DATAJUSTICE
ERC-funded project ‘Data Justice: Understanding datafication in relation to social justice’ (DATAJUSTICE) starting grant (2018-2023).

WORKING PAPER

Data Driven Policing in the Context of Europe

Fieke Jansen
JansenF[at]cardiff.ac.uk
Cardiff University

May 7th, 2018
Abstract

This report provides an overview of the data-driven policing technologies currently being integrated into European police forces. It looks at the following four data-driven policing trends: expansion of police databases, implementation of real time identification systems, use of predictive policing technology and the analysis of heterogeneous datasets. The report offers a non-exhaustive list of programs, primarily identified in Western and Northern parts of Europe.

Contents

1 Introduction into data-driven policing 2
2 Construction, merging and enhancement of police databases 2
   2.1 Investment in national police databases .............................. 2
   2.2 EU databases .......................................................... 3
3 Real Time Identification systems 5
   3.1 Automated License Plate Readers (ANPR) .......................... 5
   3.2 Automatic Facial Recognition systems ............................... 6
   3.3 Voice identification systems ......................................... 6
4 Predictive policing 7
   4.1 Predictive mapping ..................................................... 7
   4.2 Predictive identification ............................................... 8
5 Analysis of heterogeneous databases 9
6 Conclusion 10
7 References 11
1 Introduction into data-driven policing

It can be observed that European police departments are turning to the integration of data-driven decision-making to prevent and investigate crime. A growing body of research indicates that while the collection and use of data has always been a part of policing practices, the rise in the availability of data and increased sophistication of technology is shifting policing from mostly reactionary police work, solving crime after it had occurred, towards more pre-emptive policing (Brayne, 2017: 986: Brakel, 2016: Hardyns et al, 2017). While there is a clear desire to become more data or intelligence-driven the integration of data-driven policing technologies is still in its infancy in Europe. The shift in the nature of policing can predominantly be found in the language around policing and in niche programs, which are developed for or tested by individual police forces (Knobloch, 2018; Couchman, 2019). To gain an understanding of the extent to which Europe is integrating data and technology in its police practices, this report identifies the following data-driven policing trends:

1. Construction, merging and enhancement of databases
2. Real-time identification and tracking of individuals
3. Predictive policing
4. Analysis of heterogeneous datasets
5. Fighting cybercrime

The report will dive into each of these trends, offering a non-exhaustive list of programs, primarily identified in Western and Northern parts of Europe. Police strategies on cybercrime are not included in this report, as these types of crimes are native to the internet and require investigators to turn to digital forensics.

2 Construction, merging and enhancement of police databases

The collection and storing of data about crime related to individuals, objects and events are an integral part of modern police work. It is enabled by the quantification of everyday life, has significantly expanded police capabilities to collect data, and technological advances which allow for more sophisticated data analysis (Brayne, 2017; Ferguson, 2017). Legal mandates for police to process data have expanded and allowed for more data collection, longer data retention periods, and data sharing (Home Office, 2016A; European Data Protection Supervisor, 2018). The desire to implement real-time identification and predictive policing programs require higher quality police databases. To increase the quality of police data, police officers are encouraged to see paperwork or data entry as a core part of their work. The more accurate, timely and complete the data entry, the more efficient the police force can operate (Brink et al, 2017). Alongside investment in police officers’ data entry practices, individual nation states and the EU invest in the interoperability between databases, the creation of new databases with a specific mandate and enable automated searches on key functionalities, like biometric identifiers. The below section explores investments in national and European police databases.

2.1 Investment in national police databases

The U.K and Germany are making significant investments to update and increase interoperability between different police databases. The U.K has a federated law enforcement structure, which historically has provided local police forces autonomy to buy and implement their own policing technologies and has led to the proliferation of different databases and database structures. The Home Office is spearheading by the National Law Enforcement Data Programme (NLEPD), under which it aims to create the National Law Enforcement Data Service (NLEDS)\(^1\), connecting the

\(^1\)https://www.digitalmarketplace.service.gov.uk/digital-outcomes-and-specialists/opportunities/1227
databases of the different U.K police forces into one centralized system. It will bring together data from the Police National Computer (PNC), the Police National Database (PND) and Automatic Number Plate Recognition (ANPR) records (Partnership for Conflict, Crime & Security Research, 2016; Murdock, 2016; Couchman, 2019; Home Office, 2018). The Police National Computer (PNC) is a text-only computer that stores over 10 million records data about arrest, custody records and information on suspects, including links to biometric data, fingerprints and DNA. The Police National Database (PND) stores “soft intelligence” about allegations and/or investigations that did not result in an arrest (Murdock, 2016). The national ANPR database receives around 50 million ANPR ‘reads’ a day (Police.uk, 2019; Surveillance Camera Commissioner, 2016: 23). IBM has been awarded the £12,000,000 contract to assist the NLEPD with the transformation of the existing systems into the NLEDS (Home Office, 2016B). Connecting PNC, PND and ANPR data in one system is expected to change the U.K police landscape.

In Germany, the Federal Crime Police (BKA) administers various police databases to enable data sharing and processing between the 16 ‘Bundeslander’, German states, and the federal state (Monroy, 2018A). The BKA administers several databases, of which the ‘Internal Security’ databases and the Politically Motivated Crime ‘Politisch Motivierte Kriminalität’ (PMK) database are known. The PMK databases stores information about ‘relevant’ persons likely to threaten public safety’. ‘Relevant’ persons are classified as left, right, foreign ideology, and religious ideology. The databases contain data about names, addresses, biometrics, political affiliation, and email address. (Monroy, 2018A; Monroy et al, 2018; Kreckenbaum, 2017). The German police forces can access these databases through the BKA’s INPOL system, which has grown organically over time. The main critique of the current information architecture is that data is siloed into different databases, systems, and file formats, which hold ‘old’ and duplicate data (Bundeskriminalamt, 2019). Under the ‘Police 2020’ programme2 the BKA will invest in the modernization and standardization of the IT infrastructure and interoperability of databases (Bundesministerium des Innen, 2018; Bundeskriminalamt, 2019).

2.2 EU databases

The signing of the Schengen Agreement in 1985 resulted in the disappearance of internal border checks and increased the need to share data among European nations for the purpose of border control and police cooperation. The number of data-driven programmes and subsequently the number of databases has increased since then. The bulk of the existing3 and soon to be established4 databases, monitor who enters and exits the EU, are accessible to law enforcement. This report explores the databases that are of direct interest to understand the police data and technology environment, i.e. EUROPOL’s databases, European Police Registration Information System (EPRIS) and the Schengen Information System (SIS) I & II.

Europol and EPRIS

The European Union Agency for Law Enforcement Cooperation, better known as Europol, aims to “achieve a safer Europe for the benefit of all the EU citizens” (Europol, 2018A). Their main focus is to fight terrorist networks and large scale organized crime, the latter includes cybercrime, drug trafficking, and money laundering, and human trafficking. To achieve this Europol has set up the organized crime and terrorism analysis projects, under which several databases are administered. Europol’s 2015 report to the Council of Europe disclosed that its existing Analysis Work File on Terrorism included three databases: 1) Hydra; prevention and combating of terrorism-related

crimes, including data on individuals, groups, networks or organizations who evoke Islam to justify their actions, 2) Piracy; which covers “East of Africa and Gulf of Aden maritime piracy”, and 3) Dolphin; which includes all other terrorist activities, i.e. right-wing terrorism (Europol, 2015; Europol, 2018B). Europol’s website does not have any current references to the Piracy database but has added the Travellers Database, which contains information on 34,000 “foreign fighters” (EUObserver, 2016). To increase the efficient exchange of police records for criminal activities that fall outside the scope of Europol, several European Member States have successfully lobbied for the creation of EPRIS (Focant et al, 2012). Administered by Europol, the mandate of EPRIS is to provide police with a quick overview if and where police records on an individual can be found in Europe. As such Europol is developing EPRIS as an INDEX, which allows police forces to search in each other’s databases. The result will be ‘hit’/‘no hit’; in the case of a ‘hit’ the querying members state police department can then go through the proper judicial channels to gain access to the information (Jones, 2011). Similarly to the U.K Home Office NLEDS plan, the EU aims to increase interoperability between the different national databases, ensuring that available crime data is accessible and not locked away in data silos.

SIS I & II

SIS is a database containing information on criminal activity, immigration violations, and various objects and missing persons. It is most known for being the EU’s border information management system to which police officers, border patrol officers, and law-enforcement officials have access to (Jansen, 2017). A less known functionality is the use of SIS for police cooperation on criminal investigations. Under article 36 of Council Decision 2007/533/JHA police have the ability to perform discreet or specific checks, with the aim to discreetly gather information (Migration and Home Affairs, 2018). A discrete check allows the issuing police to be informed about an individual’s whereabouts, their activity and the people they are associated with when this individual gets in contact with law enforcement through a routine check, border crossing, surveillance action or investigation anywhere in Europe. The second option under article 36 is the specific check which allows police, in a different European country then of those issuing a specific check, to search luggage and vehicles of an individual, and even question them (Monroy, 2019). Discussions have been ongoing in the European Commission and the Working Party for Schengen Matters (Acquis) to expand the scope for when a discreet and specific check can be issued beyond criminal investigation, to include police action on the prevention and detection of criminal offences (Presidency, 2017). The technical capabilities to search SIS are also being enhanced, in March 2018 SIS II became the second European database that integrated the Automated Fingerprint Identification System (AFIS). Previously, the fingerprint data of SIS was only searchable through name and date of birth, now AFIS has made the fingerprint database biometrically searchable. Since the integration, it can be witnessed that the police are increasing the number of fingerprints they upload in the system (The Sofia Globe, 2018).

Investment in databases and their functionalities has met opposition from oversight commission, human rights, and public interest groups. In the U.K “The Biometrics Commissioner has warned [the police] that many of the 20 million custody photographs currently stored on their systems are being held unlawfully and might need to be destroyed” (Loeb, 2018). Merging the three databases into the NLEDS, not only accumulates these existing concerns it also creates new ones related to proportionality, legitimacy, and ownership of data. While “proportionality will be a design feature of the system with permission-based access, with a full audit trail and a description of purpose of access. There is much work to do in terms of exact detail” (Surveillance Camera Commissioner, 2016: 23). One of these exact details is around specifying who will have access to what part of the database (Big Brother Watch, 2018: 13-14). In Germany and the U.K human rights groups argue that the criteria for being included in a database are often defined too broadly in terms of ‘terrorism’, ‘violent crime’ or ‘gang-affiliated’, sweeping up to many innocent citizens (Monroy, 2018A; Amnesty International, 2018). The concerns around the growing number of individuals who are entered into various databases should be understood in relation to the increase in data collected and shared about them. It can be observed that the integration of AFIS
into SIS resulted in the collection and storage of more fingerprints than before, demonstrating that once functionality has been added police will use it (The Sofia Globe, 2018). Expanding mandates for police to share data to pre-empt crime, as in the case of EPRIS, will most likely increase the amount of data sharing between European police departments (Jones, 2011). While investments are directed towards increasing the functionalities of Europe’s police databases, it is important to note that police struggle to delete ‘old’ data. A clear example of this was during the G20 summit in Hamburg, Germany when the accreditation of 32 journalists was revoked as a result of them being labeled as ‘left motivated violent offenders’ in a BKA database (Monroy et al, 2018). Legal challenges revealed that the decision to revoke their accreditation was based on ‘old’ data which was never deleted. According to the BKA, it is the responsibility of the authorities, who enter data on suspects into the system, to delete old records. In most cases, data entry is done by the German states, which fail to remove ‘old’ data (Fiedler, 2017).

3 Real Time Identification systems

Another trend in data-driven policing practices is investments into technology that allows for the real time identification of individuals and objects, ranging from Automated License Plate Readers (ANPR), Automated Facial Recognition (AFR), and voice identification systems. These technologies can be distinguished from more established forms of police identification, such as fingerprints or DNA identification, as it allows for data to be captured in the ‘wild’ without the individuals’ knowledge or consent, which can be compared to a database in real time (Surveillance Camera Commissioner, 2016; Gates, 2011). In the case of a ‘hit’, a positive identification of an individual or object against a database, this technology allows police to monitor and follow the ‘hit’ across the infrastructure, be it highways, cities, or telecommunication networks.

3.1 Automated License Plate Readers (ANPR)

ANPR is a widely adopted real time identification system in Western Europe, which aims to monitor the movements of suspects across cities and the country. Cameras on highways and roads capture images from the front and the back of a vehicle, collecting data on the license plate, the colour and model of the vehicle, and in some cases the driver or passengers (Politie-verhoor.nl, 2019). The ANPR system scans these images for license plates of passing cars which are cross-referenced with police databases, containing lists of license plate numbers which are associated suspects or persons of interest (Surveillance Camera Commissioner, 2016: 24; Politie, 2019; NPCC, 2016). The 2015/2016 U.K’s Surveillance Camera Commissioner report showed that at that time the U.K had “approximately 8,500 cameras in use capable of capturing 35 million and 40 million ‘reads’ a day and storing upwards of 30 billion ‘reads’ a year” (Surveillance Camera Commissioner, 2016: 23). It notes that the rapidly expanding ANPR system is creating a large-scale surveillance system that operates outside the realm of parliamentary oversight and without proper legal frameworks in place (Surveillance Camera Commissioner, 2016). In the Netherlands, the police have about 450 road cameras and 150 mobile units at their disposal for ANPR. A ‘hit’ in the system will allow the police to monitor the vehicle while travelling, run additional checks or take actions (Privacy First, 2018). In January 2019 a new law on ANPR went into effect, allowing the Dutch police to capture and store the license plate, location, time and picture of the vehicle through the ANPR system for up to 28 days, regardless if it was a ‘hit’ or not. (Privacy First, 2019; Politie, 2019).

Several German states are implementing ANPR systems and compare the license plates to police databases, but unlike the Dutch and the U.K police, the practice was to delete the ‘no hit’ immediately (Gregorová, 2019). In 2019 the German ANPR systems became contested, the German Constitutional Court ruled that mass registration of license plates infringes on the right to informational self-determination and limits individual freedoms. The ruling states that ANPR system can only be used within narrow scopes of the law, for example, to detect and prevent cross border crime or crime that has a certain severity. The German states Bavaria, Baden-Württemberg, and Hessen are required to temporarily discontinue their ANPR systems and change their laws.
to comply with the court decision. The court bound the ANPR system to specific spatio-temporal horizons (Hempel, 2019: Geuther, 2019). The Danish police have modelled its ANPR system after that of the U.K., number plates are checked against a pre-compiled ‘hot list’, specifically in the border areas. The ANPR data is stored for up to 30 days and the program aims to share it with tax authorities and other Danish authorities. FOI requests from the Danish newspaper Berlingske show the police intents to use this data for investigations and data analysis (EDRi, 2014).

3.2 Automatic Facial Recognition systems

Automated Facial Recognition is used to identify individuals at soccer matches, protests, festivals or in neighbourhoods (Ferris, 2018). Mobile or fixed cameras capture images of individuals in the ‘wild’, extract machine readable facial prints, and use these to cross reference them against the police database or follow persons of interest across the city (Gates, 2011). AFR and specifically Live Facial Recognition (LFR) is not widely adopted by police in Europe, but departments are experimenting with these technologies in different pilots. The London police are testing AFR for the purpose of public order policing, experimenting with facial recognition at the Nothing Hill Carnival and with mobile camera units on the city streets for the identification of persons of interest (Ferris, 2018; London Policing Ethics Panel, 2018: 10). For personal surveillance, the city of London proposes to update the “ring of steel” surveillance system that connects the ANPR and CCTV networks in the city with amongst other facial recognition (Ferris, 2018: 31; Professional Security, 2018). In Cardiff, the South Wales Police received £2.6m from the Police Transformation Fund to lead on the testing and deployment of AFR, and in some instances LFR. “South Wales Police has admitted it has used AFR technology to target petty criminals, such as ticket touts and pickpockets outside football matches, but they have also used it on peaceful protesters” (Liberty, 2018). In Germany, AFR was developed and deployed for public order policing during the G20 summit in Hamburg and continues to be used today. Initially, the AFR software was used for the retrospective identification of ‘violent’ protesters at the G20 summit. Images captured through surveillance cameras were used to identify people as objects that could be followed across collected images. In a later stage, some facial prints were cross-matched against the BKA INPOL police system. In Hamburg, AFR is seen as increasing the likelihood violent crimes are detected and acted upon (Monroy, 2018B). The Dutch police use the software CATCH to match the images of suspects captured on a security camera, ATM machine, social media or ‘observation-units’ to images in a database that hold a million images of convicted or arrested individuals. While currently this facial recognition software is not used in real-time, it has the potential to do so (Kist, 2018; Politie 2016).

3.3 Voice identification systems

While this research was not able to pinpoint specific police program which uses voice or speaker recognition software, several developments indicate that European police forces have shown an interest in and are testing speaker recognition systems to identify suspects on phone lines, VOIP, video and social media. Interpol and the Italian, German, Portuguese and U.K.’s Metropolitan Police force (Met) were involved in the Speaker Identification Integrated Project (SiiP), a technology developed by Verint Systems Ltd and co-financed by the EU’s funding program Horizon2020\(^5\). The SiiP database operate alongside Interpol’s existing fingerprint and face database, and allow police to match voice samples taken from YouTube, Facebook, phone calls, voice-over-Internet-protocol (VOIP) recordings, and recordings from CCTV cameras to voice recordings stored in a database (Kofman, 2018; Dumiak, 2018; Monroy, 2017). “SiiP will then search local and global audio databases using key identifiers such as gender, age, language, and accent. It will also search social media channels to find matches with individuals not yet known to police” (European Commission, 2018). In 2017 the London MET and Portugal police force tested the SiiP system in the

\(^5\)
Voice identification is not a new police technique, research into the use of speaker recognition by police and secret services published by Verint Systems indicated that the technology was used in one form or another in 69 countries (Morrison et al, 2016). Known applications have been the use of speaker recognition for forensic investigation, which has been operational in the Netherlands for over 15 years (Moret, 2009).

Scholars, technologists, human rights groups and oversight committees critique the use of real time identification systems by police. The large scale capturing of data and monitoring of individuals is considered to violate people’s fundamental human rights. The German Constitutional Court ruled that the premise of the ANPR systems, tracking cars across the road infrastructure without just cause, impairs on people’s freedom and right to informational self-determination (Geuther, 2019). In the U.K Big Brother Watch (2019: 13) found that there is no legal basis for the use of live facial recognition, and raises concerns that real time identification systems operate in the grey areas of the law. These systems capture data in the ‘wild’, monitoring everyone who passes a camera or speaker system in the public space, and in the case of facial and voice recognition requires police departments to test flawed technology live on citizens. The AFR systems used in by U.K police were found to perform abnormally poorly, with “on average, a staggering 95% of ‘matches’ wrongly identified innocent people” (Ferris, 2018: 13), who were subsequently stopped and asked to prove their identity. These findings of flawed facial recognition technology are in line with research showing inherent bias in facial recognition systems towards women and people of colour (Buolamwini et all, 2018; Ferris, 2018). The independent panel advising London City Hall on ethics of policing has in its report on the London Met’s facial recognition trial acknowledged the validity of these concerns but does not question the use of the technology in itself. It has recommended that the Met “should publish its view on the legality of using facial recognition technology before any further trials … and continue working closely with the relevant commissioners to ensure proper oversight of its use” (Mayor of London, 2018b).

4 Predictive policing

Predictive policing technologies use historical and real time data to predict when and where a crime is most likely to occur or who is most likely to engage in or become a victim of criminal activity (Brakel, 2016). European police forces are investing in the development and piloting of several predictive mapping and predictive identification systems that should allow them to more efficiently allocate resources to pre-emptively intervene and deter crime. This type of policing practice shifts from investigating a crime once it has occurred or as it happens, to relying on statistical probability to intervene and take action prior to a criminal act. These predictive models are based on the assumption that when the underlying social and economic conditions remain the same crime spreads as violence will incite other violence, or a perpetrator will likely commit a similar crime in the same area (Brayne et al, 2015).

4.1 Predictive mapping

Police forces in Europe are each experimenting with their own approach to predictive mapping, also known as hotspot policing. There is a wide range of products being developed, of which all are primarily used to predict the locations of high impact crime, i.e. robbery, theft, and burglary. There are two distinct models, near repeat and time-space. Where the first predominantly relies on police data (type of crime, location and time) to predict where high impact crime is most likely to happen in the near future, the later also includes variable like weather, holidays, events, and distance to highways (Hardyns et all, 2017; Ferguson, 2017). In Germany and Switzerland, several police forces have turned to the German build near repeat model PreCobs developed by the Institut für musterbasierte Prognosetechnik (IfmPt) (Knobloch, 2018; Hardyns et all, 2017; Brakel, 2016). PreCobs uses crime data to predict where crime, specifically burglary is most likely to take place. Where most predictive policing models are based on crime in urban areas, Bavaria is experimenting with the so called ‘far repeat approach’ to predict crime in rural areas (Seidensticker et
The police in the German state of Lower Saxony developed a similar predictive policing tool for domestic burglaries called PreMap (Seidensticker et al, 2018). KLBOperativ, which is the Hessen in-house built predictive policing model, is based on near-repeat but instead of solely relying on police crime data, socio-economic data is included in the analysis, which is then mapped on a mobile phone app that allows the patrolling police officer to do searches and apply filters (Seidensticker et al, 2018).

The police departments in North Rhine Westphalia (Skala), Berlin (KrimPo) and Amsterdam (CAS) have built a time-space predictive policing model, calculating the likelihood crime will occur on the basis of a wide range of variables. Skala is built on a strong theoretical framework and includes police data, infrastructural conditions, and demographic data in its modelling. The tool is currently applied to pre-empt burglaries and car thefts (Knobloch, 2018). KrimPo is the predictive policing software built in-house by the State Office of Criminal Investigations in Berlin occasionally with support or Microsoft, which uses police crime data in combination with socio-economic data to predict where domestic burglary is most likely to happen (Seidensticker et al, 2018; Knobloch, 2018). Several police forces in the Netherlands are turning to the Crime Anticipation System (CAS) after a successful pilot in Amsterdam in 2014. This predictive policing tool is based on the computation of crime data, socio-economic data of the neighbourhood, location specific characteristics like type of retail, distance to the highway or closest known offender (Vries, 2016; Drenth et al, 2017; William, 2014). In the U.K some police forces are implementing predictive mapping systems that are developed ‘in-house’, while others are bought from commercial vendors. In 2018 Kent police announced they would discontinue the £100,000 annual contract with Predpol, a commercial predictive mapping software, stating it would look into cheaper alternatives or the possibility to build a similar tool in-house. The London MET also tested and discontinued three different commercial predictive policing software programs, PredPol, Azevea, and Palantir, as the costs were too high, and are continuing with an in-house solution (Couchman, 2019; Beckford, 2018).

4.2 Predictive identification

Predictive identification aims at predicting who is most likely to become a potential offender or potential victim of crime. The most common practice of this type of technology in Europe is risk modelling, in which individuals are identified and ranked according to the likelihood they will engage in criminal or violent activity. In the U.K. there are three distinct predictive identification programs in London, Avon and Somerset and the West Midlands (Couchman, 2019; Amnesty International, 2018). The London MET developed the Gang Matrix to identify potential gang members and score them according to the risk they pose to society. Research by Amnesty International (2018) and Scott (2018) revealed the discriminatory nature of this predictive identification program, in which the majority of individuals were young black men. The MET was ordered to radically reform the matrix within a year by the Mayor of London, and are currently working on a new program called the ‘Concern Hub’ (Mayor of London, 2018a: Dodd, 2018; Crisp, 2019). The West Midlands police are leading the National Analytic Solution project, funded by the Home Office, in collaboration with 8 other police forces. “The system called the National Data Analytics Solution (NDAS), uses a combination of AI and statistics to try to assess the risk of someone committing or becoming a victim of gun- or knife crime, as well as the likelihood of someone falling victim to modern slavery” (Baraniuk, 2018). In analysing historical police records the software found that data on an individual’s criminal history and the number of crimes committed within their social network were the strongest indicators to predict future crime (Baraniuk, 2018). In Avon and Somerset the police program Qlik aims to predict two things, the likelihood someone will commit a certain crime and offer a beat officer the ability to predict the likelihood someone will become violent during a stop (Dencik et all, 2018).

In Amsterdam in the Netherlands, there are two distinct risk modelling programs, the top 600 and the top 400 (ProKid 12). The top 600 focuses on repeat offenders who in the past 5 years have committed more than 3 crimes, and is more focused on risk assessment, whereas the top 400 focuses on repeat offenders who in the past 5 years have committed more than 10 crimes and is more focused on risk management. In the UK, Kent police announced they would discontinue the £100,000 annual contract with Predpol, a commercial predictive mapping software, stating it would look into cheaper alternatives or the possibility to build a similar tool in-house. The London MET also tested and discontinued three different commercial predictive policing software programs, PredPol, Azevea, and Palantir, as the costs were too high, and are continuing with an in-house solution (Couchman, 2019; Beckford, 2018).
years have been arrested for a high impact crime, and have had a criminal conviction at least twice for individuals under 21 years old, and three times for individuals over 21 years old (Openbaar Ministerie, 2019). The top 400 builds on ProKid 12, developed by the police department of Gelderland-Midden to score the likelihood of children under the age of 12 to become potential criminals. Once identified the idea is to holistically intervene and deter repeat offenders and potential criminals from engaging in future criminal activity (Abraham et al, 2011). In Germany the BKA developed the RADAR-iTE system to create a top 500 list, ranking potential terrorists according to the dangers these pose to society (BKA, 2018).

The use of predictive policing technologies has raised many concerns. Scholars in the USA have demonstrated how historically biased police data in predictive policing programs is perpetuating the over-policing of African American neighbourhoods (Lum et al, 2016). In the U.K Liberty found a similar negative feedback loop, people from "back, Asian and minority ethnic (BAME) communities are disproportionately more likely to be arrested, leading the program to assume, wrongly that the area in which they live or spend time are the areas where there is more crime” (Couchman, 2019: 4). The Gang Matrix is a clear example of how biases systems, when targeting or being piloted in specific neighbourhoods and not others increases stigmatization of affected communities. In the Gang Matrix, the names of ‘gang nominals’ where shared with the education, health care and housing providers, and the Driver and Vehicle Licensing Authority (DVLA). The DVLA took actions and send letters to ‘gang nominals’ in the London borough of Haringey that they are considered unfit to drive. Some letter recipients were required to return their license immediately, while others had to provide information about their drug history and some had to take drug tests (Amnesty International, 2019; Scott, 2018: 26-27; Couchman, 2019). These real-life consequences become even more concerning as the Gang Matrix did not clearly distinguish between perpetrators and victims of crime (Mayor of London, 2018a).

5 Analysis of heterogeneous databases

The last trend discussed in this paper is the analysis of heterogeneous databases. The increase of available data poses a challenge for those who try to make sense of it. Police are investing in technologies that enable them to effectively search through large quantities of heterogeneous data, from multiple sources, formats and in different databases. Police forces across Europe are mainly turning to Palantir\textsuperscript{7}, a big data analytics company from the US with a track record of contracting for police and secret services across the world, for the integration and searching of heterogeneous databases, and social graphing of suspects. The French secret service (Samama, 2018), the Danish national police (Edri, 2017), two German Bundeslander (Brühl, 2018) and a number of Dutch local police forces (Hengst-Bruggeling, 2010) have purchased data analytics tools from Palantir. The French Directorate General of Internal Security (DGSI) has bought the services of Palantir for €10 million since 2016 to fight terrorism. This contract was justified by the absence of other French or European solutions (Samama, 2018). In a public tender the Danish national police looked for a computational supplier to develop two Danish intelligence systems (called PET-INTEL and POL-INTEL), Palantir was selected in 2016 among three companies. When operational the system should enable the Danish police to access existing police and intelligence databases, exchange information with Europol, engage in open source collection of new information, as well as algorithms for pattern recognition, hotspot analysis, and social network analysis. Open source data is broadly defined and can include information from the internet and public spaces like ANPR data, but also information bought from commercial vendors (EDRi, 2017). In Hessen, Germany, police use an adapted version of Palantir’s Predictive Policing tool Gotham called Hessendata. The software is licensed to the German state for 0.01 Euro, which substantially spend 600.000 to training police officers in its use (Borchers, 2018). Hessendata builds social graphs around potential terrorist suspects, using seven data sources including police databases and Facebook profiles to create profiles and social graphs. The police have praised the tool for its interface, its ability to

\textsuperscript{7}https://www.palantir.com/
process data and find patterns that would have otherwise not be understood (Brühl, 2018).

European human rights groups, oversight committees, and scholars have been less vocal on the use of Palantir technologies by European police forces then they have been on the creation of databases, real time identification, and predictive policing systems. Concerns have been raised on the use of an American company for crunching French police data (Samama, 2018), how the use of Palantir software might not be in-line with national and European data protection regulation (EDRI, 2017; Technology World, 2018) or more fundamentally how the use of Hessendata by police to track suspects blurs the separation between the German police and secret service (Kurz, 2018). While there might be other concerns on the use of this technology, European police have not been open about the implementation of Palantir software.

6 Conclusion

This report covers the data-driven policing implementations in different countries across Europe, most prominently in Western and Northern parts of Europe. It outlines four key trends: investment in databases, real-time identification of individuals and objects, predictive policing, and the analysis of heterogeneous datasets, for which police departments are increasingly experimenting with integrating data and technologies into their policing practices. While it should be acknowledged that the adoption and integration of data-driven policing technologies are still in its infancy, pilot programs are mushrooming across Europe. The EU Horizon2020 program, the U.K Home Office Police Transformation Fund, and national police budgets support the development of new data-driven policing technologies. The development process is often a combination of in-house or public-private partnership collaborations, which is complemented with the purchasing of off-the-shelf commercial technology. Known computational vendors are IBM, Accenture, Microsoft, Palantir, PredPol and more niche real time identification companies, like Verint Systems LTD for voice recognition, and NEC for facial recognition. It is worth noting that the trend in the digital economy leans towards the establishment of monopolies over home grown or distributive systems, where niche computation vendors dominate a specific technology or section of the data-driven policing market (Srnicek, 2017). A clear example is the reliance on Palantir for analysing heterogeneous databases, in which the French secret service turns to this American company for lack of a French or EU alternatives along with several other police forces in Europe, and the facial recognition market which is also dominated by a small group of computational suppliers (Gates, 2011). In addition, there are signs which indicate how different technologies will in the near future fold into one, i.e. the proposal of the city of London to integrate AFR into the ‘ring of steel’ and the possibility to use real time identification as input for predictive policing systems. This will enable the police, with the support of a small group of computational suppliers, to collect data in real time, analyse heterogeneous databases and support the police in their investigation or predict crime.

Concerns are raised on the fundamental premise on which data-driven technologies are built, where police can monitor everyone in a public space and take action on the basis of this data. For specific programs, human rights groups and oversight committees question the proportionality and scope of these technologies, the shortcomings in responsible data practices, the lack of clarity on the criteria and processes for including individuals in databases and the failure to delete ‘old’ data. There is a clear need to explore how these data-driven policing programs are currently being applied, where, on which crime and question the potential negative impact on poorer neighbourhoods and affected communities. For this, it is crucial to question these pilots in light of historical, social, political and economic context of the particular country and Europe. Future work under the ERC-funded project ‘Data Justice: Understanding datafication in relation to social justice’ (DATAJUSTICE) starting grant (2018-2023) will look at how and why police are integrating certain technologies, who is financing it and for what purpose, and how the use of these technologies impact the ability of impacted communities to make justice claims.
7 References


Kreickenbaum, M. (2017). German federal police illegally collect data to blacklist journalists and activists


Morrison, Geoffrey Stewart ; Sahito, Farhan Hyder ; Jardine, Gaëlle ; Djokic, Djordje ; Clavet, Sophie ; Berghs, Sabine ; Goemans Dorny, Caroline (2016). INTERPOL survey of the use of speaker identification by law enforcement agencies. Forensic Science International, June 2016, Vol.263, pp.92-100


April 2016 https://www.npcc.police.uk/documents/NPCC%20NUG%20Terms%20of%20Reference%20April%202016.pdf


Seidensticker, K., Bode, F. & Stoffel, F. (2018). Predictive Policing in Germany. https://kops.unicolnanz.de/bitstream/handle/123456789/43114/Seidensticker_2-14svvox1tik0z06.pdf?sequenc e=5&isAllowed=y


