Royal United Services Institute
for Defence and Security Studies

Occasional Paper

Big Data and Policing

An Assessment of Law Enforcement Requirements, Expectations and Priorities

Alexander Babuta

Big Data and Policing

An Assessment of Law Enforcement Requirements, Expectations and Priorities

Alexander Babuta

RUSI Occasional Paper, September 2017

Royal United Services Institute
for Defence and Security Studies

185 years of independent thinking on defence and security

The Royal United Services Institute (RUSI) is the world's oldest and the UK's leading defence and security think tank. Its mission is to inform, influence and enhance public debate on a safer and more stable world. RUSI is a research-led institute, producing independent, practical and innovative analysis to address today's complex challenges.

Since its foundation in 1831, RUSI has relied on its members to support its activities. Together with revenue from research, publications and conferences, RUSI has sustained its political independence for 185 years.

London | Brussels | Nairobi | Doha | Tokyo | Washington, DC

Published in 2017 by the Royal United Services Institute for Defence and Security Studies.

RUSI Occasional Paper, September 2017. ISSN 2397-0286 (Online); ISSN 2397-0278 (Print).

Printed in the UK by Stephen Austin and Sons, Ltd.

Royal United Services Institute
for Defence and Security Studies
Whitehall
London SW1A 2ET
United Kingdom
+44 (0)20 7747 2600
www.rusi.org

RUSI is a registered charity (No. 210639)

Contents

Acknowledgements

The author would like to thank the police officers and staff who gave up their valuable time to contribute to this project, without whom the research would not have been possible. The author is also very grateful to a number of individuals who provided invaluable comments and insights on an earlier version of this paper, particularly Calum Jeffray, Shane Johnson and Sir Jon Murphy, as well as RUSI colleagues (past and present) Malcolm Chalmers, Clare Ellis, Andrew Glazzard and Cathy Haenlein. The author would also like to thank the RUSI Publications team for their expert editorial input and patience in producing this report, particularly Emma De Angelis, Sarah Hudson, Joseph Millis, Zenab Hotelwala and Charlie de Rivaz. Finally, the author is very grateful to Unisys for helping to make this research possible.

Executive Summary

IN RECENT YEARS, big data technology has revolutionised many domains, including the retail, healthcare and transportation sectors. However, the use of big data technology for policing has so far been limited, particularly in the UK. This is despite the police collecting a vast amount of digital data on a daily basis.

There is a lack of research exploring the potential uses of big data analytics for UK policing. This paper is intended to contribute to this evidence base. Primary research in the form of interviews with 25 serving police officers and staff, as well as experts from the technology sector and academia, has provided new insights into the limitations of the police's current use of data and the police's priorities for expanding these capabilities.

The research has identified a number of fundamental limitations in the police's current use of data. In particular, this paper finds that the fragmentation of databases and software applications is a significant impediment to the efficiency of police forces, as police data is managed across multiple separate systems that are not mutually compatible. Moreover, in the majority of cases, the analysis of digital data is almost entirely manual, despite software being available to automate much of this process. In addition, police forces do not have access to advanced analytical tools to trawl and analyse unstructured data, such as images and video, and for this reason are unable to take full advantage of the UK's wide-reaching surveillance capabilities.

Among the numerous ways in which big data technology could be applied to UK policing, four are identified as key priorities. First, predictive crime mapping could be used to identify areas where crime is most likely to occur, allowing limited resources to be targeted most efficiently. Second, predictive analytics could also be used to identify the risks associated with particular individuals. This includes identifying individuals who are at increased risk of reoffending, as well as those at risk of going missing or becoming the victims of crime. Third, advanced analytics could enable the police to harness the full potential of data collected through visual surveillance, such as CCTV images and automatic number plate recognition (ANPR) data. Fourth, big data technology could be applied to open-source data, such as that collected from social media, to gain a richer understanding of specific crime problems, which would ultimately inform the development of preventive policing strategies.

There are at present a number of practical and organisational barriers to implementing these technologies. Most significantly, the lack of coordinated development of technology across UK policing is highly problematic for big data, which relies on effective nationwide data sharing and collaboration. Financial cuts in recent years have also severely hindered technological development, as the majority of police IT budgets is spent supporting existing legacy systems, with little funding available to invest in new technology. Finally, there are significant legal and ethical constraints governing the police's use of data, although these are not a main focus of this report.

These barriers are by no means insurmountable, and it is expected that, in the coming years, advances in the police's use of technology will enable the successful development of big data policing tools. It is imperative that such development is informed according to specifically identified requirements and priorities, and this report identifies several areas of particular interest that warrant further investigation. Despite the budget cuts of the kind imposed since 2010, it is crucial to invest in new technology, as the costs of the initial investment will be more than recuperated by the efficiency savings made in the long term.

Recommendations

Police Forces, Police and Crime Commissioners

- **Any major development in big data technology should be informed by direct consultation with a representative sample of operational police officers and staff.**

  This has several positive outcomes for both the developer and the end user. From a development perspective, technological investment will be directly targeted to address specific requirements, rather than being based on ad hoc and speculative research. From the police perspective, officer buy-in will be improved both at the front line and within management.

- **UK police forces should prioritise exploring the potentials of predictive mapping software.**

  Predictive hotspot mapping has been shown time and again to be significantly more effective in predicting the location of future crimes than intelligence-led techniques. However, few forces have integrated the practice into current patrol strategies. The police collect a large amount of historic crime data that could be used to predict where crime is likely to occur, allowing limited resources to be directly targeted to where they are most needed.

- **The digital aspects of any serious or long-term investigation should be managed by a digital media investigator, who takes responsibility for becoming the technical lead for specific operations.**

  Almost all police investigations have a significant digital component, but investigations are often conducted without a coherent digital strategy, with officers reporting a lack of coordination between teams working on different types of data. Digital media investigators, which most police forces are yet to fully utilise, can fill this void and represent a potentially highly valuable resource for digital investigations.

- **Analytical tools that predict the risks associated with individuals should use national, rather than local data sets.**

Predictive analytics makes it possible for police forces to use past offending history to identify individuals who are at increased risk of reoffending, as well as using partner agency data to identify individuals who are particularly vulnerable and in need of safeguarding. Analysis of this kind is currently carried out using local police datasets, but the use of national datasets is necessary to gain a full understanding of these risks.

Home Office, College of Policing and the Police ICT Company

- **A national big data procurement strategy should be developed to coordinate technological investment across all UK police forces.**

  The highly localised structure of UK policing has resulted in wide variation in the levels of technological development between different police forces. Forces pursue technological change in isolation, with little coordination at the national level. The Home Office should issue clear national guidance for procurement of big data policing technology to ensure that future investment in this area is not wasted.

- **A standardised glossary of common terminology should be developed for entering information into police databases.**

  When retrieving information from police databases, investigators are required to perform keyword searches, which involves guessing every potential synonym for a particular topic of interest. This makes it particularly difficult to collate and cross-reference information collected from different sources. A standardised lexicon would address this issue, while also enabling the development of useable text-mining software.

- **Shared MASH (Multi-Agency Safeguarding Hub) databases should be created to allow for better data sharing between the police and partner agencies.**

  Local authorities, social services and the police should collaborate closely when identifying vulnerable individuals in need of safeguarding. Shared MASH databases would facilitate this while also giving the police quick access to information that could prove vital for ongoing investigations. At present, data-sharing deficiencies mean that the police's understanding of vulnerability is somewhat one-dimensional.

- **A clear decision-making framework should be developed at the national level to ensure the ethical use of big data technology in policing.**

  There is currently no clear decision-making framework governing the ethical use of big data technology by public sector organisations in the UK. This must be addressed as a matter of urgency, to ensure that organisations such as the police are able to make effective use of these new capabilities without fear of violating citizens' right to privacy.

Software Developers

- **Provision of new analytical software should be accompanied by an officer training session and/or instructional video delivered by the software developer.**

  Analytical tools are only as effective as the individual operating them, and investment will be wasted if officers are unable to effectively use the new technology. As modern software solutions are highly intuitive and can typically be adopted with ease, training in this context could be provided in the form of a brief presentation or an instructional video that officers can refer to at their convenience, demonstrating how a new piece of technology works and why it is effective.

- **All data applications should include an event log feature that is permanently enabled, documenting any changes that are made to a data set.**

  When presented with a new dataset, analysts should be able to view a corresponding event log, documenting how and when the data has been modified, and by whom. This will ensure continuity and prevent duplication from one user to the next.

- **Developers of predictive policing software should conduct further research into the use of network-based models for generating street segment-based crime predictions.**

  Recent research suggests that a calibrated network-based model which generates street-segment predictions delivers a significantly higher level of predictive accuracy than traditional grid-based predictions. Street-segment predictions are more useful for policing purposes than arbitrary grids, and these preliminary findings suggest that further research is needed to refine such models and integrate them within existing predictive policing software.

Further Research

- **Further research is needed to explore the potential uses of Risk Terrain Modelling (RTM) for identifying areas most at risk of experiencing crime.**

  Current predictive mapping methods rely on past criminal events alone to predict future crimes, and are indifferent to the underlying geographical and environmental factors that make certain locations more vulnerable to crime. RTM takes account of these underlying environmental factors to provide a comprehensive analysis of spatial risk, and some studies have shown that RTM has better predictive power than retrospective hotspot mapping.

- **Further research is needed to explore the use of harm matrices to assess the harms caused by different types of crime.**

  At present, in most forces, the prioritisation of police resources is based primarily on the total volume of crime in a given area, as opposed to the harms caused by different types of crime. Tools such as the MoRiLE Matrix demonstrate that it is possible to use data to understand harm in a much deeper way, by taking into account factors such as the harms caused to individuals, communities and the economy. Further research is needed to explore these potentials in greater detail, and develop more sophisticated analytical methods for evaluating the harm caused by different types of crime.

- **Further research is needed to explore the police's potential uses of big data collected from the Internet of Things.**

  It is likely that in the coming years there will be a significant increase in the amount of data being transmitted from sensors in the urban environment. While such data can be used to enhance the performance and efficiency of urban services, such as transportation networks, hospitals and schools, it could also transform the way urban environments are policed.

Introduction

IN RECENT YEARS, the rise of big data technology has revolutionised the domains of retail, healthcare, finance and many others. If used effectively, big data analytics has the potential to transform many aspects of policing. As sophisticated technologies become available at increasingly low cost, the effective use of big data will become a top priority for the police and other law enforcement agencies.

Until now, there has been only limited research exploring the uses of big data for policing, despite UK forces collecting vast amounts of digital data on a daily basis. The purpose of this paper is to identify specific ways in which big data analytics could enable UK police forces to make better use of the data they collect, allowing officers to act more efficiently and effectively.

Drawing directly from policing expertise and experience, this paper provides an overview of the current state of police technology in England and Wales, before exploring how big data could support policing at the operational, strategic and command levels. Potential challenges are identified, and practical solutions are proposed. Note that while the findings of this paper almost certainly extend to all the police forces in the UK, the research only examined the 43 territorial police forces in England and Wales.

Research Rationale

'We're sitting on absolutely monumental amounts of information collected from different sources. What we lack is the technological capability to effectively analyse it'.[1] These are the words of a detective inspector interviewed as part of this paper's research, and this sentiment recurred throughout all conversations that followed. While police officers differ in opinion regarding what technological development is needed, the importance of data analytics and what the priorities should be for enhancing data capabilities in the future, there is clear consensus over this fundamental fact: that the police lack the technical capability to effectively analyse the data they collect.

Few organisations in the UK collect data on the same scale as the police, and fewer still have such wide-reaching powers to acquire data from other sources. Yet the police make use of only a very small proportion of this data. At present, the analysis of police data is a laborious task, as forces do not have access to sophisticated data-mining tools and infrastructure. If the police were able to effectively apply such technology to the data they collected, they would greatly enhance their operational efficiency and crime-fighting capabilities.

It is imperative that any future technological development is informed by specific, clearly identified requirements and priorities. The purpose of this paper is to provide such an evidence

1. Author interview with a detective inspector, conducted by telephone, 5 April 2017.

base, revealing new insights into the police's current use of data, the key limitations, and the main requirements to expand these capabilities.

Methodology

Research for this paper was carried out in three phases. The first comprised a review of existing academic literature, government policy documents, law enforcement strategies and private sector reports on the police's use of data. This included recent and current initiatives to further develop data capabilities.

In the second phase, semi-structured interviews were conducted with 25 serving police officers and staff from four forces, as well as five experts from the technology sector and academia. Interviews were carried out in London in April and May 2017; most were conducted in person, with three taking place by telephone. All were conducted on a confidential basis, allowing respondents to speak openly about sensitive or contentious issues. Throughout this paper, where interviews are referenced as taking place in London, it does not necessarily mean that the officer belongs to a London police force.

The third phase of research involved a half-day workshop in London, which gathered together representatives from five police forces, as well as the Home Office, College of Policing and academia. This provided an opportunity to triangulate and validate the findings from the first two phases of research, and generated an informed discussion on future technological development, as well as ethical and legal considerations regarding the police's use of data.

This paper is divided into four chapters. Chapter I provides an overview of big data, outlining its primary uses, key limitations and practical applications to crime analysis. Chapter II explores the current state of policing technology in the UK, focusing on the issues raised in interviews. Chapter III discusses several specific ways in which big data technology could be applied to policing, based on the requirements identified during the research. Chapter IV highlights the main organisational barriers that need to be overcome to achieve these changes, as well as ethical and legal challenges concerning the police's use of data.

I. What is Big Data?

Definitions and Uses

DEFINITIONS OF BIG data vary considerably, and industry experts have yet to reach a consensus on the topic.[1] However, nearly all definitions make reference to an analytical process in which a large number of basic units (data points) are processed to produce a finished product. The purpose of this product is to answer questions, solve problems and tell stories. 'Big' data cannot be defined purely in terms of the size of a dataset, but rather the 'capacity to search, aggregate, and cross-reference large data sets'.[2] In other words, big data analytics (advanced analytics) becomes necessary when data is collected on such a large scale that it cannot be analysed with traditional data-management tools and methods.

A key feature of advanced analytics is the use of algorithms, which increasingly incorporate Artificial Intelligence (AI) methods underpinned by machine learning. As AI entails that the machine processing the data learns new rules through experience, the processing methods and calculations involved are often opaque to a human observer.[3] For the purposes of this paper, the term 'big data' does not necessarily entail the use of machine learning, unless specified.

Big data can incorporate a wide range of analytical techniques,[4] and the data used can take many different forms, originating from a virtually limitless number of potential sources. The majority of information that organisations hold is in an unstructured format, such as free text, images and video. A key distinction between advanced analytics and traditional data analytics is that the latter is unable to make sense of unstructured data. The former, however, is able to extract meaningful information from both structured and unstructured data. This allows the data to be interrogated in a much deeper way than with traditional data science methods, yielding significantly richer products.

Big data has countless practical applications, which have been discussed at length elsewhere. Major retail corporations use advanced analytics to inform pricing strategies and inventory

1. For a list of more than 40 definitions from industry experts, see Jennifer Dutcher, 'What is Big Data?', Berkeley School of Information, 3 September 2014, <https://datascience.berkeley.edu/what-is-big-data/>, accessed 6 July 2017.
2. Danah Boyd and Kate Crawford, 'Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon', *Information, Communication & Society* (Vol. 15, No. 5, 2012), p. 663.
3. Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection', March 2017, p. 9.
4. James Manyika et al., 'Big Data: The Next Frontier for Innovation, Competition, and Productivity', McKinsey Global Institute, May 2011, pp. 27–31.

control, as well as to better direct advertising.[5] GPS navigational systems use big data to provide drivers with real-time traffic information, and even to predict flows of traffic using deep learning methods.[6] In healthcare, routine use of electronic health records has generated vast datasets of personal information, and big data analytics has the potential to improve care, save lives and lower costs.[7] In astronomy, analysis of data collected by telescopes has helped to create the most detailed three-dimensional maps of the universe ever made.[8] Big data has also transformed the domains of weather forecasting, education, cyber security and fraud detection,[9] as well as telecommunications, politics and public sector administration.[10]

Big data is big business. Worldwide revenues for big data analytics are forecast to exceed $150 billion in 2017, rising to more than $210 billion by 2020.[11]

Limitations

Despite its widely transformative capabilities, big data is not without limitations. Advanced analytics is by no means objective, but rather relies on a considerable amount of interpretation. For instance, the data-cleaning process – which involves the manual elimination of variables and data points that are deemed either irrelevant or unreliable – is inevitably subjective. As computer scientist and statistician Jesper Andersen puts it, 'the moment you touch the data, you've spoiled it'.[12]

Additionally, datasets used for advanced analytics are often unreliable, containing data from disparate sources, which has been collected using varied and sometimes questionable methods. Increasingly, the data used originates from the internet, and 'large data sets from Internet sources are often unreliable, prone to outages and losses, and these errors and gaps are magnified when multiple data sets are used together'.[13]

As the amount of available data and the number of potential sources increase, so too does the potential for error – and particularly the possibility of detecting relationships where none exist (false positives). Statistician and mathematician David Leinweber is famed for providing a contrived yet excellent illustration of this. He developed a regression model that predicts the

5. David Bollier, *The Promise and Peril of Big Data* (Washington, DC: Aspen Institute, Communications and Society Program, 2010).
6. Yisheng Lv et al., 'Traffic Flow Prediction with Big Data: A Deep Learning Approach', *IEEE Transactions on Intelligent Transportation Systems* (Vol. 16, No. 2, 2015), pp. 865–73.
7. Wullianallur Raghupathi and Viju Raghupathi, 'Big Data Analytics in Healthcare: Promise and Potential', *Health Information Science and Systems* (Vol. 2, No. 3, 2014), p. 3.
8. SDSS, 'The Sloan Digital Sky Survey: Mapping the Universe', <http://www.sdss.org/>, accessed 6 July 2017.
9. Steve Mills et al., 'Demystifying Big Data: A Practical Guide to Transforming the Business of Government', TechAmerica Foundation, 2012.
10. Manyika et al., 'Big Data: The Next Frontier for Innovation, Competition, and Productivity', pp. 27–31.
11. International Data Corporation, 'Worldwide Semiannual Big Data and Analytics Spending Guide', 2016.
12. Bollier, *The Promise and Peril of Big Data*, p. 13.
13. Boyd and Crawford, 'Critical Questions for Big Data', p. 668.

annual fluctuations of the S&P stock index with 99% accuracy using three input variables: cheese production in the US; butter production in the US and Bangladesh; and the sheep population in the US and Bangladesh.[14]

These examples serve as an important reminder that a failure to understand the inherent limitations present in a dataset can easily lead to misinterpretation, and that 'data analysis is most effective when researchers take account of the complex methodological processes that underlie the analysis of that data'.[15] This is especially true in the context of big data, where even a small oversight can have a dramatic impact on the findings.

Another limitation is the high technical requirement of big data analytics, both in terms of computing power and storage. Data complexity has evolved dramatically in both volume and structure, and typical organisations do not have the infrastructure required to store and process data at the rate at which it is now being collected. Traditional data warehouses are not sufficient to meet the requirements of advanced analytics and so organisations must instead invest in creating data lakes – centralised data stores that can accommodate all forms of data, both structured and unstructured.

Due to the prohibitive costs involved, it is unfeasible for public sector organisations to invest in dedicated in-house infrastructure to perform big data analytics, leading to an increase in partnerships with third-party organisations that provide the service within a secure private cloud environment or private data centre, a model known as 'Infrastructure as a Service'.[16]

Many public sector organisations are understandably apprehensive of storing sensitive data in the cloud or in off-premise facilities, especially given the restrictive nature of data-protection legislation. This is especially true for the police, who handle large volumes of personal data. However, the security of private cloud networks has improved considerably over time, and the use of cloud-based data services is increasingly becoming the norm in both the private and public sectors.

Another key limitation lies in the legal and ethical constraints on the use of big data, and these issues are briefly discussed in Chapter IV. It is, however, beyond the scope of this paper to address in detail the many ethical and legal concerns arising from the use of big data. Moreover, this paper is concerned primarily with the *analysis,* rather than the *collection* of digital data. The legislation governing digital surveillance, interception of communications and other means of data collection – such as the Regulation of Investigatory Powers Act 2000 and the Investigatory Powers Act 2016 – are not a focus of this paper.[17]

14. David J Leinweber, 'Stupid Data Miner Tricks: Overfitting The S&P 500', *Journal of Investing* (Vol. 16, No. 1, 2007), pp. 15–22.
15. Boyd and Crawford, 'Critical Questions for Big Data', p. 668.
16. See Michelle Boisvert, 'Infrastructure as a Service (IaaS)', TechTarget, January 2015, <http://searchcloudcomputing.techtarget.com/definition/Infrastructure-as-a-Service-IaaS>, accessed 19 July 2017.
17. For a more detailed discussion of the ethical and legal constraints on the collection of digital data, see Michael Clarke et al., 'A Democratic Licence to Operate: Report of the Independent

Big Data and Crime Analysis

While big data has been employed in a diverse range of domains, its use in policing has been limited. The three main ways in which the police use big data are through the collection and storage of DNA information, mass surveillance, and predictive policing.[18] Police forces have the capability to cross-reference DNA samples against a database of millions of other records with a high degree of accuracy. However, the police's ability to trawl and analyse large volumes of data collected through mass surveillance is far more limited. This is because DNA information is stored as numbers, which is far more straightforward to analyse than large volumes of unstructured data, such as images and video.

Predictive policing is defined as 'taking data from disparate sources, analyzing them and then using results to anticipate, prevent and respond more effectively to future crime'.[19] Predictive policing is based on the notion that analytic techniques used by retailers to predict consumer behaviour can be adapted and applied to policing to predict criminal behaviour.[20]

The use of quantitative analysis to make predictions about crime levels is by no means a new approach.[21] In the 1990s, the New York Police Department (NYPD) was at the forefront of the intelligence-led policing revolution. The CompStat (Compare Statistics) system, originally conceived by then Deputy Commissioner Jack Maple in 1994, was used to track spatial crime patterns and identify hotspots by sticking pins in maps. Twice-weekly CompStat meetings for commanding officers became compulsory.[22] The system allowed the performance of each officer's precinct to be quantitatively measured, providing a level of accountability never before seen in law enforcement.

The dramatic reductions in crime rates that followed within precincts that implemented CompStat led almost every law enforcement agency in the US to adopt the practice of automated mapping and statistical analysis of crime data,[23] and CompStat is now considered in the US to be 'part of the institutional DNA of policing'.[24] The practice of hotspot policing – where crime hotspots are identified using spatial analysis and police activity is targeted to these areas accordingly

---

Surveillance Review', *Whitehall Report*, 2-15 (July 2015).

18. Elizabeth E Joh, 'Policing by Numbers: Big Data and the Fourth Amendment', *Washington Law Review* (Vol. 89, No. 1, March 2014).
19. Beth Pearsall, 'Predictive Policing: The Future of Law Enforcement?', National Institute of Justice, No. 266, June 2010, p. 16.
20. Jennifer Bachner, 'Predictive Policing: Preventing Crime with Data and Analytics', IBM Centre for the Business of Government, Improving Performance Series, 2013, p. 4.
21. Walter L Perry et al., *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations* (Santa Monica, CA: Rand Corporation, 2013).
22. David Weisburd et al., 'Reforming to Preserve: Compstat and Strategic Problem Solving in American Policing', *Criminology & Public Policy* (Vol. 2, No. 3, 2003), pp. 421–56.
23. Bachner, 'Predictive Policing: Preventing Crime with Data and Analytics', p. 9.
24. Bureau of Justice Assistance, US Department of Justice, and Police Executive Research Forum, *Compstat: Its Origins, Evolution and Future in Law Enforcement Agencies* (Washington, DC: Police Executive Research Forum, 2013), p. vii.

– has been shown to be a highly effective crime prevention strategy in a range of different environments and for many different crime types.[25]

The technological capabilities of the police have advanced immeasurably since the days of CompStat, and law enforcement agencies worldwide now have access to both a much greater volume of useable data and far more sophisticated and efficient methods of analysis. Research for this paper found that UK police forces have access to a vast amount of digital data, but currently lack the technological capability to use it effectively. With mass surveillance and the large-scale collection of data now becoming a matter of daily routine, the challenge ahead lies in devising accessible, affordable systems that can be used to process and analyse these quantities of data efficiently and reliably.

25. Anthony A Braga, 'The Effects of Hot Spots Policing on Crime', *The Annals of the American Academy of Political and Social Science* (Vol. 578, No. 1, November 2001), pp. 104–25.

II. Police Technology in the UK Today

THE RESEARCH FOR this paper has found that UK police forces are several years away from being able to effectively implement big data technology. There are, at present, significant deficiencies in the police's core data infrastructure, making it difficult to carry out even basic data entry tasks.

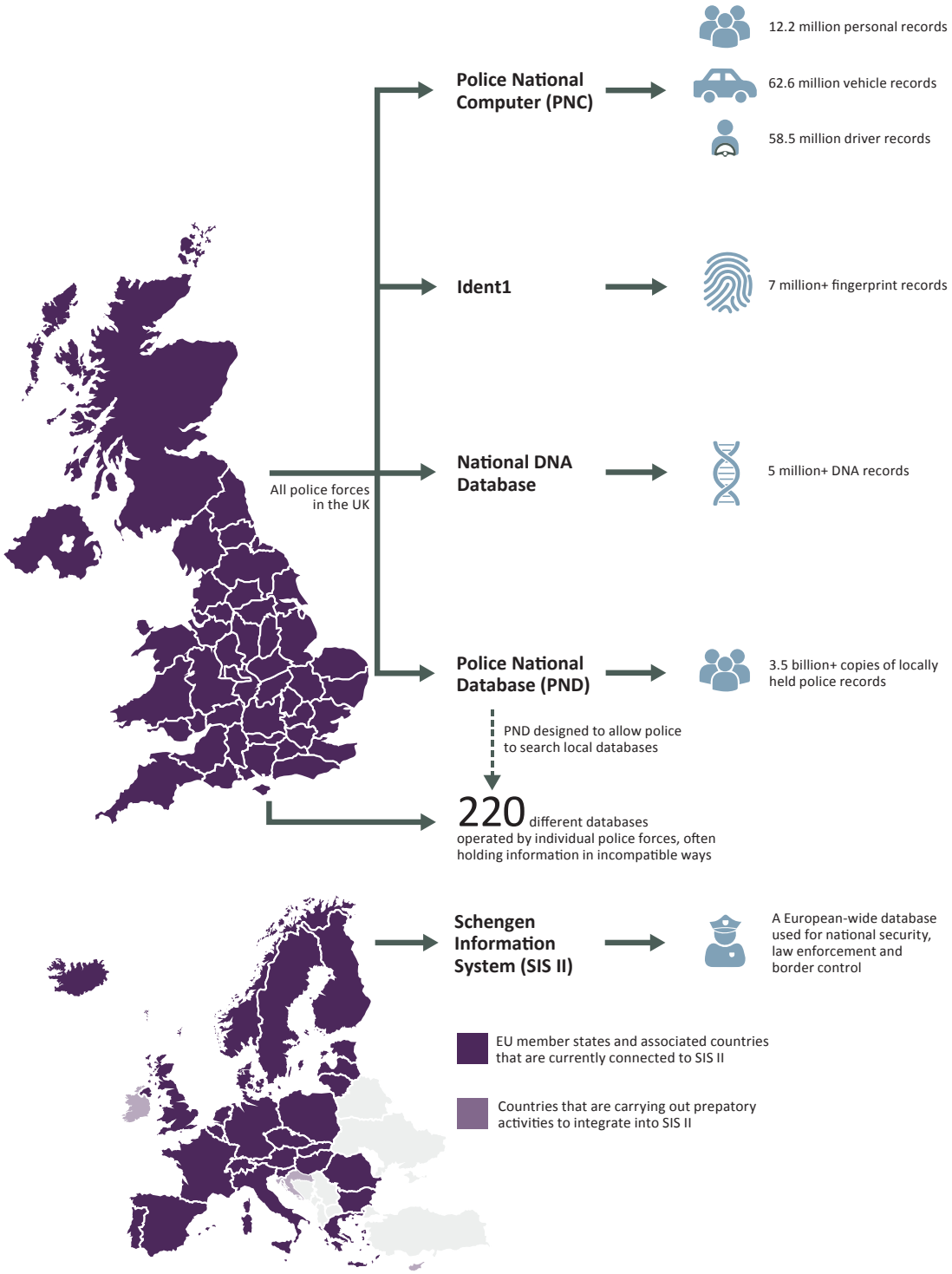
Many officers interviewed by the author were aware of technology being used in other sectors that would allow them to perform their jobs more efficiently, but felt that this was 'out of reach' of the police. It is hoped that within the next few years, the fundamental deficiencies in core data infrastructure will be rectified, enabling forces to make effective use of big data technology.

All police forces in the UK have access to the Police National Computer (PNC), a nationwide database, which as of May 2017 contained more than 12.2 million personal records, 62.6 million vehicle records and 58.5 million driver records (see Figure 1).[1] This is distinct from the Police National Database (PND), a national intelligence-handling system containing copies of locally held police records. PND was created in 2011 to allow police officers to search across the 220 different databases operated by individual police forces in the UK.[2] Some of these local databases hold huge amounts of data: the Metropolitan Police Service (MPS) in London operates an Automatic Number Plate Recognition (ANPR) network which receives approximately 38 million reads a day.[3]

As well as PNC and PND, all UK police forces also have access to Ident1, the central national database for biometric information, which contains more than 7 million fingerprint records,[4] and the National DNA Database, which holds the DNA information of more than 5 million individuals.[5] In 2015, the UK connected into the second-generation Schengen Information System (SIS II), a European-wide governmental database used for national security, law enforcement and border control.[6]

1. Figures provided by the Home Office to the author under Freedom of Information Request 43411, email, 'Total Number of PNC Records', 4 May 2017.
2. Datalynx, 'Police National Database', <http://www.datalynx.net/case-studies/police-national-database/>, accessed 30 August 2017; CGI, 'Case Studies: Police National Database', <https://www.cgi-group.co.uk/case-study/police-national-database-joins-forces>, accessed 14 August 2017; Paul Crowther, 'Oral evidence: Police National Database', oral evidence given before the Home Affairs Committee, HC 960, 20 January 2015, Q63.
3. Metropolitan Police, 'One Met: Total Technology 2014–17', 2014.
4. *Ibid.*
5. Home Office, 'National DNA Database Statistics', <https://www.gov.uk/government/statistics/national-dna-database-statistics>, accessed 6 July 2017.
6. Home Office, 'Second Generation Schengen Information System (SISII): General Information', 13 April 2015, <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/421540/SISII_General_Information_document.pdf>, accessed 6 July 2017.

Figure 1: Major Databases to which UK Police Forces Have Access



Source: HMIC, 'Map of Police Forces in England, Wales, and Northern Ireland', 21 February 2013; Home Office, 'Second Generation Schengen Information System (SISII): General Information', 13 April 2015.

These are just a few examples of the many databases to which the police forces have access. Once the necessary infrastructure is in place to effectively manage and analyse this huge volume of data, its potential applications to policing are virtually limitless.

Database Management

As in many public service domains, law enforcement data is typically stored across multiple fragmented databases, with no unified system for querying several databases simultaneously. Officers are required to input the same data multiple times, accessing several individual systems to manually collate information and manage cases. Data applications operate in isolation, with no facility to copy information between different systems. This problem has created demand for systems that allow users to access and input information across multiple databases simultaneously.

In the UK, police force productivity is still greatly impeded by the fragmentation of databases. Speaking to the London Assembly Budget and Performance Committee in June 2013, Metropolitan Police Assistant Commissioner for Specialist Operations Mark Rowley explained that 'if an officer is dealing with a crime from start to finish in terms of arrest and putting a file together, they will input the names of both the suspect and the victim 10 or 12 times'.[7]

Addressing this problem is of great importance to police forces across the country as well as central government. The UK Digital Strategy, launched in March 2017, lists consolidating databases as a top priority for the coming years, promising to 'create a linked ecosystem of trusted, resilient and accessible canonical datastores (known as registers) of core reference data. These registers will make government data easier to create, maintain, and put to use'.[8]

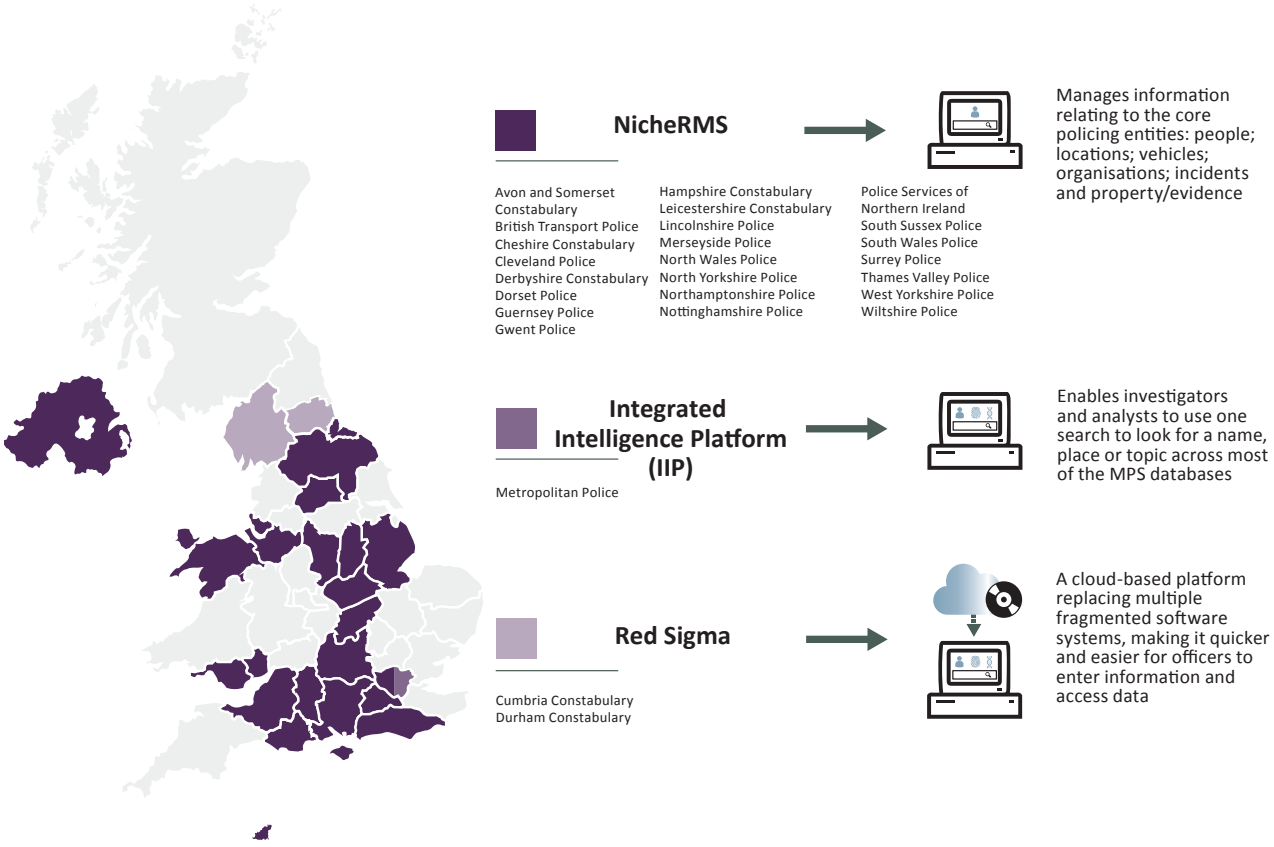
To date, limited progress has been made in the unification of police databases. MPS officers have access to the Integrated Intelligence Platform (IIP), a single portal through which officers can search across all major MPS systems. However, MPS officers interviewed as part of this research reported that searches made through IIP return only a small proportion of all relevant information, so they instead resort to retrieving information from each individual database separately.[9] Twenty-three forces in the UK use NicheRMS,[10] a software system that unifies data entry applications in a similar way, but the majority of territorial forces in England and Wales have yet to implement such technology.[11]

Durham and Cumbria Constabularies recently adopted a cloud-based platform, Red Sigma, to replace the multiple fragmented software systems previously used to perform routine policing tasks.[12] Stuart Grainger, then head of ICT at Durham Constabulary, explained how the new

7. London Assembly, Budget and Performance Committee, *Smart Policing: How the Metropolitan Police Service Can Make Better Use of Technology* (London: City Hall, 2013).
8. Department for Digital, Culture, Media & Sport, 'UK Digital Strategy: 7. Data – Unlocking the Power of Data in the UK Economy and Improving Public Confidence in its Use', 1 March 2017.
9. Author interviews with MPS officers, London, 19 April 2017.
10. This includes British Transport Police and Police Service of Northern Ireland, which were not studied as part of this research.
11. NicheRMS, 'Who We Serve', <http://nicherms.com/who-we-serve/>, accessed 27 July 2017.
12. Durham Constabulary, 'New IT System Rolled Out in Fight Against Crime', 14 August 2017.

system has reduced data overlap and duplication, making significant time savings.[13] Crucially, Durham Constabulary's strategy did not require any formal training for officers, but focused instead on making the technology intuitive and easy to use.

Figure 2: Major Database Management Systems Currently Being Used by UK Police Forces



Source: NicheRMS, 'Who We Serve', <http://nicherms.com/who-we-serve/>, accessed 27 July 2017; Robert Milne, Forensic Intelligence *(Boca Raton, FL: CRC Press, 2013), p. 3; Dan Worth, 'Durham Police Turn to the Cloud to Improve Crime Fighting', V3, 11 March 2015.*

Another issue highlighted by officers is the lack of a standardised lexicon or naming conventions for police databases. For example, an officer recording a firearms offence may use the words 'shoot', 'fire' or 'discharge' when entering case information into a database. This complicates the search process for investigators and analysts who are later trying to retrieve such data. Instead of filtering all firearms offences within a certain area during a specific timeframe, investigators must instead use a keyword search to return all relevant information, and even then there is no way to be sure that they have not omitted any lesser-used synonyms. With police hours becoming an increasingly scarce resource, it is more important than ever that valuable time is

13. Dan Worth, 'Durham Police Turn to the Cloud to Improve Crime Fighting', *V3*, 11 March 2015.

not wasted carrying out routine administrative tasks. Some forces have started to address the problem locally (see Figure 2), but there has been little progress at the national level. Only when a unified national infrastructure is in place for centrally managing all police data will forces be able to make effective use of big data technology.

Collaboration, Compatibility and Continuity

The research for this paper found that there is wide variation in the levels of technological development across forces. Different forces invest in different software for the same policing purposes, resulting in IT systems that are not mutually compatible. It was also highlighted that forces invest in developing specific tools for short-term requirements, but are then often unable to re-use such technology in the future.[14]

In general, the officers interviewed were unaware of technological initiatives that had been pursued by other forces nationwide. For instance, as already mentioned, some forces have adopted sophisticated software packages to unify data-management applications, while most others have yet to identify a solution to this problem. Cambridgeshire and Durham constabularies have implemented software that allows officers to remotely access custody images using an app on their smartphone.[15] However, officers from other forces were unaware that such a capability existed in the UK, and were confused as to why it had not been made available to them (this is discussed further in Chapter IV).

There is similar variation within forces. For example, facial matching seems to be used within certain departments, but not others. While detectives from one unit explained how they were able to cross-reference CCTV images with custody photographs for facial matching,[16] other detectives from a separate unit claimed that they had never had access to such technology, and were not aware of it being used by police forces in the UK.[17]

The Police ICT Company,[18] established in 2015, was conceived to address this lack of cohesion and to 'enable collaboration at a local, regional and national level'.[19] The company estimated at the time that there were more than 2,000 individual police software systems in use across the UK, and aimed to consolidate these by promoting a national approach to procurement.[20] However, inaction on the part of government to mandate particular systems has meant that there is little to prevent forces from procuring individual pieces of software in isolation from other forces, and as a result little progress has been made in improving national compatibility.

14. Author interviews with a superintendent and detective chief inspector, MPS, London, 18–19 April 2017.
15. Author interviews with detectives from Durham and Cambridgeshire constabularies, London, 3 May 2017.
16. Author interviews with a detective inspector and a detective sergeant, London, 18 April 2017.
17. Author interviews with a detective sergeant and a detective constable, London, 19 April 2017.
18. Police ICT, 'About the Police ICT', <https://ict.police.uk/about/>, accessed 6 July 2017.
19. Home Office, 'Have You Got What It Takes? The Police ICT Company', 2015.
20. *Ibid*.

There is also poor continuity between individuals working on police data, with analysts reporting that they spend a great deal of time checking the work carried out previously by colleagues.[21] When presented with a new dataset, analysts have no way of knowing how the data has been previously modified, or by whom, leading to duplication of work and potential errors.

Departments also suffer from a lack of coordination when using data for ongoing investigations. Forces typically operate separate teams for analysing different types of data, with little collaboration between them. For instance, the ANPR unit may be working with data pertinent to an investigation being carried out by the communications data team, but each may be unaware that the other has data of interest.

To address this issue, several interviewees suggested that serious investigations should be overseen by an individual who assesses data requirements and serves as the main point of contact for all digital aspects of the investigation. Digital media investigators (DMIs) were conceived to fulfil precisely this role, and the College of Policing offers a DMI course for officers of forces that participate in the programme. However, in 2015, Her Majesty's Inspectorate of Constabulary (HMIC[22]) found that in all but one force, the introduction of DMIs was still in the planning stage.[23] Officers interviewed as part of this research suggested that investigative units would greatly benefit from the presence of a DMI, and this report suggests that all forces in the UK should consider taking part in the initiative.

Predictive Analytics

Despite their proven effectiveness, few UK police forces have adopted crime prediction tools as part of their digital strategies. A great deal of officer time is still taken up with traditional beat policing, despite the fact that hotspot policing has been repeatedly shown to be significantly more effective at predicting where crime will happen, thereby allowing limited resources to be deployed to where they are most needed.[24]

Police mapping techniques are currently retrospective, identifying hotspots on the basis of recorded crimes and phone calls made by the public relating to crime and disorder.[25] Hotspot maps are produced manually by analysts, and senior officers report that 'by the time we get it [the hotspot map], it's out of date'.[26] This need not be the case, given that automated predictive crime mapping tools are readily available at relatively low cost.

21. Author interviews with four police analysts, London, 18 April 2017.
22. Note that, as of July 2017, HMIC has changed its name to Her Majesty's Inspectorate of Constabulary and Fire and Rescue Services (HMICFRS).
23. HMIC, *Real Lives, Real Crimes: A Study of Digital Crime and Policing* (London: Her Majesty's Inspectorate of Constabulary, 2015), p. 51.
24. Kate J Bowers et al., 'Spatial Displacement and Diffusion of Benefits among Geographically Focused Policing Initiatives: A Meta-Analytical Review', *Journal of Experimental Criminology* (Vol. 7, No. 4, December 2011), pp. 347–74; Braga et al., 'The Effects of Hot Spots Policing on Crime', pp. 633–63.
25. College of Policing, 'The Effects of Hot-Spot Policing on Crime: What Works Briefing', September 2013.
26. Author interview with a superintendent, London, 18 April 2017.

HMIC's most recent annual police effectiveness (PEEL) inspection highlighted the fact that forces have not yet made effective use of predictive crime mapping tools:

> As analytical resources across the service are apparently shrinking, HMIC encourages forces to make better use of opportunities available through new and emerging technology. Innovative analytical techniques should be used to help the service to make decisions about where to target resources. HMIC found that most forces have not yet explored fully the use of new and emerging techniques and analysis to direct operational activity at a local level. Forces need to develop a greater awareness of the benefits available from sophisticated crime-mapping software and enhanced analysis in predicting and preventing crime within local communities. Intelligent use of such technology could improve effectiveness, release officer capacity, and is likely to be cost effective.[27]

The PEEL inspection found that the most high-performing forces were those that use 'structured and evidence-based approaches to tackling local problems and [provide] preventative and targeted foot patrols'.[28] One such force is Kent Police, which introduced predictive policing software in 2013. An operational review carried out by Kent Police in 2014 found that the software produced a hit rate of 11%, making it '10 times more likely to predict the location of crime than random patrolling and more than twice as likely to predict crime [than] boxes produced using intelligence-led techniques'.[29]

Despite this impressive level of accuracy, the review found that only 25% of boxes generated by the software were actually visited by police officers. Patrol officers reportedly did not understand why they were being sent to particular locations, as they had not been adequately briefed on how the software worked or why it was effective. As a result, the majority of officers simply chose to ignore the predictions.

Speaking on the subject to the London Assembly Budget and Performance Committee in 2013, Her Majesty's Inspector of Constabulary (HMI) for the MPS, Stephen Otter, said, 'it works; it is evidenced; it is professional practice'.[30] More recently, the issue was raised in Europol's 2017 Serious and Organised Crime Threat Assessment, which stated that:

> Technology is also a significant aid to law enforcement authorities in the fight against serious and organised crime. This includes the use of advanced digital forensics tools, the deployment of predictive policing software driven by Big Data as well as drones for the monitoring of areas and large events.[31]

27. HMIC, *PEEL: Police Effectiveness 2016: A National Overview* (London: Her Majesty's Inspectorate of Constabulary, 2017), p. 33.
28. *Ibid*., p. 10.
29. Kent Police, 'PredPol Operational Review [Restricted and Heavily Redacted]', Corporate Services Analysis Department, <http://www.statewatch.org/docbin/uk-2014-kent-police-predpol-op-review.pdf>, accessed 6 July 2017.
30. London Assembly, Budget and Performance Committee, *Smart Policing*.
31. Europol, *Serious and Organised Crime Threat Assessment 2017: Crime in the Age of Technology* (The Hague: Europol, 2017), p. 25.

Recommendation 1 of HMIC's most recent PEEL report is that by December 2017, national guidance should be issued to police forces on seven specific topics, including 'analytical capability to support effective and targeted preventative policing'.[32] This paper reiterates this pressing need, concluding that UK police forces have not taken advantage of tried-and-tested technologies that could significantly improve the efficiency of local operational activity – in particular predictive hotspot mapping. Cuts to personnel do not diminish the need for this analysis; on the contrary – predictive hotspot policing allows forces to make most efficient use of the limited resources available to them. The use of predictive policing software is discussed in more detail in Chapter III.

Mobile Policing

While smartphones and mobile tablet devices have transformed many aspects of people's day-to-day lives, police forces are yet to benefit from this mobile revolution. In its 2013 report, the London Assembly Budget and Performance Committee concluded the following about the Met's use of mobile technology:

> Tablet and smartphone technology is commonly available and relatively cheap. Many Londoners now have smartphones in their pockets, giving instant access to travel information, bar and restaurant reviews, news and much more. Yet a police officer has to radio back to base to find out simple background information about, for example, previous crime reports or information about particular suspects. It seems incredible that officers have this modern technology at home yet when they arrive at work they take a step back in time.[33]

While the committee was investigating the MPS specifically, its findings extend to the majority of forces across the country. At present, most officers do not have the technology required to access or input data remotely, greatly restricting their operational capabilities. One officer recalls an occasion when she had to wait by the side of the road with a suspect detained in handcuffs for around 30 minutes for a call back from base to verbally deliver their PNC record over the radio.[34] A police car's onboard computer would deliver this information immediately.

In response to this problem, the Mayor of London's draft Police and Crime Plan for London 2017–2021 lists mobile policing as a top priority, promising to 'equip frontline officers with mobile data tablets to enable them to work on the move, without having to return to the station to access or input information'.[35] Many forces across the country have already implemented this initiative,[36] and most others intend to do the same in the coming years.

32. HMIC, *PEEL: Police Effectiveness 2016*, p. 23.
33. London Assembly, Budget and Performance Committee, *Smart Policing*.
34. Author interview with a police constable, London, 19 April 2017.
35. Greater London Authority, 'A Safer City for all Londoners: Police and Crime Plan 2017–2021', March 2017, p. 47.
36. Airwave, 'Pronto Policing', <https://www.airwavesolutions.co.uk/airwave-smartworld/smart-applications/pronto/pronto-policing/>, accessed 30 August 2017.

The initiative's timing coincides with a major upgrade to the emergency services mobile communications system, due to begin later this year. At present, the emergency services rely on Airwave – an archaic communications system with a data transfer speed of 0.00095 mbps, approximately 30,000 times slower than the average home internet connection. 'Over Airwave you can talk to each other, have talk groups, send text messages, but you can't do much more than that', according to Deputy Chief Constable Richard Morris.[37]

The new Emergency Services Network (ESN) will provide superfast 4G data transfer speeds, allowing emergency services to remotely access a range of data services, such as PNC. There are potential issues surrounding network coverage, as nationwide 4G coverage is significantly poorer than that of Airwave. However, there will be a gradual transition from Airwave to ESN over the coming years, by which time nationwide 4G coverage will probably have improved considerably. Effective use of mobile technology has the potential to revolutionise day-to-day policing, but it is imperative that mobile devices are supported by an efficient core infrastructure and reliable communications network.

The provision of mobile technology will allow officers to perform simple tasks remotely, without having to manually radio back to base to retrieve information. However, at present, when retrieving information using a police car's computer, data is often returned in the form of many pages of free text, and it is difficult for officers to extract meaningful information from these during a live operation.[38] Therefore, when mobile policing becomes a reality, it is essential that officers are able to view a clear and concise overview of an individual's history and past offences, as opposed to being presented with many pages of free text.

Communications Data

Where there is an immediate threat to life, communications data can be provided instantly by communications service providers (CSPs) on the verbal authority of a senior officer. However, for reactive investigations (where there is no immediate threat to life) there are often delays of up to 28 days in gaining access to communications data. When data is eventually provided, the datasets are often so large that they are difficult to interpret, and are provided in the form of free text that must be converted into spreadsheets prior to analysis.

Analysts estimate that they spend at least 50% of their working time cleansing data, and that 'a lot of time is taken having to manually read through free-text files'. While there are programs available for automatically converting free-text data into a useable format, the software is at present rudimentary and unreliable.[39]

A detective chief inspector described how it may take days for investigators and analysts to identify connections within large sets of communications data, but that the technology is available to do this

37. Oliver Smith, 'Faster, Data-Driven: Police & Ambulance Crew Get Life-Saving Upgrade', *The Memo*, 2 March 2016.
38. Author interview with a police constable, London, 19 April 2017.
39. Author interviews with four police analysts, London, 18 April 2017.

within minutes.[40] A limited number of tools are available for the police to perform basic descriptive analytics on communications data,[41] but a detective sergeant explained that most officers do not know that these tools exist, and fewer still have the technical ability to make effective use of them.[42] Instead, detectives manually trawl through vast amounts of data, often unaware that they have access to automated tools that could save them a significant amount of time. When asked why this may be the case, the detective sergeant replied simply, 'because training is non-existent'.

After communications data has been used and is no longer needed, it is stored in a central database and rarely revisited. According to a superintendent, 'we have shed loads of data that is used once and then never looked at again'.[43] Such information could provide valuable insights into the behavioural patterns and associates of known individuals, but analysing the data is such a painstaking process that forces are prevented from making effective use of it.

Open-Source Data

The police's use of open-source data is currently heavily restricted by technological deficiencies as well as organisational barriers. Interviews revealed that police forces spend a significant amount on buying social media data, but there are often no meaningful outputs from analysis. There are no automated systems in place to analyse this data, and manual analysis of social media is a laborious process.

The software required to perform open-source analysis covertly (without leaving a digital trace) is provided on a licence basis, and is installed only on a small number of computers. Investigators are required either to switch to a separate machine to perform open-source analysis, or to request that a specialist team carry this out on their behalf.

For proactive investigations, the time constraints are often too great for investigators to go through this process, and detectives report evading the process by using their own personal devices to perform open-source analysis.[44] Police forces' use of open-source data would be significantly enhanced if investigators were able to perform such analysis covertly on their own computer terminal.

Where forces have made effective use of social media data, it is typically for large events with many individuals involved, such as protests and riots. The use of open-source analysis for smaller-scale investigations is much more limited. Officers explained that the tools used for social media analysis typically operate on a large amount of data in a wide geographical area, but that individual forces are responsible for responding to crime problems only within their local jurisdiction. Therefore, the wide geographical nature of open-source analytics is not compatible with the highly localised structure of UK policing.

40. Author interview with a detective chief inspector, London, 18 April 2017.
41. These are typically macro scripts within Microsoft Excel, which in many cases have been written by current or previous analysts.
42. Author interview with a detective sergeant, London, 18 April 2017.
43. Author interview with a superintendent, London, 18 April 2017.
44. Author interviews with detectives, London, April 2017.

III. Big Data and the Future of Police Technology

THE PREVIOUS CHAPTER illustrates that UK police forces are several years away from being able to effectively implement big data technology due to deficiencies in existing data infrastructure. Addressing these issues has become a priority for various government organisations in recent years. It is hoped that in the near future, forces in the UK will have access to the infrastructure necessary to implement advanced analytical tools and methods. This chapter highlights some of the ways in which big data could support policing in the future, based on the requirements identified in the course of interviews with police officers and staff.

Predictive Hotspot Mapping

The emergence of advanced analytics has enabled the development of sophisticated predictive crime mapping tools that use statistical models to identify areas that are at increased risk of experiencing crime, based on past criminal events. The practice of using past crimes to predict future crime is founded on the observation that repeat victimisation accounts for a large proportion of all crime.[1] In addition, crime is often contagious, with the risk of victimisation for crimes such as burglary increasing for houses near a burgled property in the aftermath of the initial crime.[2]

Based on research first published in 2004,[3] various field trials have since demonstrated that predictive mapping software is in most cases significantly more effective at predicting the location of future crimes than traditional intelligence-led techniques.[4]

The software has since been commercialised in the form of PredPol, a crime prediction tool developed in 2011 by mathematicians and social scientists at UCLA and Santa Clara University in

1. For reviews on repeat victimisation, see Graham Farrell and Ken Pease (eds), *Repeat Victimization: Crime Prevention Studies, Volume 12* (Monsey, NY: Criminal Justice Press, 2001); Ken Pease, *Repeat Victimisation: Taking Stock*, Crime Detection and Prevention Series, Paper 90 (London: The Stationery Office, 1998).
2. Kate J Bowers, Shane D Johnson and Ken Pease, 'Prospective Hot-Spotting the Future of Crime Mapping?', *British Journal of Criminology* (Vol. 44, No. 5, September 2004), pp. 641–58; Shane D Johnson, 'Repeat Burglary Victimisation: A Tale of Two Theories', *Journal of Experimental Criminology* (Vol. 4, No. 3, 2008), pp. 215–40.
3. Bowers, Johnson and Pease, 'Prospective Hot-Spotting the Future of Crime Mapping?'.
4. Shane D Johnson et al., *Prospective Crime Mapping in Operational Context,* Final report (London: The Stationery Office, 2007); Pearsall, 'Predictive Policing: The Future of Law Enforcement?'; Bachner, 'Predictive Policing: Preventing Crime with Data and Analytics'; Perry et al., *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*; Kent Police, 'PredPol Operational Review'.

collaboration with the Los Angeles Police Department (LAPD) and Santa Cruz Police Department, based on work conducted earlier that year.[5]

Since being implemented across various jurisdictions in California, PredPol has proved effective in reducing the incidence of property crimes, most notably burglary. Santa Cruz Police Department (which first implemented PredPol in July 2011) reported a 14% reduction in the number of burglaries from January to June 2012 when compared with the same time period from the previous year.[6] The LAPD's Foothill Division reported a 20% reduction in predicted crimes from January 2013 to January 2014. Similar results have been seen elsewhere, for instance in Alhambra and Modesto, in California, and Norcross, Georgia.[7]

The system initially processes several years of crime data to lay down a 'background' level of crime, using an Epidemic Type Aftershock Sequence model – a self-learning algorithm based on seismological models that predict earthquake aftershocks.[8] The software uses three data points as the input for forecasting: crime type; crime location; and crime date and time. These are deemed to be the most objective variables, since they do not include any personal data.

Using self-exciting point process modelling (a time series model in which the occurrence of past points makes the occurrence of future points more probable), the data is used to pinpoint small (500 x 500 feet) boxes that indicate the times and places where crime is most likely to occur. These boxes are automatically generated for each shift of each day, allowing officers to respond to potential crime locations in real time.

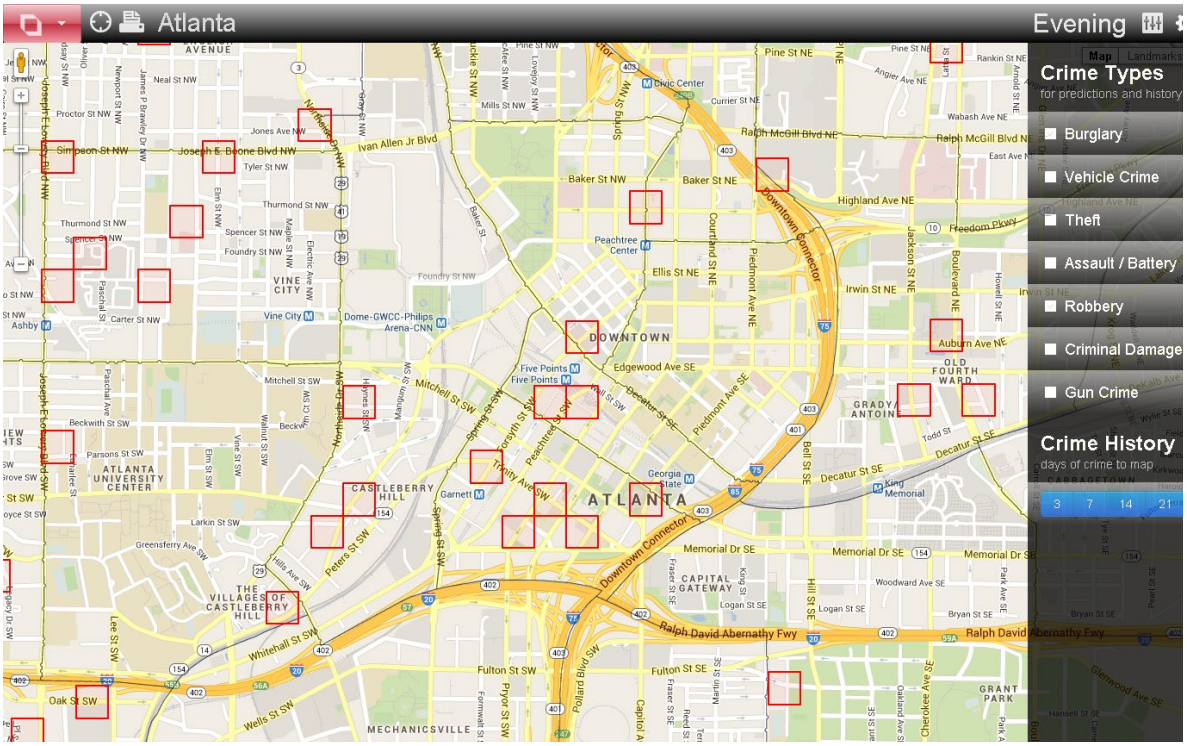
One of the main attractions of the software is that it is affordable and requires minimal training.[9] This is crucial, as a primary purpose of crime analysis tools is to enable police departments to do more with less, in other words to achieve tangible results even in the face of significant cuts to personnel. Predictive mapping achieves this by allocating resources more efficiently. In 2012, Greater Manchester Police developed a proprietary version of predictive mapping software that was shown to be effective at reducing burglary, demonstrating that in many cases such software can be implemented at no extra cost.[10]

5. G O Mohler et al., 'Self-Exciting Point Process Modelling of Crime', *Journal of the American Statistical Association* (Vol. 106, No. 493, March 2011), pp. 100–08.
6. Bachner, 'Predictive Policing: Preventing Crime with Data and Analytics', p. 26.
7. Predpol, 'Proven Crime Reduction Results', <http://www.predpol.com/results/>, accessed 6 July 2017.
8. For a mathematical explanation of how this type of model works, see Mohler et al., 'Self-Exciting Point Process Modelling of Crime'.
9. Bachner, 'Predictive Policing: Preventing Crime with Data and Analytics', p. 26.
10. Matthew Fielding and Vincent Jones, '"Disrupting the Optimal Forager": Predictive Risk Mapping and Domestic Burglary Reduction in Trafford, Greater Manchester', *International Journal of Police Science & Management* (Vol. 14, No. 1, 2012), pp. 30–41.

Beyond Predictive Mapping

Despite being shown in the UK to be around ten times more effective in predicting crime than random patrolling,[11] predictive hotspot mapping is not without its limitations. As depicted in Figure 3, software such as PredPol generates grid-based predictions, which are indifferent to the geographical landscape. More recent research has shown that a calibrated network-based model that generates street segment predictions delivers a significantly higher level of predictive accuracy than traditional grid-based methods.[12] Street segment predictions are more useful for policing purposes than arbitrary grids, and these promising findings suggest that further research is needed to refine such models and integrate them within existing predictive policing software.

Figure 3: Example of PredPol Crime Predictions



Source: Michell Eloy, 'Concerns Arise Over New Predictive Policing Program', WABE90.1, 23 September 2013.

Another shortcoming of tools such as PredPol is that they rely entirely on the analysis of past events to predict future crime. They do not take into account the many complex and dynamic environmental factors that make certain places more vulnerable to crime. This 'environmental

11. Kent Police, 'PredPol Operational Review'.
12. Gabriel Rosser et al., 'Predictive Crime Mapping: Arbitrary Grids or Street Networks?', *Journal of Quantitative Criminology* (Vol. 33, No. 33, September 2016), pp. 569–94.

backcloth'[13] – resulting from the interplay between routine behavioural patterns and the physical geography of urban environments – is equally important as past crimes when examining the distribution of hotspots.[14]

These underlying environmental factors are more difficult to analyse than past criminal events. Perhaps the most successful attempt is Risk Terrain Modelling (RTM), which takes into account any factors that are believed to relate to a particular outcome, and assigns a value corresponding to the presence, absence and intensity of each of these factors to different areas in geographical space.[15] Each value forms a map layer, and these are then combined using a geographic information system to produce a composite map, showing a combined risk factor for the particular outcome of interest. This resulting map is a risk terrain map.

RTM shows promise in predicting where crimes are likely to occur, and has displayed better predictive power than retrospective hotspot mapping.[16] However, RTM and predictive mapping are not mutually exclusive, as RTM generates predictions about where crime is most likely to occur over a period of several months, whereas predictive mapping generates predictions for the next few days or weeks. The two therefore serve different purposes: RTM can be used to gain an understanding of which locations are most at risk of experiencing crime over a sustained period of time, while predictive mapping can be used to predict where crime is most likely to occur within a relatively short timeframe.

RTM presupposes a much more detailed insight into the complex risk factors that contribute to different crime problems, and as such it requires a significantly broader body of data from which to compute. Nevertheless, analysis of this kind is becoming increasingly viable given the availability of large volumes of complex data that the big data revolution provides.

It is important to also bear in mind that the use of predictive mapping results in a prioritisation of resources that is based simply on crime volume, rather than harm caused. In any given area, crimes that occur most frequently will receive a greater police response than those that occur less frequently, despite some crimes causing more harm than others.

Some forces allocate resources using more sophisticated measures of risk assessment, for instance using models that take into account a range of factors including: reported crime; number of households; night-time economy; population demographics; deprivation factors; and professional judgement. The MoRiLE (Management of Risk in Law Enforcement) Matrix is a tool that calculates the total harm caused by different types of crime, by taking into account

13. Paul J Brantingham and Patricia L Brantingham (eds), *Environmental Criminology* (Beverly Hills, CA: Sage, 1981).
14. Brantingham and Brantingham, 'Criminality of Place: Crime Generators and Crime Attractors', *European Journal on Criminal Policy and Research* (Vol. 3, No. 3, 1995), pp. 5–26.
15. Joel M Caplan, Leslie W Kennedy and Joel Miller, 'Risk Terrain Modeling: Brokering Criminological Theory and GIS Methods for Crime Forecasting', *Justice Quarterly* (Vol. 28, No. 2, 2011), pp. 360–81.
16. Leslie W Kennedy, Joel M Caplan, Eric Piza, Risk Clusters, 'Hotspots, and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocation Strategies', *Journal of Quantitative Criminology* (Vol. 27, No. 3, September 2011), pp. 339–62.

various factors, such as the harms caused to an individual, an organisation, and the community, as well as factors such as public expectation and financial damage.[17] Applying big data to these models would allow crime problems to be understood in much greater depth, allowing risk to be assessed in terms of total harm caused, rather than simply the volume and spatial distribution of crime.

Predictive Risk Assessment of Individuals

The police's current use of data analytics is heavily location-based, but the uses of big data extend far beyond spatial analysis. Data can also be used to calculate the risks associated with particular individuals. Forces routinely collect information on known offenders, and this data could prove valuable in identifying potential repeat offenders, especially when analysed in combination with data from partner agencies.

While analysts currently use matrix predictions to manually identify high-risk offenders – such as priority firearms offenders – this process is arduous and time-consuming, with analysts reporting that 'it could take weeks'.[18] Algorithmic risk assessment tools can now be used to automate much of this process. Such technology has already been implemented in the US: in 2016, the Supreme Court of Wisconsin used a risk assessment tool – COMPAS – to conclude that the defendant in question posed a great enough risk to society to be ineligible for probation.[19]

Durham Constabulary is currently in the process of developing an AI-based system to assess the risk of individuals reoffending.[20] The Harm Assessment Risk Tool (HART) uses as its input variables an individual's past offending history, their age, postcode and other background characteristics. Predictive algorithms are then used to classify each individual as being at low, medium or high risk of re-offending.

The system was tested initially in 2013 and the results monitored over the following two years. The model was found to predict low-risk individuals with 98% accuracy, and high-risk individuals with 88% accuracy.[21] This disparity reflects the fact that the model favours classifying individuals as medium- or high-risk in order to reduce the likelihood of false negatives. These preliminary results suggest that algorithmic risk assessment tools are able to predict the likelihood of an individual re-offending with a high degree of accuracy, and therefore warrant further research.

However, the HART model only uses data from the individual force in question, and does not incorporate data from other forces, PNC or partner agencies. For example, a repeat offender,

17. For more information on the MoRiLE Matrix, see Amanda Huggins, 'MoRiLE: Management of Risk in Law Enforcement', October 2015, <http://www.excellenceinpolicing.org.uk/wp-content/uploads/2015/10/1-3_MoRiLE.pdf>, accessed 20 July 2017.
18. Author interviews with four police analysts, London, 18 April 2017.
19. Joe Palazzolo, 'Wisconsin Supreme Court to Rule on Predictive Algorithms Used in Sentencing', *Wall Street Journal*, 6 June 2016; State of Wisconsin vs. Eric L Loomis, Supreme Court of Wisconsin, 2015AP157-CR, 13 July 2016.
20. Chris Araniuk, 'Durham Police AI to Help with Custody Decisions', *BBC News*, 10 May 2017.
21. *Ibid*.

who has recently moved from another jurisdiction will not have their offending history recorded in the model, and therefore will not appear to be at increased risk of re-offending. The impoverished nature of these very local datasets will inevitably increase the occurrence of false negatives, at worst causing dangerous individuals to slip through the net. The use of national datasets is essential to explore the full potential of such predictive tools.

Another issue to address is how predictions should be acted upon when an individual is identified as posing an increased risk. Acting upon such predictions may result in negative social effects, such as perpetuating bias or racial discrimination.[22] A *ProPublica* investigation into the COMPAS sentencing algorithm found that only 20% of individuals identified as likely to commit a violent crime actually did so, and that black defendants were twice as likely to be deemed at risk of offending than white defendants.[23]

Systems underpinned by machine learning will inevitably reproduce the inherent biases present in the data they are provided with – if particular minorities have been disproportionately targeted by police action in the past, the algorithm will disproportionately assess those individuals as posing an increased risk. Acting on these predictions will then result in those individuals being disproportionately targeted by police action, creating a 'feedback loop' by which the predicted outcome simply becomes a self-fulfilling prophecy.[24] Perhaps partly for this reason, Durham Constabulary's HART system is intended to function purely as an 'advisory' tool, with officers retaining ultimate responsibility for deciding how to act on the predictions. Nevertheless, it could be argued that individual-level analysis of this kind should instead be carried out by probation services, the judiciary and other third-party organisations, which are able to maintain a greater degree of separation from the data, and therefore a higher level of objectivity when interpreting the inherent biases present in a dataset.

Similarly, predictive risk analysis could be used to identify potential victims and vulnerable individuals for safeguarding purposes. Much police time is taken up investigating missing persons' cases, especially children. Partner organisations hold a large amount of relevant data which could be analysed to better understand missing persons' problems, and ultimately form the basis of predictive risk assessment tools.

There are, however, major barriers preventing partner agencies from sharing data with the police, with officers reporting that they have very limited access to data from social services. Even when partner agencies are able to share data, social services and local authorities have no evening or weekend provision, meaning that such information can only be accessed during office hours. This is especially problematic considering most missing persons investigations occur outside office hours.

22. Ryan Calo and Kate Crawford, 'There is a Blind Spot in AI Research', *Nature* (Volume 538, No. 7625, October 2016); Lee Rainie and Janna Anderson, 'Code-Dependent: Pros and Cons of the Algorithm Age', Pew Research Center, 8 February 2017, p. 57.
23. Julia Angwin et al., 'Machine Bias', *ProPublica*, 23 May 2016.
24. Stephen Buranyi, 'Rise of the Racist Robots – How AI is Learning All our Worst Impulses', *The Guardian*, 8 August 2017.

Since 2011, Multi-Agency Safeguarding Hubs (MASHs) have been set up in some areas to promote inter-agency collaboration and provide a single point of contact for the safeguarding of children and young people. While the initiative has proved successful overall, senior officers report that data sharing is limited, and there is no shared MASH database.[25] As a result, local authorities are usually tasked with performing multi-agency analysis on behalf of the police. It is clear that individual-level analysis of this kind would be significantly more effective if the necessary data-sharing agreements were implemented.

Visual Surveillance

More than ten years ago, George Washington University law professor Daniel J Solove introduced the notion of the 'digital dossier', describing how passive surveillance and digital data collection allow every aspect of our lives – daily routines, shopping preferences, travel habits and so on – to be preserved forever 'in vast databases with fertile fields of personal data'.[26] As the costs of data storage have fallen year on year and the density of storage capacity is continually expanding,[27] collection of this kind is now a matter of daily routine for governments, law enforcement and the private sector.

In addition, increased information sharing between the public and private sectors has resulted in data that was once kept separate being merged to form more complete databases. This pulls together information from many different sources to create comprehensive profiles of individuals and their activities – moving towards what has been described as 'perfect memory'.[28]

The UK is one of the world's foremost surveillance states, with one of the highest numbers of CCTV cameras per capita of any country,[29] and – as of November 2016 – the most comprehensive powers of digital surveillance ever seen in a democracy.[30] The Investigatory Powers Act 2016 introduced new powers allowing the UK intelligence agencies and police to carry out not just targeted interception of communications, but also indiscriminate 'bulk' collection and analysis of communications data. It also requires CSPs to store all UK citizens' internet connection records and provide these records on request to the police, intelligence agencies and other authorities without requiring a warrant. In December 2016, the European Court of Justice ruled that the general and indiscriminate collection of electronic communications is illegal under EU law.[31] However, it is at present unclear whether these provisions of the Act are currently being implemented by authorities in the UK.

25. Author interviews with a superintendent and a chief inspector, London, 19 April 2017.
26. Daniel J Solove, *The Digital Person: Technology and Privacy in the Information Age* (New York, NY: NYU Press, 2004), p. 1.
27. Patricia L Bellia, 'The Memory Gap in Surveillance Law', *University of Chicago Law Review* (Vol. 75, No.1, Winter 2008), p. 143.
28. *Ibid*., p. 141.
29. Tom Reeve, 'How Many CCTV Cameras Actually Are There in the UK?', *arc24* blog, 18 May 2016.
30. 'Investigatory Powers Act 2016 (UK)'.
31. Madhumita Murgia, George Parker and Jim Brunsden, 'EU's Highest Court Declares UK Surveillance Powers Illegal', *Financial Times*, 21 December 2016.

Despite these far-reaching surveillance capabilities, law enforcement agencies are often overwhelmed by the sheer volume of data collected through digital surveillance methods, and lack the technological capabilities to use it for operational purposes.[32] As a result, there has been increased demand for reliable and inexpensive software solutions that the police can use to manage the vast quantities of data collected through this type of surveillance.

One example of such software is the Domain Awareness System (DAS), first developed in 2012 by the NYPD in collaboration with Microsoft.[33] The DAS analyses data from multiple sources, such as CCTV cameras, automatic number plate readers and radiation sensors, to detect potential security threats in real time.[34] The DAS is able to automatically cross-reference this data with information held on police databases in order to provide officers with immediate access to individuals' known criminal histories and recent movements.

Systems such as the DAS can detect connections between seemingly disparate pieces of data, identifying threats that would easily go unnoticed by individual crime analysts. Crucially, this can be done in real time, for instance at public events where large numbers of people make policing even small areas significantly more difficult. For example, in November 2013, following the Boston Marathon bombings in April of that year, the NYPD deployed hundreds of temporary cameras along the New York City Marathon route, allowing them to monitor nearly every portion of the race route in real time using the DAS.[35]

As part of their partnership with Microsoft, the NYPD will receive 30% of all revenue generated by the sale of DAS programs to other agencies nationwide.[36] Additionally, the initial costs could well be outweighed by the efficiency savings made by streamlining police practice, but these savings are much harder to quantify.

Adopting systems such as the DAS would allow police forces to take full advantage of the UK's extensive digital surveillance network; at present the police's analysis of such data is almost entirely manual.

Open-Source Analytics

In 2015, the Policing Hate Crime project was launched in the UK, funded by the Police Knowledge Fund, and implemented by a consortium formed of the Metropolitan Police, the think tank Demos, the University of Sussex, CASM Consulting and Palantir Technologies.[37] The project aims

32. Joh, 'Policing by Numbers', p. 48.
33. Michael Endler, 'NYPD, Microsoft Push Big Data Policing into Spotlight', *DARKReading*, 20 August 2012.
34. Joh, 'Policing by Numbers', p. 49.
35. Michael Schwirtz, 'After Boston Bombings, New York Police Plan Tight Security at Marathon', *New York Times,* 1 November 2013.
36. Rebekah Morrison, 'New York's Domain Awareness System: Every Citizen Under Surveillance, Coming to a City Near You', blog of the *North Carolina Journal of Law & Technology*, 23 February 2016.
37. University of Sussex, 'Social Media's "Big Data" Could Help Police Understand Hate Crime', 9 October 2015.

to use big data software to reveal insights from social media that could be used to anticipate and prevent hate crime incidents.

A preliminary report by Demos examined the use of migration-related language on Twitter in the run-up to the June 2016 referendum on Britain's membership of the EU.[38] Analytical algorithms were used to draw connections between tens of thousands of tweets, with results showing the patterns of anti-migration comments over time, as well as geolocation information and spatial trends.

This type of study demonstrates that data from social media can provide a detailed insight into the beliefs and activities of many thousands of individuals. When combined with data from closed sources, analysts have at their fingertips digital dossiers of the kind described by Solove for many millions of citizens.[39] Use of advanced analytics would allow the police to develop proactive crime-fighting strategies to make effective use of this massive amount of data.

Additionally, the rise of the Internet of Things brings with it the emergence of 'smart cities', which use real-time monitoring systems and electronic sensors to continuously transmit urban telematics data.[40] Such data can be used to enhance the performance and efficiency of urban services – such as transportation networks, hospitals and schools – but could also transform the way cities are policed.

For instance, data transmitted from smart cities could be used to anticipate crime problems and pre-emptively respond to emerging threats before they develop. The 2015 Spring Budget included £140 million investment for research into smart cities and the Internet of Things.[41] In the coming years, police forces and developers of policing technology should pay close attention to new opportunities presented by the growth of the Internet of Things and the emergence of smart cities.

38. See Demos, 'Hate Speech After Brexit', 11 July 2016.
39. Solove, *The Digital Person*.
40. Tim Sandle, 'Telematics is Shaping the '"Smart City"', *Digital Journal*, 16 August 2017.
41. HM Treasury, 'Budget 2015: Some of the Things We've Announced', 18 March 2015.

IV. Challenges and Solutions

THIS CHAPTER OUTLINES some of the main challenges currently preventing police forces from making effective use of big data technology. These include organisational and cultural barriers to developing and implementing new systems, limitations resulting from financial cuts to police resources and ethical and legal barriers concerning the police's use of personal data. Potential solutions to these challenges are identified, based on feedback received in the course of interviews.

Isolation of Local Police Forces

'If it's not in our borough, we're not interested'.[1] While the superintendent making this claim was referring to his team specifically, this sentiment was echoed by all officers interviewed. Crime prevention strategies are developed in response to specific local crime problems, with little collaboration between forces, and virtually non-existent national oversight. Frontline officers report frustration at not being able to see the bigger picture of crime problems, as responses are restricted according to arbitrary geographical boundaries, while criminals, of course, are not.

There are 43 territorial police forces in England and Wales (see Figure 4), each headed by a chief constable (with the exception of the Metropolitan Police and the City of London Police, whose chief officers hold the rank of commissioner). Until 2012, forces outside London were governed by police authorities – public bodies responsible for overseeing all force operations. In 2012, in response to concerns over a perceived lack of accountability of police forces, police authorities were abolished and replaced with publicly elected Police and Crime Commissioners (PCCs).

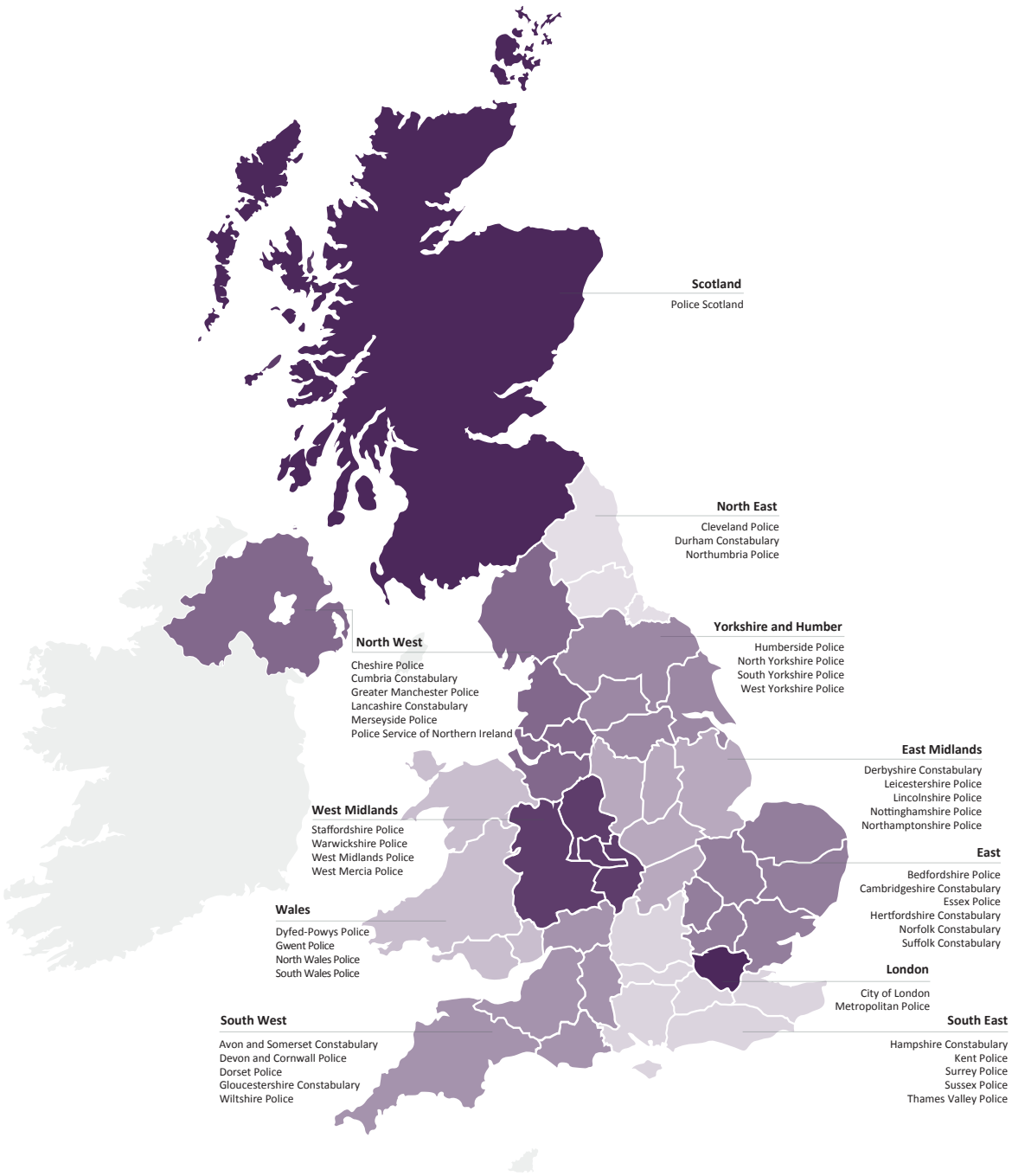
The role of PCCs is to hold forces and their chief constables to account, ensuring they are operating effectively and efficiently. This includes formulating police and crime plans which set policing objectives, as well as taking responsibility for all budget decisions.[2]

As a result of this highly localised, devolved governance structure, forces exist in relative isolation, with chief constables and PCCs retaining a great degree of autonomy over local policing strategies, organisational change, and crucially for this report – technological development. While national organisations such as the National Police Chiefs' Council and HMIC routinely scrutinise police use of technology and provide specific recommendations to individual forces, the chief constable and PCC are ultimately responsible for deciding whether or not to act on these recommendations.

---

1. Author interview with a superintendent, London, 18 April 2017.
2. Association of Police and Crime Commissioners, 'Role of the PCC', <http://www.apccs.police.uk/role-of-the-pcc/>, accessed 7 July 2017.

Figure 4: Territorial Police Forces in the UK



Source: HMIC, 'Map of police forces in England, Wales, and Northern Ireland', 21 February 2013.

Because of this, chief constables typically pursue new technological initiatives independently, with little or no collaboration with other forces. This results in the wide variation in forces' technological development discussed in Chapter II. As time goes on and individual forces pursue different forms of technological change, collaboration between forces becomes increasingly difficult. This is especially problematic in the context of big data analytics, which relies upon effective nationwide data sharing and collaboration.

As forces across the country begin to make use of predictive analytics for assessing the risks associated with individuals, they will need to make more effective use of national data. Local datasets provide only an incomplete understanding of these risks, as criminals and victims are not restricted according to arbitrary geographical boundaries. Moreover, the police's procurement of big data technology must be coordinated at the national level to ensure that future investment in this area is not wasted.

Culture, Buy-In and Training

The research for this paper identified significant issues surrounding officer buy-in and attitudes towards digital policing. In the case of predictive hotspot mapping, several forces in the UK have successfully implemented such software, only to find that the predictions generated are not translated into operational activity. A senior officer described a 'clash in perspectives' between those officers keen to embrace the opportunities presented by new technologies, and those who feel that the use of such tools diminishes the importance of their professional judgement.[3]

This clash in perspectives became apparent through the course of interviews. Several officers recognised and appreciated the value of predictive hotspot mapping, while others believed that such tools were 'gimmicks' that had no place in local policing strategies. This is likely due to the fact that officers had not received any training material or information explaining how such software works, why it is useful and how effective it has proven to be.

Improving buy-in to such initiatives therefore requires engaging officers from the outset, during the planning stages of the programme. This would give officers a better understanding of why new technologies are desirable, increasing the likelihood of them buying-in to such initiatives. Front line officers complained that they had never been involved in conversations regarding technological development, with one sergeant explaining that 'it's really noticeable that the systems we use were not designed through consulting with police officers, and certainly not operational police officers'.[4]

Training is another area of major concern. Technological training is virtually non-existent for police officers, and constables explained how they are required to 'learn on the job' and be taught by more experienced officers how to use technology as fundamental as their radios and car computers.[5] Senior officers have repeatedly highlighted this as a problem. For example,

3. Author interview with a detective inspector, conducted by telephone, 5 April 2017.
4. Author interview with a police sergeant, London, 19 April 2017.
5. Author interviews with police constables, London, 19 April 2017.

Assistant Commissioner Rowley, when speaking in 2013 on the issue of mobile policing, told the London Assembly's Budget and Performance Committee that 'the intention is to do pretty much zero training'.[6]

Investment in new technology will not be cost effective if officers on the ground are not trained how to use it, and the costs of initial training are likely to be outweighed by the efficiency savings made in the long term.

Analytics to Action

In recent years, financial cuts have significantly reduced the analytical capabilities of UK police forces. For example, in the MPS, restructuring changes from 2012 to 2013 saw the elimination of Borough Intelligence Units from local police departments. In the past, if the police wanted to better understand a particular crime problem in their local area, an in-house analyst would be tasked with producing a report and providing recommendations for action. Instead, local police must now request such analytical products from a centralised body, and if a request 'did "not fit the control priorities and the properties of the Met", it may be acted on more slowly or not at all', according to Chief Superintendent Simon Laurence, Borough Commander of Hackney Police.[7]

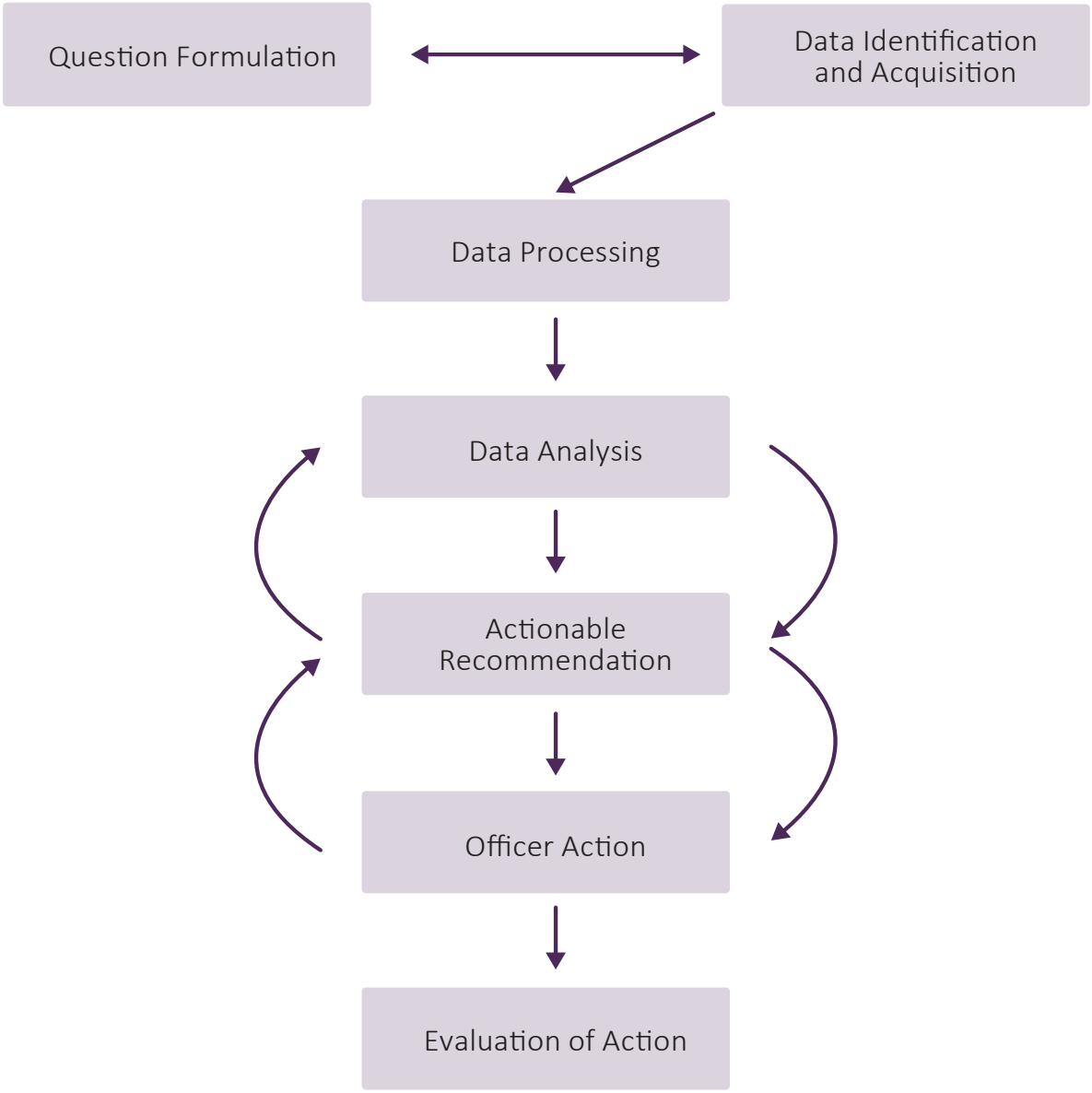
This centralisation has divorced the analytical and operational processes. For the crime analysis process to function successfully, it relies on regular communication between officers and analysts.[8] The process is iterative rather than linear, as depicted in Figure 5 below, a model of crime analysis.

With this model in mind, the problems with adopting a centralised analysis system are clear. Far from communicating and working collaboratively to inform each other's work, officers and analysts operate entirely independently, in separate buildings, and will likely never meet. Analytic products are not refined and updated as local crime problems change and evolve, but rather are delivered weeks or even months after they are requested (if at all), by which time the crime problem under investigation has changed, requiring new analysis. This is a far cry from the real-time analytics of the kind that modern software provides, as it is updated multiple times a day as new data is collected.

This problem may be remedied in part by the provision of mobile technology to officers. As the London Assembly Budget and Performance Committee highlighted, '[o]fficers using mobile devices with access to the force's ICT systems may eliminate the need for staff to duplicate tasks such as data entry. As a result, the force might need fewer back-office staff as some tasks become automated'.[9] The automation of basic data tasks would release back office staff, allowing them to spend more time carrying out crime analysis in closer collaboration with police officers.

6. London Assembly Budget and Performance Committee, *Smart Policing*, p. 25.
7. Josh Loeb, 'Yet Another Stabbing in Hackney – But Police do not Believe Recent Incidents of Knife Attacks Are Related', *Hackney Citizen*, 30 January 2017.
8. Bachner, 'Predictive Policing: Preventing Crime with Data and Analytics', p. 13.
9. London Assembly, Budget and Performance Committee, *Smart Policing*.

Figure 5: Model of Crime Analysis



Source: Bachner, 'Predictive Policing: Preventing Crime with Data and Analytics', p. 11.

Nevertheless, it is important to reiterate that new technologies should not be seen as a replacement for traditional policing, but rather as a supplement used to inform and support traditional policing strategies. Software packages are merely another tool in a police officer's arsenal, while 'humans remain – by far – the most important elements in the [predictive policing] process'.[10]

10. Perry et al., *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations,* p. 117.

Cybercrime and Big Data-Enabled Crime

Another important challenge to address is the issue of criminal innovation. While big data can be harnessed by law enforcement agencies seeking to prevent and disrupt criminal activity, it can also prove valuable to criminals, enabling them to commit more sophisticated cyber-enabled crimes.

Organised crime groups are now able to exploit big datasets to carry out complex types of fraud on a much larger scale, a factor listed as a 'key driver for change' in Europol's 2015 report on the future of organised crime,[11] and more recently in their 2017 Serious and Organised Crime Threat Assessment. The latter report states:

> For almost all types of organised crime, criminals are deploying and adapting technology with ever greater skill and to ever greater effect. This is now, perhaps, the greatest challenge facing law enforcement authorities around the world, including in the EU. ... Network intrusions for the purpose of illegally acquiring data have significant impact globally, resulting in the loss of intellectual property and the compromise of mass amounts of data which can be used for further criminality including fraud and extortion.[12]

As the number of wirelessly connected devices constantly increases, so do the opportunities for hackers and criminals to surreptitiously intercept and gather personal data.

Biometric information, such as that used for identification technologies, is becoming cheaper to collect and process. As such information becomes more readily available to state authorities and the private sector, it also becomes more accessible to organised crime groups and cybercriminals seeking to exploit the data for illegal activities.

Chief Constable Stephen Kavanagh of Essex Police recognises that '[c]riminals are exploiting technology, and the tools to preserve anonymity online, more quickly than law enforcement is able to bring new techniques to bear'.[13] Great care must therefore be taken to ensure that sufficient security measures are in place to prevent law enforcement data and technologies developed for public service agencies from falling into the wrong hands.

Legal and Ethical Use of Data

The big data revolution brings with it complex ethical questions and the ethical implications of big data are as yet poorly understood.[14] As the use of such technology becomes increasingly

11. Europol, *Exploring Tomorrow's Organised Crime* (The Hague: Europol, 2015), p. 19.
12. Europol, *Serious and Organised Crime Threat Assessment 2017*, pp. 24, 28.
13. College of Policing, National Crime Agency and NPCC, 'Digital Investigation and Intelligence: Policing Capabilities for a Digital Age', April 2015, p. 6.
14. Boyd and Crawford, 'Critical Questions for Big Data', p. 672.

widespread, legislative frameworks must expand to incorporate 'new rules to regulate the societal cost of our new tools without sacrificing their undeniable benefits'.[15]

Ethical concerns around the use of data typically focus on the collection, analysis and dissemination of personally identifiable information. In the UK, personal data is subject to the Data Protection Act 1998, which dictates that personal information must be used fairly, lawfully, and for limited, specifically stated purposes.[16] The EU General Data Protection Regulation[17] will introduce further rules governing the collection and use of personal data, and Directive EU 2016/680[18] will legislate for the processing of personal data for policing purposes. Both will come into effect in May 2018, and both apply to the automated analysis of personal data as well as manual analysis.

Crucially, data protection legislation in all its forms does *not* apply to anonymised datasets.[19] For this reason, the Law Enforcement Directive suggests that organisations should aim to 'pseudonymise' datasets as early as possible, to facilitate 'the free flow of personal data within the area of freedom, security and justice'.[20] Pseudonymisation is defined as:

> the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.[21]

Data protection legislation restricts large-scale exploratory analysis of personal data, such as the predictive risk assessment methods discussed in Chapter III. For this reason, pseudonymisation is likely to be the only way to perform big data analytics on personal datasets while complying with data protection laws. The use of advanced analytics has the potential to increase anonymity in this regard, as algorithmic machine learning-based systems do not require the input of a human observer during the analytical process. Data can be anonymised prior to analysis, and the identities of individuals of interest only revealed once the analytical process has concluded.

---

15. Neil M Richards and Jonathan H King, 'Big Data Ethics', *Wake Forest Law Review* (Vol. 49, May 2014), p. 409.
16. 'Data Protection Act 1998 (UK)'.
17. Council of the European Union, 'Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)', *Official Journal of the European Union* (L 119/1, 4 May 2016).
18. Council of the European Union, 'Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data by Competent Authorities for the Purposes of the Prevention, Investigation, Detection or Prosecution of Criminal Offences or the Execution of Criminal Penalties, and on the Free Movement of Such Data, and Repealing Council Framework Decision 2008/977/JHA', *Official Journal of the European Union* (L 119/89, 27 April 2016).
19. *Ibid*., para. 21.
20. *Ibid*., para. 53.
21. *Ibid*., chapter 1, article 3, para. 5.

However, issues of accountability arise from the use of such exploratory analysis techniques, as the purposes and scope of analysis are often not defined in advance. Moreover, the use of machine learning to automate analysis means that the analytical process itself is typically opaque to an observer.[22] This means that organisations may be unable to defend the decisions made on the basis of analytical outputs, as it is impossible to explain exactly how these analytical conclusions were reached.

Analysis of anonymised *crime* data (rather than personal data) is far less problematic from an ethical and legal perspective, and this is partly why this paper focuses heavily on the use of predictive mapping software for crime prevention purposes. Analysis of this kind does not require the use of any personal information, and as such is significantly more straightforward for the police to implement than other big data technologies.

Different considerations arise when using data from public sources, such as social media. When an individual broadcasts information in the public domain, this does not entail that they automatically consent to this information being collected and analysed without their knowledge.[23] Privacy is not a binary concept, but rather 'virtually all information exists in intermediate states between completely public and completely private'.[24]

Internet users create data in specific, context-sensitive spaces, and do not expect it to be used in other contexts, for unrelated purposes. The use of this data must therefore be clearly justified by specific aims and requirements. This is partly why there are restrictions on the police's use of open-source analysis software, as discussed in Chapter II. However, when such analysis *is* justified and necessary for a specific policing purpose, there is no reason why investigators should not be able to carry it out themselves, using their own computer terminal, rather than requiring it to be carried out on their behalf, or having to use a dedicated terminal.

At present, while the police's use of data is legally governed by data protection legislation, there is no clear decision-making framework for the ethical use of big data technology in law enforcement. Data protection legislation is highly complex and nuanced, and practitioners have no source of accessible and practical guidance on what constitutes the appropriate use of data. The UK government's 'Data Science Ethical Framework',[25] published in May 2016, is a short and deficient document outlining six broad principles for the ethical use of data, which seem to focus on public opinion and perception of organisations rather than ethics themselves. The document provides little in the way of practical advice about how to tackle common ethical problems surrounding the use of data, and offers no insight into what constitutes inappropriate use of data analytics.[26]

22. Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection', p. 9.
23. Boyd and Crawford, 'Critical Questions for Big Data', p.672.
24. Richards and King, 'Big Data Ethics', p. 413.
25. Cabinet Office, 'Data Science Ethical Framework, Version 1', 19 May 2016.
26. Charles Raab and Roger Clarke, 'Inadequacies in the UK's Data Science Ethical Framework', *European Data Protection Law Review* (Vol. 2, No. 4, December 2016).

A more coherent decision-making framework for the use of big data in policing is Algo-care, developed in collaboration with Durham Constabulary,[27] which provides a clear and comprehensive list of criteria that must be met to justify the application of algorithmic methods to personal data sets. The framework aims to ensure that algorithmic big data technologies can be used for policing purposes without violating data protection legislation.

Given the increasing power and availability of advanced analytical tools and techniques, it is imperative that practitioners are provided with an adequate decision-making framework to ensure the ethical use of big data technology for law enforcement purposes. The EU General Data Protection Regulation and Directive EU 2016/680 will have significant implications for the ways in which police forces are able to legally use data, and by the time these laws come into effect in 2018, the government should have developed clear guidelines for the benefit of police officers and staff who are required to handle sensitive data.

27. Marion Oswald and Sheena Urwin, 'Algorithms in Policing: Take Algo-Care™: Written evidence submitted by Marion Oswald, Senior Fellow in Law and Director of the Centre for Information Rights, University of Winchester, and Sheena Urwin, Head of Criminal Justice, Durham Constabulary (ALG0030) ', written evidence before the House of Commons Science and Technology Committee, April 2017.

Conclusion

THE RESEARCH FOR this paper found that police forces in the UK have access to a vast amount of digital data, but lack the technological capabilities to make effective use of it. Big data analytics has already proved revolutionary in a range of other domains, and similarly has the potential to transform many aspects of policing.

Use of big data technology could automate currently arduous and time-consuming tasks – such as the manual analysis of communications data. Effective use of predictive analytics would allow UK police forces to develop proactive crime-fighting strategies, targeting resources to where they are most needed, rather than simply responding to crime events when they occur. Algorithmic risk assessment tools could be used to predict the risks associated with individuals – for instance, to predict reoffending, or to identify vulnerable individuals who are at increased risk of going missing or coming to harm. Open-source data could be used to gain a deeper understanding of different crime problems across a very large section of society.

Deficiencies in the police's core IT infrastructure currently present a significant barrier to the implementation of big data technologies. However, these issues are by no means insurmountable. It is expected that in the coming years, technological advances will mean that forces have access to the infrastructure required to successfully incorporate big data technology into their policing strategies.

Perhaps more concerning are the organisational and cultural barriers discussed in this paper. When new technologies have been adopted by police forces in the UK, they have often not been implemented consistently and officers have not received adequate information about what the technology is or training in how it operates, meaning they choose not to make use of it.

But most problematic of all is the highly localised structure of UK policing. Forces pursue technological change independently in response to local requirements, with little inter-force coordination. Although there are regional structures and partnerships in place, the wide variation in the level of technological development makes it difficult for forces to collaborate when designing new technology. It is imperative that national policies and strategies are developed to create coherence between forces seeking to implement new technologies. Only when this is achieved can police forces in the UK hope to make effective use of big data technology.

Many other issues are beyond the scope of this paper. Perhaps most significant is the challenge of investing in new technologies while faced with budget cuts in excess of 20%, equating to a reduction in staff since 2010 of around 20,000 police officers across England and Wales.[1] In spite of this, it must be stressed that efficient IT systems will allow forces to make more effective

1. Rowena Mason and Peter Walker, 'Under-Fire Theresa May Hits Back over Police Cuts', *The Guardian*, 5 June 2017.

use of the limited resources available to them. For this reason, despite the budget cuts of the kind imposed since 2010, it is crucial to invest in new technology, as the costs of the initial investment will be more than recuperated by the efficiency savings made in the long term.

About the Author

Alexander (Sacha) Babuta is a Research Analyst specialising in policing and organised crime within the National Security and Resilience group at RUSI. His research focuses on policing in the digital age, transnational organised crime and cross-border trafficking. He holds an MSc in Crime Science from University College London (UCL), where his dissertation examined the nature, characteristics and police response to missing child incidents. He also holds a Bachelor's degree in Linguistics from UCL. Prior to joining RUSI, Sacha worked for a Member of Parliament.