HOUSES OF PARLIAMENT
PARLIAMENTARY OFFICE OF SCIENCE & TECHNOLOGY

POSTNOTE

Number 470 July 2014

Big Data, Crime and Security

Recent technological advances in processing and analysing large and complex data offer new opportunities and challenges for police and security agencies. This POSTnote examines the use of large and varied data in three key areas: crime prevention, crime detection and national security. It also covers regulatory issues and public perception about privacy, civil liberties and social benefits.

Overview

- Big data is being used by police forces to identify locations more likely to experience crime and by HMRC to detect tax fraud.
- Analysing big data could provide police and security agencies with additional tools to predict and detect crime; however there is little evidence on the effectiveness of particular applications.
- Big data technologies enable bulk collection and analysis of electronic communications data. There is debate about the legality, necessity and proportionality of this.
- Public support for the use of personal data is likely to vary depending on the specific use and the perceived risks and personal and social benefits associated with the use.

Background

Law enforcement agencies (LEAs) and security agencies (Box 1) routinely collect large amounts of data in the course of their work preventing and detecting crime and gathering intelligence. The development of electronic communications has led to a large increase in the amount and type of data available about people and their activities that can be used by these agencies. The proliferation of large and complex structured and unstructured data (see below) is known as 'big data'. Advances in computer power, combined with new technical and methodological approaches to capture, process and analyse large and complex data ('big data analytics'), are opening up new opportunities to use big data to gain insights into criminal activity and to use resources more efficiently.[1]

Current and 'Big' Approaches to Data

How data are collected and stored typically varies between local police forces. To address this, a number of national databases have been established to enable sharing of records and intelligence between local forces and with national agencies such as the National Crime Agency (Box 2). These databases contain large quantities of information including:

- **Structured data**: information that follows a set format, such as the location and type of crime reported, DNA profiles, or personal details of an individual who has been arrested or charged.
- **Unstructured data**: text that does not follow a set format, including police and witness statements. Context can be more important to extract meaning from these data.

Currently these systems will only return information based on a user's search terms. For example, whether a particular name is associated with any convictions or whether DNA from a crime scene matches a profile in the database.

Box 1. UK Law Enforcement and Security Agencies

Law Enforcement Agencies (LEAs)

- **Local crime agencies**. These include the 43 territorial police forces across England and Wales, and a single Police Service in each of Northern Ireland and Scotland.
- **National bodies.** These include special police forces such as British Transport Police, and the National Crime Agency (NCA), who deal with serious and organised crime.

Security Agencies

- **Security Service (MI5).** MI5 is responsible for gathering intelligence within the UK to protect against threats such as terrorism and espionage.
- **Secret Intelligence Service (MI6).** MI6 is responsible for gathering secret information outside the UK to support national security and the economic well-being of the UK.
- **Government Communications Headquarters (GCHQ).** GCHQ is responsible for monitoring electronic communications for national security, military operations and law enforcement activities. Provide advice and assistance on protecting the Government's communication and information systems.

Box 2. National Computer Databases

A number of databases are run by the Home Office to share information between local police forces and national agencies:

Police National Computer (PNC)

This was established in the mid-1970s and contains several databases containing criminal records, such as convictions, cautions, final warnings and reprimands. The PNC is primarily used as a record keeping tool to enable national checks of criminal records.

National DNA Database and IDENT1 Fingerprint Database

These databases hold electronic records of DNA profiles and fingerprints taken from arrested individuals or collected from crime scenes. In March 2013 the National DNA Database held 6,737,973 profiles from individuals, and 428,634 profiles from crime scenes.[2]

Police National Database (PND)

This was established in 2011 to enable intelligence to be shared nationally. The software used can convert disparate methods of recording data into a format that can be used by the PND. Data from a force's local database are automatically uploaded to the PND.

Big data analytics can be used to process and analyse these structured and unstructured data automatically to identify patterns or correlations.[3] Advanced computer software can also be used to link big data in internal datasets with each other or with other datasets, such as publically available data from social media (see POSTnote 460). Big data analytics can then be used to look for new insights. These patterns and correlations can be used to highlight areas for further investigation, give a clearer picture of future trends or possibilities and target limited resources. Technical and organisational barriers to maximising use of big data are covered in POSTnote 472.[4]

There is currently limited evidence on how effective these approaches are because they are relatively new. It can also be difficult to systematically compare new approaches with existing techniques or tactics, because of a lack of information about what techniques are currently used and how they are implemented in particular police forces. The evidence base may increase as a What Works Centre for Crime Reduction, hosted by the College of Policing, was launched in 2013 to provide access to evidence about what works in policing and crime reduction.[5]

Crime Prevention

Data are increasingly used by LEAs to map out crime as it occurs and is reported.[6] For example, controllers of police vehicles and ambulances use computer systems designed to capture, store and analyse geographical data (GIS data). These are used to identify the crime incident location and the closest emergency personnel who are able to respond.[7] GIS can also store historical information and be used to look for incident patterns and black spots. Big data analytics can potentially exploit data further; enabling LEAs to target resources to areas where crime is more likely to occur or informing strategic planning.

Predictive Policing

Forecasting Crime Location 'Hotspots'

Certain types of crime, such as burglary, street violence, theft from vehicles and anti-social behaviour have elements of regularity in their occurrence. For example, when a house is burgled there is an increased risk of nearby houses being burgled for a short time afterwards.[8,9,10] Algorithms (a series of calculations) can be developed that use these elements of predictability to forecast 'hot-spots' where the probability of crime will be greater, relative to the surrounding area, based on historic location data. Up to date hot spots can be generated as often as new crime reports are added to the computer system. This information can be used to inform police decisions about which areas to visit on foot patrol. Software designed for this purpose was used to forecast the locations of burglaries over a 7-day period across areas of the East Midlands as part of a Home Office pilot project. Its projections were accurate in 78% of cases, compared to 51% accuracy using traditional techniques that rely on extrapolating historic patterns into the future.[11] Kent and West Yorkshire Police are two forces using predictive policing to forecast crime location hot-spots in the UK in an operational context (Box 3).[12]

There is limited evidence of the effectiveness of predictive policing, but an analysis of several studies into geographically-focused policing initiatives suggests that hot-spot policing does not displace crime to other locations but in fact has a diffusive effect, acting as a deterrent in areas surrounding the hot-spot.[13] In the US, the use of predictive policing by the Los Angeles police force has been reported to have reduced crime by up to a quarter, although these figures have not been independently verified.[14] Further, differences in policing culture, such as the preference for car patrols in the US and foot patrols in the UK, mean that large drops in crime may be unlikely in the UK.

Statewatch, a non-profit organisation that monitors civil liberties, has expressed concern that predictive policing could amplify existing biases in reported crime data. This is because certain crimes are more likely to be reported to the police and certain socio-economic groups are more likely to report crimes. This could lead to a biased picture of where crimes are taking place, leading to forecasts that favour these locations. Police deployment to those areas could then create a cycle of biased crime reports and predictions.[15] However, police forces note that although predictive policing can identify hot-spots more quickly and frequently than manual approaches, it builds on existing techniques and is not used in place of police officer judgement about where to allocate foot patrols.

Estimating Individual's Risk of Crime

Pattern analysis of multiple data sources can create a fuller picture of a suspect's activities around the time of a crime, or be used to make predictions about possible future crimes. No LEAs in the UK have reported using big data analytics to estimate risk at the individual level to date. Using big data analytics to target individuals rather than geographical areas is more controversial, because it could lead to discrimination against individuals who share particular characteristics with people involved in crime, but who are not, and may never be, involved in crime themselves. Datasets may also be incomplete or inaccurate leading to misleading results.[16]

Box 3. Predictive Policing Case Studies

PredPol

PredPol, a predictive policing software package, has been deployed by Kent Police since early 2013. Twice a day it automatically analyses real-time recorded crime location data about burglary, street violence, theft from vehicles and anti-social behaviour, supplemented with historic location data from the past five years. PredPol produces a map of 500 square foot 'hot-spots' where there is a higher probability of crime taking place relative to other local areas over the next 12 hours. Officers use these maps to incorporate hot-spots into their daily patrols. Kent Police force is positive about its use and reports that since its implementation there have been small drops in crimes of this type and anti-social behaviour.

Prospective Mapping

West Yorkshire Police have developed their own software, based on the work of researchers at University College London, to forecast hot-spots for burglary and theft from vehicles. It produces analyses once every two days, based on crime location data about burglary over the past three weeks. Officers patrol hot-spots more heavily and provide advice on crime prevention to residents living in close proximity to a house that has been burgled. A peer-reviewed evaluation of a similar approach used by Greater Manchester Police found that there was a greater reduction in crime in areas where the technique was used compared to similar areas where it was not used.[17]

In the US, some police forces have used big data analytics to estimate the probability that a particular person will be involved in criminal activity. For example, in Philadelphia, police used big data analytics to predict parolee's risk of re-offending and used this to inform decisions about levels of supervision.[18,19] In Chicago, police are piloting a program involving officers visiting the homes of individuals they identify as likely victims or perpetrators of crime, based on the use of big data analytics. However, the approach has prompted concerns around privacy and racial profiling.[20]

Informing Strategic Planning

Big data analytics could be used for strategic planning by LEAs in the medium to long term. For example, the pan-European project, ePOOLICE, is developing a prototype system linking police data to social media data to identify new correlations and highlight emerging crime trends in cybercrime, human trafficking and drug trafficking.[21] Researchers at the University of Strathclyde are using public data and crime location data to look at how weather, lighting, location of street furniture and traffic flows affect crime levels. This could be used to influence the design of the built environment to discourage and reduce crime.[22]

Crime Detection

Some types of criminal investigations, such as fraud, involve dealing with increasingly large amounts of data. Other types of police work require data to be shared between international crime and security agencies, for example in combating organised criminal gangs operating across the EU.[23] However, as well as technical difficulties around sharing data collected and stored in different formats there are complex legal frameworks for sharing personal data across different countries (Box 4).[24,25] This means that data are typically shared only on a case-by-case basis for a specific purpose. Advanced computer software and big data analytics could make these processes more efficient and help to detect or solve cases by highlighting new connections or patterns. Two examples are given below.

Financial Crime

Financial transactions, such as electronic bank transfers and credit card purchases, generate a large amount of data. Some transactions may indicate fraud or money laundering. However, detecting potentially criminal activity is difficult because of the vast amount of data collected. In addition, suspicious patterns of behaviour often only emerge when a number of disparate pieces of data are connected. To increase its ability to detect tax fraud and evasion, HM Revenue and Customs (HMRC) has implemented a big data system called Connect. This has allowed HMRC to bring together and analyse the majority of its internal data (over 1 billion pieces) to find patterns and connections. As of April 2013, HMRC reported that, with an initial investment of £45 million (including five years running costs), it has been able to recover £2.6 billion through Connect because:[26]

- simple searches can visually present related information in minutes, which could previously have taken months
- complex analysis is not confined to one dataset or limited by the skills of individual analysts, rather the same analysis can be applied to all data across the UK to deliver consistency of customer handling
- profiles can be built of data patterns indicative of a certain type of crime, which can then be used to identify cases that may warrant further investigation.

Sharing Ballistics Data

In 2011, Project Odyssey ran as a research project to create a system for sharing ballistics data between LEAs across Europe. The software enabled Member States to share data more effectively by converting and comparing data from the different ballistic systems and standards used across Europe, without compromising the security of personal data held on national databases. The system was successful but it ceased operating at the end of the project, in part because of a lack of central leadership, co-ordination and funding.[27]

Box 4. Key Legislation Relevant to Sharing Criminal Information

Data Protection Act 1998 (DPA)

The DPA implements Directive 95/46/EC and regulates the processing of personal data.[28] It covers data use by LEAs and security agencies but there is a general exemption for national security and qualified exemptions for crime prevention and detection.[29]

European Legislation

Protection of personal data across the EU is based on the Treaty on the Functioning of the European Union, the European Convention on Human Rights, the Council of Europe Convention 108 and the EU Charter of Fundamental Rights.[30] There are a number of EU Instruments that are relevant to the sharing of criminal information and intelligence in the EU, including Directive 95/46/EC and Framework Decisions 2006/960/JHA[31] and 2008/977/JHA[32] as well as the case law of the European Court of Human Rights and of the Court of Justice of the European Union. Data protection reforms are under consideration, including a General Data Protection Regulation, intended to replace Directive 95/46/EC as well as a new General Data Protection Directive, which will provide for data protection in the areas of police and judicial cooperation.[33]

National Security

Security agencies collect intelligence as part of their activities to protect national security. Traditionally, this involves investigating specific individuals and organisations suspected of potential activity seriously damaging to the national interest to obtain information. Intelligence may be collected by covert operations, following or observing targets, eavesdropping or intercepting communications. Modern communication technologies leave behind a large 'data footprint', such as the 'who, where, how and when', as well as the actual message. Big data analytics provides the capability to store and analyse vast amounts of electronic communications data, in order to identify patterns or connections that may indicate suspicious behaviour.[34]

Bulk Data Collection

In May 2013, Edward Snowden released classified documents alleged to show that the UK Government Communications Headquarters (GCHQ) (Box 1) had been bulk collecting data about internet communications and storing it for 30 days under a program called 'Tempora'. In January 2014, an inquiry by the European Parliament Committee on Civil Liberties, Justice and Home Affairs (LIBE) reported that there was evidence that security agencies in the US and UK had been collecting, storing and analysing communication and location data 'on an unprecedented scale and in an indiscriminate and non-suspicion-based manner'.[35]

Several governments have argued that security agencies need to have timely and accurate intelligence in order to be able to protect national security. In the UK, agencies activities are subject to Article 8 of the European Convention on Human Rights (ECHR), which protects the right to privacy. It is a qualified right, which means that interference can be justified in particular circumstances where it is lawful, proportionate and necessary.[36] However, some non-governmental groups, including Liberty and Big Brother Watch, consider the alleged bulk collection of data to be an unjustified intrusion into the privacy of individuals who are not under suspicion. A number of inquiries are examining these allegations and the legality, proportionality and necessity of such programmes. These include the UK Parliament's Intelligence and Security Committee (ISC)[37] and Home Affairs Select Committee,[38] the EU's LIBE Committee and the Royal United Services Institute.[39]

Legislation and Oversight

In the UK, monitoring of communications is governed by a number of pieces of legislation which interact, including the 2000 Regulation of Investigatory Powers Act (RIPA), the Intelligence Services Act 1994 and Article 8 of the ECHR (see POSTnote 436).[40] There is ongoing debate as to the effectiveness of RIPA considering how methods of communications have changed since 2000. A case brought in January 2014 against the UK Government at the European Court of Human Rights is currently pending. It alleges that the use of powers under RIPA by GCHQ is not compatible with Article 8 of the ECHR.[41] However, the 2013 annual report by the Interception of Communications Commissioner states that RIPA still offers adequate safeguards for privacy and was designed to be technology neutral.[42] There is also debate as to whether the current oversight mechanisms are credible and effective.

In April 2014 the European Court of Justice struck down the 2006 EU Data Retention Directive that required communication service providers to retain communications data for up to two years on the grounds that it entails a serious interference with the fundamental rights to respect for private life and to the protection of personal data.[43] In July 2014, the Government announced emergency legislation, which it stated was necessary to retain existing powers in the UK. The Data Retention and Investigatory Powers Act was passed by Parliament in July 2014.[44] It contains a sunset clause, which means that the laws will lapse at the end of 2016. The Act committed the Government to complete a review of investigatory powers, including RIPA, before May 2015.[45] Some civil liberties groups are seeking a judicial review of the Act.

Privacy, Civil Rights and Social Benefits

Research on public views of big data in the UK suggests that awareness of data collection and use by government and companies is quite high, but that the level of understanding of what this means in practice is much lower. It also suggests that perceptions are likely to change as understanding about potential implications increases.[46] There is limited research specifically on the use of big data for law enforcement and intelligence.[47,48] In general, concerns appear to centre on the use of personal data and a potential loss of privacy or control over how their data are used. People seem to be more willing to trade-off personal privacy concerns when data cannot be used to identify and target particular individuals and when they perceive personal or social benefits from the use.[49]

In the US, a review of big data and privacy by the Executive Office of the President reported that big data technologies can provide effective tools for LEAs and security agencies and contribute to the public good. However, it also stressed that big data poses difficult questions about how to protect personal privacy and civil rights, ensure fairness and prevent discrimination.[50] For example, big data approaches can help to catch criminals and protect national security but they can also focus scrutiny on particular individuals with little or no human intervention or sweep up detailed personal information about people who are not subjects of an investigation. The review concluded that big data had the potential to transform every sphere of life, and that citizens should be involved in shaping the policies and laws to govern big data in a way that protects their core values.

Endnotes

See overleaf.

POST is an office of both Houses of Parliament, charged with providing independent and balanced analysis of policy issues that have a basis in science and technology. POST is grateful to Daniel Rathbone for researching this briefing, to EPSRC for funding his parliamentary fellowship, and to all contributors and reviewers. For further information on this subject, please contact the co-author, Dr Abbi Hobbs.

1 Olesker, A. (2012). *Big Data Solutions for Law Enforcement*, CTO Labs.
2 Home Office, *National DNA Database Strategy Board Annual Report 2012-13*.
3 SAS Software. (2013). *Managing the Intelligence Life Cycle: A More Effective Way to Tackle Crime*.
4 POSTnote 472, *Big and Open Data in Transport*. To be published July 2014.
5 College of Policing (2014). *What Works Centre for Crime Reduction*.
6 Bang, M. and Weirs, R. (2007). *The use of Geographic Information Systems by crime analysts in England and Wales*, Home Office.
7 Ordnance Survey (2014). *GIS in Use*.
8 Bowers, K. *et al* (2004). Prospective Hot-Spotting: The future of crime mapping? *British Journal of Criminology*, 44: 641.
9 Johnson, S. (2008). Repeat burglary victimisation: a tale of two theories. *Journal of Experimental Criminology*, 4: 215-240.
10 Bowers, K. and Johnson, S. (2010). Permeability and burglary risk: are cul-de-sacs safer? *Journal of Quantitative Criminology*, 26: 89-111.
11 Johnson, S. *et al*. (2007). *Prospective crime mapping*, Home Office.
12 HMIC. (2014). *Policing in austerity: Rising to the challenge*. Compendium.
13 Bowers, K. *et al*. (2011). Spatial displacement and diffusion of benefits among geographically focused policing initiatives: a meta-analytical review, *Journal of Experimental Criminology*, 7:347–374.
14 Greengard. S. (2012), Policing the future, *Communications of the ACM*, 55: 3.
15 The Economist, *Predictive Policing: Don't even think about it*, July 2013.
16 Statewatch (2014). *Note on big data, crime and security: Civil liberties, data protection and privacy concerns*.
17 Fielding, M. and Jones, V. (2011). 'Disrupting the optimal forager': predictive risk mapping and domestic burglary reduction in Trafford, Greater Manchester, *International Journal of Police Science and Management*, 14(1): 30.
18 Ritter, N. (2013). Predicting Recidivism Risk: New Tool in Philadelphia Shows Great Promise, *National Institute of Justice Journal*.
19 Barnes, G. and Hyatt, J. (2012), *Classifying Adult Probationers by Forecasting Future Offending*, National Criminal Justice Reference Service.
20 Erbentraut, J (2014). Chicago's Controversial New Police Program Prompts Fears Of Racial Profiling, *The Huffington Post*.
21 European Commission *EPOOLICE project*.
22 University of Strathclyde project on *Using big data analytics and genetic algorithms to predict street crime and optimise crime reduction measures*.
23 House of Commons Home Affairs Select Committee (2013). Ninth Report of Session 2013-14. *Pre-Lisbon Treaty EU police and criminal justice measures: the UK's opt-in decision*.
24 EU Framework Decision 2006/960/JHA.
25 EU Framework Decision 2008/977/JHA.
26 Capgemini Consulting (2013). An interview with Mike Hainey.
27 Yates, S. *et al*. (2011). A platform for discovering and sharing confidential ballistic crime data, *International Journal of Knowledge and Web Intelligence*, 2 (2/3): 202-218.
28 Data Protection Act 1998.
29 Directive 95/46/EC.
30 European Court of Human Rights (2014). Data Protection Handbook.
31 EU Framework Decision 2006/960/JHA.
32 EU Framework Decision 2008/977/JHA.
33 Decision pursuant to Article 10 of Protocol 36 to The Treaty on the Functioning of the European Union, HM Government, July 2013.
34 Pavlin et al. (2013). Exploiting Intelligence for National Security, Ch.15, *Strategic Intelligence Management*, Butterworth-Heinemann.
35 EU Parliament Committee on Civil Liberties, Justice and Home Affairs. (2014), Draft report on the US NSA surveillance programme.
36 European Court of Human Rights (2014). Data Protection Handbook.
37 UK Parliament Intelligence and Security Committee, Press release, May 2014.
38 UK Parliament House of Commons Home Affairs Select Committee (2014). Seventeenth Report of Session 2013–14. *Counter-Terrorism*.
39 Royal United Services Institute (2014) Announcement of Independent inquiry.
40 European Convention on Human Rights.
41 Application No. 58170/13 to the European Court of Human Rights.
42 Rt Hon. Sir Anthony May. (2014), *2013 Annual Report of the Interception of Communications Commissioner*.
43 Court of Justice of the European Union, *Press release, No 54/14*, 8 April 2014.
44 Data Retention and Investigatory Powers Act 2014.
45 Independent Reviewer of Terrorism Legislation (2014). *Investigatory Powers Review*.
46 Sciencewise. (2014). *Big Data: Public views on the collection, sharing and use of personal data by government and companies*.
47 YouGov (2013). *Little appetite for scaling back surveillance*.
48 ComRes (2013). *Big Brother Watch Surveillance Study*.
49 Ipsos Mori. (2014). *Public Attitudes to Science 2014*.
50 Executive Office of the President. (2014), Big Data Seizing Opportunities, Preserving Values.