Note on big data, crime and security:
Civil liberties, data protection and privacy concerns

3 April 2014

Following the publication of the Snowden files and related media stories, it is clear that the main users and adopters of 'big data' approaches amongst state institutions are the security agencies, in particular GCHQ in the UK and the NSA in the US. The extent of surveillance undertaken by these two agencies has now been widely-reported and includes the capturing and storage of vast amounts of internet and telecommunications information (both content and metadata). It appears that this information is subsequently used in operations against known or suspected terrorists, hackers (overlapping and/or confused with online political activists), cyber-security and industrial espionage. However, there is much yet to be uncovered about the role of GCHQ and other state security agencies in their collection of 'big data' and against whom it is used.

Widespread concern has been raised over the indiscriminate, mass collection of data by these agencies, although little has so far been done to rein in their activities. Laws governing the collection and use of data were drawn up in an era before the massive increase in digital information generated by smartphones, computer, tablets, and other devices connected to the internet and other information networks. Hundreds of individuals and organisations across the globe have signed up to a set of principles aimed at laws and regulations to ensure privacy and data protection in the face of mass state surveillance programmes,[1] and there are also specific national campaigns.[2] Although many details of these programmes remain unknown, what seems abundantly clear is their incompatibility with basic human rights standards.

Predictive policing

Police forces are also known to be increasingly attempting to make use of 'big data', in a variety of ways. *Statewatch* reported last year on Kent Police's trial of "predictive policing", in which analytics software is used in an attempt to ascertain areas in which crimes are most likely to occur, based on "several years' worth of crime data and human behaviour to predict the areas in which offences are likely to take place."[3] Greater Manchester Police, West Midlands Police and West Yorkshire have also undertaken trials using the same system, PredPol, designed and first used by US police forces. The Metropolitan Police have also begun using the system.[4]

Use of the technology appears to be more developed in the US. Chicago Police Department, for example, has developed a "heat list". Rather than attempting to predict places in which crime will occur, this system uses information held in police records to generate "an index of

[1] https://en.necessaryandproportionate.org/
[2] Amongst others, UK: https://www.dontspyonus.org.uk/; USA: http://stopwatching.us/
[3] http://database.statewatch.org/article.asp?aid=32180
[4] http://www.ibtimes.co.uk/predictive-policing-predpol-future-crime-509891

the roughly 400 people in the city of Chicago supposedly most likely to be involved in violent crime." These are not necessarily people who have committed crimes in the past, but rather those who have associations or connections with those who have. Miles Wernick, an academic who helped develop the system, has said: "It's not just shooting somebody, or being shot… It has to do with the person's relationships to other violent people." Those who end up on the "heat list" receive a visit from the police to warn them: "if you commit any crimes, there will be major consequences."[5]

Claims are frequently made about the effectiveness of predictive policing systems in reducing crime and bringing offenders to justice, and the statistics look impressive: in August 2013 Kent Police claimed that its use of the PredPol system led to a reduction in street violence in Medway of 6%.[6] Greater Manchester Police reported that burglary rates in Trafford dropped by 26% between May 2010 and May 2011.[7] A report on the effectiveness of Chicago Police Department's "heat list" is due in 2016.

Other data collection methods employed by police forces in the UK include the harvesting of mobile phone data from individuals placed under arrest,[8] the use of IMEI catchers to gather mobile phone data including intercepting SMS messages and phone calls,[9] and even the personal data of people calling 999 to report crimes.[10]

Passenger Name Record

On a wider scale, Passenger Name Record (PNR) systems are increasingly being adopted by states in order to facilitate the surveillance of travellers, currently primarily those travelling by air. PNR systems take travel reservation and booking information (for example from airline tickets) and run them against databases and profiling systems in order to ascertain whether an individual is or may be involved in terrorism or serious crime. The policing element of the UK's e-Borders system is an example of this, and the EU currently has agreements with Australia, Canada and the USA which see all individuals travelling from the EU to those countries screened. There is also a proposed EU PNR system on the table (although negotiations have stalled),[11] which would at a minimum cover flights coming into the EU. Some Member States, including the UK, are in favour of such a system also covering internal flights along with sea and rail transport.[12]

Terrorist Finance Tracking Program

Financial data is also the subject of state surveillance. Under the EU-US Terrorist Finance Tracking Program (TFTP), for example, data on vast numbers of international bank transfers are passed to the US Treasury for analysis and assessment. Europol is responsible for managing data transfers from the EU to the US, and in March 2013 the agency's Joint Supervisory Board (JSB) said:

[5] http://www.theverge.com/2014/2/19/5419854/the-minority-report-this-computer-predicts-crime-but-is-it-racist
[6] http://www.bbc.co.uk/news/uk-england-kent-23689715
[7] http://www.ibtimes.co.uk/predictive-policing-predpol-future-crime-509891
[8] http://www.bbc.co.uk/news/technology-18102793
[9] http://www.theguardian.com/uk/2011/oct/30/metropolitan-police-mobile-phone-surveillance
[10] http://www.bbc.co.uk/news/uk-12104215
[11] http://www.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2011/0023%28COD%29&l=en
[12] http://database.statewatch.org/article.asp?aid=29118

> "For the avoidance of doubt, the JSB repeats that, in view of the nature of the TFTP and the scope of the agreement there is a massive, regular, data transfer from the EU to the US.
>
> "There is a clear tension between the idea of limiting the amount of data to be transmitted by tailoring and narrowing the requests and the nature of the TFTP.
>
> "These are political issues and it is up to the legislators to balance the transfer of massive data sets – mostly of non-suspects – with proportionality."[13]

MEPs have recently called for suspension of the TFTP due to revelations about the activities of the US National Security Agencies.[14]

Data Retention Directive

Another piece of European legislation enhancing the 'big data' capabilities of law enforcement agencies is the highly controversial Data Retention Directive, the lawfulness of which will be judged by the European Court of Justice (ECJ) on 8 April.[15] It has already been condemned by an Advocate-General of the ECJ in a non-binding opinion delivered as part of the case.[16] This mandates the collection and storage by telecommunications firms of data generated through the use of landlines and mobile phones, fax machines, and the internet, "in order to ensure that the data are available for the purposes of the investigation, detection and prosecution of serious crime".[17] Companies are obliged to retain a wide range of data:

- The source of a communication;
- The destination of a communication;
- The date, time and duration of a communication;
- The type of a communication;
- Users' communication equipment or what purports to be their equipment; and
- The location of mobile communication equipment.

The UK has consistently been amongst one of the highest-ranked countries in terms of requests from law enforcement authorities for telecommunications data. In 2012, 570,135 requests were made.[18]

However, while the Data Retention Directive was passed in 2006, as far back as 1997 BT met "ever-increasing demands for details of customers' calls by installing an automated computer-to-computer 'interface' to feed call information out."[19] Attempts by state agencies to collect and make use of vast amounts of data are not new; however, the technology available with which to gather and process data has advanced significantly in recent years.

---

[13] http://www.statewatch.org/news/2013/apr/eu-europol-jsb-tftp-assessment.pdf
[14] http://www.europarl.europa.eu/news/en/news-room/content/20131021IPR22725/html/MEPs-call-for-suspension-of-EU-US-bank-data-deal-in-response-to-NSA-snooping
[15] Digital Rights Ireland & Seitlinger and Others, http://curia.europa.eu/jcms/jcms/Jo1_6581/?dateDebut=8/04/2014&dateFin=8/04/2014
[16] http://www.statewatch.org/news/2013/dec/eu-ecj-advocate-general-opinion-data-retention-C-293-12.pdf
[17] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:105:0054:0063:EN:PDF
[18] http://www.statewatch.org/uk-tel-tap-reports.htm; figures for a number of EU Member States can be found in the Appendix to a recent Statewatch analysis of the Data Retention Directive: http://www.statewatch.org/news/2013/dec/secile-data-retention-directive-in-europe-a-case-study.pdf
[19] http://www.duncancampbell.org/menu/journalism/guardian/Cops.pdf

Civil liberties, privacy and data protection concerns

There are a range of civil liberties, privacy and data protection issues raised by these methods. *The Economist* has said of predictive policing systems:

> *Misuse and overuse of data can amplify biases. It matters, for example, whether software crunches reports of crimes or arrests; if the latter, police activity risks creating a vicious circle. And report-based systems may favour rich neighbourhoods which turn to the police more readily rather than poor ones where crime is rife. Crimes such as burglary and car theft are more consistently reported than drug dealing or gang-related violence.*[20]

Similarly, the "heat list" used by Chicago Police could perpetuate or amplify bias through its targeting of those who have contact with convicted criminals. Hanni Fakhoury of the Electronic Frontier Foundation has said of this method "Are people ending up on this list simply because they live in a crappy part of town and know people who have been troublemakers?" Given the over-representation of black and ethnic minority individuals in the criminal justice system,[21] there is a clear possibility for the perpetuation of racist and discriminatory policing through the use of such systems.

Such methods also appear to turn the principle of 'innocent until proven guilty' on its head. The harvesting of data to try and ascertain who is likely to commit a crime places individuals who have done nothing wrong under suspicion. The same problem is inherent in PNR systems, through which every traveller is screened for involvement in or association to crime and terrorism, regardless of the likelihood of their actual involvement. The Data Retention Directive also has the same effect: to place everybody under suspicion, with details of their communications stored 'just in case' they are required.

In terms of data protection, the increasing drive to make available to law enforcement information collected for commercial purposes (as with PNR, telecommunications data retention, and the TFTP), the key data protection principle of purpose limitation is diminished. This stipulates that data should not be used for purposes other than that which it is initially collected. The gathering of commercial data by law enforcement and security agencies without a clear legal basis clearly violates legal principles. The passing of legislation to permit the gathering and use of data collected prior to the enactment of the law (as has happened with Eurodac, the EU-wide database of asylum-seekers' fingerprints)[22] clearly offends the principle of purpose limitation. Those whose fingerprints were taken by the authorities prior to recent legislative amendments were unable to consent to the potential use of their data by police forces. Even after laws have been passed permitting law enforcement agencies to gather and use data collected for non-law enforcement purposes, questions over proportionality remain, especially in a society in which more and more personal data is potentially available for use by state agencies.

It is clear that harnessing and analysing vast data sets may simplify the work of the police. However, this in itself is not a justification for their use. There are all sorts of powers that could be given to law enforcement agencies, but which are not, due to the need to protect individual rights and the rule of law – effectiveness should never be the only yardstick by which law enforcement powers are assessed. The ends of crime detection, prevention and

[20] http://www.economist.com/news/briefing/21582042-it-getting-easier-foresee-wrongdoing-and-spot-likely-wrongdoers-dont-even-think-about-it
[21] https://www.justice.gov.uk/statistics/criminal-justice/race
[22] http://www.europarl.europa.eu/news/en/news-room/content/20130422IPR07522/html/MEPs-back-deal-with-Council-on-police-access-to-asylum-seekers%E2%80%99-fingerprints

reduction cannot in themselves justify the means of indiscriminate data-gathering and processing.

Furthermore, the drive towards such systems may lead to overlooking the fact that enhancing the repressive capabilities of law enforcement agencies does nothing to solve the complex socio-economic causes of crime. To quote *The Economist* again:

> *Predicting and forestalling crime does not solve its root causes. Positioning police in hotspots discourages opportunistic wrongdoing, but may encourage other criminals to move to less likely areas. And while data-crunching may make it easier to identify high-risk offenders—about half of American states use some form of statistical analysis to decide when to parole prisoners—there is little that it can do to change their motivation.*[23]

Corporate data-gathering and use

Finally, it is worth noting that the 'big data' systems and software employed by law enforcement and security agencies are frequently developed by and adapted from the private sector, where their use is far more advanced.

As far back as 2005, the collection and processing of vast amounts of data by credit ratings agencies was highlighted by the journalist Jon Ronson.[24] Catherine Crump and Matthew Harwood have recently written at some length about the ongoing development of big data, the 'Internet of Things', and the potential "surveillance of everything".[25] Future developments in the field of big data will almost certainly be driven primarily by the interests of companies, with state agencies subsequently attempting to adopt either the methods used by these companies, or the data they collect. Evgeny Morozov has argued that data collection about one group of individuals is likely to lead to detrimental effects for other groups:

> *[M]y decision to disclose personal information, even if I disclose it only to my insurance company, will inevitably have implications for other people, many of them less well off. People who say that tracking their fitness or location [through their smartphone, for example] is merely an affirmative choice from which they can opt out have little knowledge of how institutions think. Once there are enough early adopters who self-track—and most of them are likely to gain something from it—those who refuse will no longer be seen as just quirky individuals exercising their autonomy. No, they will be considered deviants with something to hide. Their insurance will be more expensive.*[26]

While law enforcement and security agencies have been granted access to more and more datasets (telecommunications data, travel data, identification data, financial data etc.), the security risk posed by individuals is just as likely to be assessed by private organisations, data aggregators and vetting agencies than the state. Big data-bases have long been established, for example, by:

(1) the insurance industry (shared fraud data), financial services industry (e.g. SWIFT), travel industry (shared reservation databases), telecommunications firms (shared customer data and network usage), ISPs (cyber-security) etc.;
(2) regulatory bodies such as the Charity Commission, Companies House and many other registers of professionals;

[23] http://www.economist.com/news/briefing/21582042-it-getting-easier-foresee-wrongdoing-and-spot-likely-wrongdoers-dont-even-think-about-it
[24] http://www.theguardian.com/money/2005/jul/16/creditcards.debt
[25] http://www.tomdispatch.com/blog/175822/
[26] http://www.technologyreview.com/featuredstory/520426/the-real-privacy-problem/

(3) law enforcement and security agencies (Criminal Records Bureau, DVLA, e-borders etc.);
(4) government service providers (welfare, health, e-gov etc.); and
(5) private companies specialising in risk assessment and vetting services such as World-Check, Lexis-Nexis and Experian as well as a new generation of big data analysts such as Detica.

In this context, for most people it may matter less what information has been collected by state agencies directly and more about the routine decisions taken about their 'digital twin(s)' (the profile(s) generated by private firms) by myriad private actors. It is less a case of Big Brother / Little Brother than the emergence of a series of "apps" and "plug-ins" that allow private entities and governments alike to assess risk or conduct due diligence: credit checks, employment checks, CRB checks, social network analysis, behavioural analysis, risk profiling, and so on – anything that turns big data into "actionable intelligence". Since this transformation is inevitably based on statutory or arbitrary interpretations of who or what constitutes a "risk" or "threat", the risk of discrimination against certain groups and denial of individual political, economic and social rights may be increasing with the emergence of every new application/service.