the Brennan Center for Justice were important, and achieved some concrete results (at least in the short term), but the problem is ongoing. After Twitter cut off access to its Firehose of data for certain companies, the ACLU acknowledged the importance of this result, but cautioned that:

> […] social media surveillance is just a piece of the surveillance puzzle. Companies and communities will need to take further steps in the months and years ahead to build in much stronger transparency, accountability, and oversight for government surveillance and make sure that rights are properly protected.[108]

Achieving transparency in this area is made more challenging by the number and nature of different actors involved. First, there are police services. These public authorities have some obligations to be transparent with respect to some of their activities. Moreover, two tiers of private sector companies are involved. One set of companies are the social media companies that have direct relations with the individuals that supply user-generated content to their platforms. It is this content that those companies provide paid premium access to. These companies have a duty to be transparent about their policies and practices to those who provide their personal information. There are also the private sector companies who access and use the social media data and who provide access – along with data analytics tools – to police services. These companies currently have very limited transparency obligations. They are responsible to their shareholders, but not to a broader public, and they have no direct commercial relationship with the people who provide the data that is used in the analytics. Their own records are largely protected from scrutiny as confidential commercial information.

---

[108] Ozer, *supra* n. 25.

The problem is also complicated by the nature of police activity. While the transparency of public actors is an important democratic value, it is also recognised that police activities and investigations may require a certain amount of secrecy in order to be effective. The level of secrecy afforded to police and law enforcement agencies has been increasing since the 9-11 terrorist attacks. Legislation has since been enacted in a number of countries, such as the US and Canada,[109] which makes it easier for authorities to gather data and to pursue investigations with little or no public scrutiny and often fairly limited oversight.

Another challenge lies in the ease with which vocabulary can be used as a barrier to transparency. As discussed above, social media companies do not permit the use of their data for "surveillance" purposes, and many intermediaries continue to explain that they provide services, not for surveillance, but for "public safety" or "law enforcement".[110] It must be recognised that there is considerable ambiguity in these terms. After all, surveillance is often a part of law enforcement, and controlling protests and demonstrations can be linked to public safety. Even after the ACLU disclosures therefore, it remains difficult to identify the precise nature of the uses being made of social media data by police services. This may be an instance where vocabulary and communications change but the underlying activity does not.

Vocabulary is a challenge for transparency in other respects as well. For example, the Brennan Center for Justice flags as problematic the fact that "[w]ithout informing city government or the public, the Seattle Police Department paid $12,900 to software company Geofeedia, violating a provision mandating City Council approval for any purchase of surveillance

---

[109] See, e.g., USA PATRIOT Act, Public Law 107-56 (2001); *Anti-Terrorism Act*, S.C. 2001, c. 41.

[110] For example, Snaptrends indicated that it was no longer marketing its services for law enforcement and policing, but that it would continue to provide services oriented towards "public safety". See Utt, *supra* n. 15.

equipment."[111] While the contract with Geofeedia may be understood as problematic in light of the ACLU revelations, and while it is clear that the Seattle City Council sought to maintain some level of oversight over police surveillance activities, it is not clear that a data analytics service falls within the definition of "surveillance equipment". In fact, it is unclear how any given municipality would characterise data analytics services for procurement purposes, and it might differ from one jurisdiction to another. This may make it more difficult for those seeking transparency to do so through public record access requests.

In spite of these complicating factors, some transparency/oversight mechanisms have long been part of the legal framework governing such activities. For example, the requirement for police to seek warrants or production orders for data with privacy implications is long-standing. In some instances, the failure to follow proper procedure and to obtain a warrant will lead to the exclusion of evidence and the acquittal of accused persons. Yet although constitutional rights claims in individual prosecutions may test specific data gathering activities of police, policing based on social media analytics is more difficult to address through these means. This is in part because the data used in the analytics might be categorised as "public" rather than "private" or, at the very least, it because the data subject has provided consent for analytics use of such data. In addition, the harm that flows from this type of surveillance may implicate more than due process and privacy rights; it may touch on other values such as the freedoms of expression and association and the right to be free from discrimination.[112] These harms are more collective than individual; the traditional model of judicial privacy oversight of police activities is focussed on individuals and the reasonable expectations of privacy in specific contexts.

---

[111] Brennan Center for Justice, *supra* n. 31.

[112] *Supra* n. 89, ch. 4.

What is proposed here is a three-part transparency approach: 1) transparency as policy; 2) transparency as oversight; and 3) transparency in redress. These transparency measures apply both to the private sector and public-sector companies involved.

10 Transparency as policy

Three levels of transparency via policy are required. The first is police transparency; the second is the transparency of platforms vis-à-vis their user base; and the third is the transparency of platforms vis-à-vis their developer communities.

10.1 Police transparency

The public must be made aware of how the information they share with social media platforms is further shared and used for law enforcement purposes. This requires police services that use social media data analytics to be transparent about these uses by posting clear policies articulating their approach to social media in policing. While many police services have general social media policies, many of these do not address data analytics and relatively few of these policies are publicly available. For example, the Brennan Center for Justice reported that only 18 of the 157 jurisdictions it canvassed had publicly available social media policies.[113]

Social media policies should be posted on police service websites and must be easily accessible by the public. They should not be deeply buried within the site, or listed under obscure headings. They must also be updated as techniques and technologies evolve. Not only does this give the public clear notice of how their published personal information may be used by local,

[113] Brennan Center for Justice, *supra* n. 31.

regional and national police services, but also the policies will articulate norms of conduct against which investigatory practices in specific instances can be measured. The International Organisation of Police Chiefs has recommended that police services develop social media policies, and has provided some examples of policies used by particular police services.[114]

10.2 Social media platform transparency

Social media companies must also be transparent with their user base about how the personal information which they collect as well as all user-contributed content may be used for law enforcement purposes. The privacy policies of these companies are relatively explicit about the possibility that information about specific users may be shared with law enforcement or national security officials where there are requests made for such information, or where there is a court order that requires its disclosure.[115] They are generally clear that the data they collect or that is provided by users can be shared with other companies, including third party app developers. These policies, however, must connect the dots between general use in analytics and the potential for law enforcement use of analytics. The platform privacy policies should clearly articulate which law

---

[114] Global Advisory Committee, *supra* n. 43. The sample policies are found in the Appendices to the report.

[115] For example, Twitter provides in its Privacy Policy that "we may preserve or disclose your information if we believe that it is reasonably necessary to comply with a law, regulation, legal process, or governmental request […]" (Twitter, "Privacy Policy", *supra* n. 52.) Facebook's Policy provides: "We may access, preserve and share your information in response to a legal request (like a search warrant, court order or subpoena) if we have a good faith belief that the law requires us to do so. This may include responding to legal requests from jurisdictions outside of the United States where we have a good faith belief that the response is required by law in that jurisdiction, affects users in that jurisdiction, and is consistent with internationally recognized standards. We may also access, preserve and share information when we have a good faith belief it is necessary to: detect, prevent and address fraud and other illegal activity; to protect ourselves, you and others, including as part of investigations; or to prevent death or imminent bodily harm." (Facebook, "Data Policy" (2016), available at https://www.facebook.com/policy.php (accessed 14 June 2017).)

enforcement purposes are supported and which are not.

10.3 Transparency with developers

In its letters to social media companies following the publication of its report, the ACLU set out specific recommendations for those companies. These recommendations demanded transparency between the platforms and developers who make use of the data. Thus, the ACLU recommended that the companies adopt "clear and transparent public policies that prohibit developers from using [social media] data to facilitate surveillance".[116] They recommended that "[t]hese policies should also appear prominently in specific materials and agreements with developers."[117]

The major social media companies tend to now be quite explicit with their developer communities that information provided to them should not be shared for "surveillance" purposes.[118] However, in order for these policies to be truly transparent – for developers and for the public who may seek to understand the parameters of acceptable use of data by developers – it is necessary for these policies to define key terms such as "surveillance". The language used should also be clear and transparent. For example, Twitter's policy for developers states:

> You will not knowingly: 1) display, distribute, or otherwise make available Content to any entity to investigate, track or surveil Twitter's users or their Content, or to obtain information on Twitter users or their Content, in a manner that would require a subpoena, court order, or other valid legal

[116] Letter from the American Civil Liberties Union of California to Twitter (10 October 2016), available at https://www.aclunc.org/sites/default/files/20161010_ACLU_CMJ_Color_of_Change_Joint_letter_Twitter.pdf (accessed 14 June 2017).

[117] *Ibid*.

[118] *Ibid*.

> process or that would otherwise have the potential to be inconsistent with our users' reasonable expectations of privacy; […][119]

While this prohibition is a positive step, it is not as clear as it should be. For instance, the limitation that the content must not be provided in a manner that would require a subpoena or court order introduces uncertainty. If user content is considered to be in the public domain, then no court order or subpoena is required to access it. Yet analytics may create a new level of privacy threat that raises an expectation of privacy. The uncertainty is problematic. The catch-all phrase, "or that would otherwise have the potential to be inconsistent with our users' reasonable expectations of privacy", might suffice to capture the use of social media analytics for surveillance, but it is ambiguous.

Developer policies should also contain strong statements barring the use of data for targeting or profiling in ways that are discriminatory. Currently the Twitter Developer Policy provides:

> You will not knowingly […] 2) display, distribute or otherwise make available Content to any person or entity that you reasonably believe will use such data to violate the Universal Declaration of Human Rights (located at http://www.un.org/en/documents/udhr/), including without limitation Articles 12, 18, or 19. […] You will not conduct and your Services will not provide analyses or research that isolates a small group of individuals or any single individual for any unlawful or discriminatory purposes.[120]

This policy could be made clearer by articulating the specific human rights values that should not be infringed rather than by referring to them by number

[119] Twitter, "Developer Agreement and Policy" (2017), art. VII.A, available at https://dev.twitter.com/overview/terms/agreement-and-policy (accessed 14 June 2017).

[120] *Ibid.*; Article 12 of the UNDHR deals with the right to privacy. Articles 18 and 19 deal with the rights to freedom of association and freedom of expression respectively.

with reference to a document hosted on another site. It is worth noting that what is prohibited is the display, distribution or making available of content to someone "that you reasonably believe will use such data" in an inappropriate manner. Since presumably one is permitted to trust that law enforcement officials will use data in a lawful and appropriate manner, these provisions actually do not prevent the provision of content and data analytics for law enforcement purposes. It would seem to be necessary to have first demonstrated that an inappropriate use has taken place before it is no longer reasonable to assume that the use will be appropriate. Thus, while it is important to have such clauses in developer agreements, and while it would be beneficial to make them as clear as possible, they are of limited value on their own. This is why oversight is also an important element of transparency.

11 Transparency as oversight

The second level of transparency is referred to here as transparency of oversight. Oversight is required to ensure that police and social media companies comply with their established policies. Transparency frameworks already exist that can assist in providing transparency in oversight. For example, freedom of information legislation allows journalists, civil society, and the broader public to request public records, including those outlining government expenditures, policies, programs and activities. This transparency mechanism led the ACLU to uncover the extent of police contracting for the services of companies such as Geofeedia. These mechanisms function best when they are relatively quick and inexpensive to use. The area of procurement, for example, calls out for proactive disclosure of data rather than the more cumbersome process of filing access requests. At a minimum, proactive disclosure should identify the companies with whom police services are contracting, the nature of the products or services

provided, and the dollar values of such contracts.

Proactive disclosure is important, but it is not always enough to ensure proper transparency. Where it is made, proactive disclosure can provide a baseline for information upon which access requests can be built. Access requests can reveal details of how policies are (or are not) implemented or complied with. For example, the kind of investigation carried out by the ACLU required freedom of information requests in order to document exchanges between company representatives and police services and in order to understand how the purchased services might be deployed. It was this information that revealed the truly problematic – and more revealing – exchanges of emails regarding the uses to which the services could be put. This is a more time-consuming and cost-intensive investigation.

Oversight can be assisted by requiring public explanations of why data analytic services are procured and for what purposes. City councillors should be pushed to ask such questions in reviewing police expenditures. Public authorities can provide explanations in annual reports, in minutes of city council meetings, or in other contexts that create a public record. The likely future explosion of the use of data analytics services and the potential impacts of some forms of big data analytics may require by-laws, policies, regulations or even laws that mandate privacy impact assessments to be carried out prior to contracting for such services. The privacy impact assessment can in turn serve as a public document that helps provide a measure of transparency in oversight.

As noted earlier, police social media policies should provide a level of transparency regarding how social media data analytics are carried out. This transparency should reveal the oversight mechanisms that are in place. For example, it might be that data analytics can only be used once formal approval is received from a certain level within the chain of command. This approval should be documented, and could include a requirement to specify the objectives of the

analytics and the search criteria being used. Internal record keeping of this kind, even if never fully publicly disclosed, can at least be available for internal audits or reviews by oversight bodies of how data analytics are being used and whether improper considerations enter into the search criteria.

Transparency as oversight can also be facilitated where social media companies make some of their activities more transparent. Emerging voluntary transparency norms for the private sector see a growing number of companies recognising their role in the responsible management of the personal information they hold. Due to increasing public frustration over the government's relatively easy access to data, there has been a growing movement, spurred by civil society towards voluntary transparency reporting.[121] Thus, for example, we see voluntary transparency reporting by a broad range of corporations in sectors where there has been a strong interest demonstrated by public authorities in accessing data. These include telecommunications companies,[122] social media companies,[123] and sharing economy platforms.[124] Faced by increased scrutiny from civil society, these companies have, to differing extents, embraced voluntary transparency reporting. This has focussed on providing information

---

121 For a discussion of voluntary transparency reporting in Canada, see Office of the Privacy Commissioner of Canada, "Transparency Reporting by Private Sector Companies" (2015), available at https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2015/transp_index/ (accessed 14 June 2017). In the US, the Electronic Frontier Foundation publishes transparency reports from major internet-based companies. See Electronic Frontier Foundation, "Who Has Your Back? Protecting Your Data from Government Requests" (2017), available at https://www.eff.org/who-has-your-back-government-data-requests-2015 (accessed 14 June 2017). See also AccessNow, "Transparency Reporting Index" (2017), available at https://www.accessnow.org/transparency-reporting-index/ (accessed 14 June 2017).

122 See, Electronic Frontier Foundation, *supra* n. 121. In Canada, some of the major telecommunications companies now issue transparency reports. See Rogers, "2015 Rogers Transparency Report" (2016), available at http://about.rogers.com/2016/06/27/2015-rogers-transparency-report/ (accessed 14 June 2017); Telus, "Transparency" (2016), available at https://sustainability.telus.com/en/ (accessed 14 June 2017).

123 Electronic Frontier Foundation, *supra* n. 121.

124 See *ibid*.

about individual police requests for access to information in their possession and details on how these are handled by the company. Specifically, it would be helpful to have transparency with respect to any oversight measures in place to ensure that licensees are not misusing the data streams to which they have access, any complaints regarding certain licensees or certain practices, and any steps taken in response to complaints.

Transparency in this context is complicated by the way in which private corporate interests intersect with government activity. Just as the law enforcement and national security agencies have an interest in shielding their activities from scrutiny, so too do private sector corporations have a stake in maintaining secrecy regarding many of their activities and transactions. The private sector urge for secrecy may be motivated in part by a desire to protect confidential algorithms, and in part by a desire to avoid public backlash over services or clients. These interests may coincide in the case of the use of social media data for surveillance purposes.

12 Transparency in enforcement

Developer policies for social media platforms typically rely upon two primary means of enforcement. The first, and in all likelihood, the dominant means, is reactive. Social media companies react once a particular situation is drawn to their attention. In the case of the use of information for improper purposes, awareness of the transgression could come as a result of a formal complaint or after revelations in the media. This is what took place with the ACLU's report on the activities of Geofeedia and similar companies. The media attention given to the explosive report produced immediate responses from social media companies.

Because of the covert nature of the surveillance analytics, the complaints-based model would seem to require either whistleblower involvement (i.e. disclosure of a company's activities by an insider) or the sustained engagement of a civil society organisation such as the ACLU. In ACLU's case, their investigation was extensive and resource-intensive, involving a large volume of access to data requests. Of course, once public authorities and the companies with which they contract become aware that public records disclosure may be used to reveal surveillance activities, the communications between such authorities may become more difficult to track. Phone calls and in-person meetings, for example, may be used instead of email, leaving much less of a trail to uncover.

The second avenue of enforcement is proactive. This would involve monitoring or oversight by the company supplying the data. In the wake of the Geofeedia scandal, a letter from the ACLU to Twitter stated: "Twitter should institute both human and technical auditing mechanisms designed to effectively identify potential violations of company policies, both by the developers and their end users, and take swift action for violations."[125] While there was some indication from Twitter and Facebook that monitoring efforts might be stepped up after the Geofeedia scandal,[126] there was little information about what such

[125] *Supra* n. 116. Note that social media companies have increasingly been called upon to take measures to address problems to which their platforms contribute. This includes addressing fake news (see, e.g., Josh Halliday, "Facebook and Twitter Should Do More to Combat Fake News, Says GCHQ" (The Guardian, 14 March 2017), available at: https://www.theguardian.com/media/2017/mar/14/facebook-twitter-gchq-combat-fake-news (accessed 27 November 2017)) and hate speech (see, e.g., Crispian Balmer, "Top Italian Official Says Facebook Must Do More Against Hate Speech" (Reuters, 12 February 2017), available at http://www.reuters.com/article/us-italy-boldrini-facebook/top-italian-official-says-facebook-must-do-more-against-hate-speech-idUSKBN15R0J1 (accessed 27 November 2017)).

[126] For example, in a statement released after the ACLU report, Facebook indicated it had stepped up its actions to enforce its policies against developers who use its data for surveillance. See Facebook, "Facebook Platform Policy", *supra* n. 97. Twitter announced it would take steps to crack down on developers using its data for surveillance activities. See

monitoring would entail or how effective it would be. While companies could have their own internal monitoring in place to ensure compliance with their policies, it is important to remember that enforcement of policies that limit the uses to which data can be put would be against clients – and likely against the clients for some of the social media companies' most expensive products or services. There is little incentive to go after clients or to deny them access to for-fee services – unless there is a risk of reputational harm to the social media companies should their activities become public. Where monitoring or auditing simply involves asking how information is or will be used, its efficacy will depend upon the forthrightness of the respondent.

More concrete information is required about what measures are put in place by social media companies to audit developers' uses of their data. It may be that it is unworkable for social media companies to share full details of audit measures in order to increase their effectiveness. Yet some level of disclosure should be possible.

13 Conclusion

The 2016 reports by the ACLU and the Brennan Center for Justice cast light on the use of social media data analytics by police services. The outcry raised by these reports had an immediate and significant impact on companies such as Geofeedia, but there is reason to be concerned that broader impacts will not be lasting. This is in part due to the challenges of monitoring and detecting inappropriate uses of social media data analytics within a dynamic and evolving marketplace for such services.

Darren Pauli, "Twitter to Crack Down on Spies Wielding Its APIs" (The Register, 24 November 2016), available at: https://www.theregister.co.uk/2016/11/24/twitter_to_ramp_up_crack_down_on_dev_api_spies/ (accessed 14 June 2017).

Legal responses to the use of social media data analytics discussed here are complicated by the fact that the social media content is made publicly available with the consent of those who contributed the content, and there is also consent to reuse of the content. While it might be possible to argue that there are privacy interests in being free from state surveillance based on data analytics, the focus in this paper has been on the issue of transparency. Without transparency regarding the use of data analytics in state surveillance and monitoring programs, it becomes impossible to know of such activities and to devise appropriate limits and oversight. Further, without transparency, it is impossible to know whether social media data is being used contrary to the stated policies of social media platforms.

This paper has argued that new approaches to transparency are required in order to shed light on police use of social media data analytics and to ensure that such practices remain within acceptable boundaries. It recommends transparency of policy, oversight and enforcement for the main actors: police, social media companies and data analytics intermediaries. These transparency measures are increasingly important as existing privacy laws and paradigms may be insufficient to address surveillance practices with broad impacts, and because the use of social media data analytics increasingly implicates other values such as the freedoms of association and expression and the right to be free from discrimination.

Acknowledgements

I gratefully acknowledge the Social Sciences and Humanities Research Council of Canada for their support of the Geothink project of which this research is a part, as well as the Canada Research Chairs Program. Many thanks to Stephanie

Tadeo for her research assistance, and to Charles Sanders for his comments on an earlier draft of this paper. I am grateful as well to the anonymous peer reviewers for their thoughtful and helpful comments. All errors or omissions are my own.