WRR

THE NETHERLANDS SCIENTIFIC COUNCIL FOR GOVERNMENT POLICY

WRR-Policy Brief 6

Big Data and Security Policies: Serving Security, Protecting Freedom

2017

Big Data analytics in national security, law enforcement and the fight against fraud can reap great benefits for states, citizens and society but require extra safeguards to protect citizens' fundamental rights. This requires a crucial shift from regulating Big Data *collection* to regulating the phases of *analysis* and *use*.

Dennis Broeders, Erik Schrijvers & Ernst Hirsch Ballin

SUMMARY

In order to benefit from the use of Big Data analytics in the field of security, a framework has to be developed that adds new layers of protection for fundamental rights and safeguards against erroneous use. Additional regulation is needed at the levels of analysis and use, and the oversight regime is in need of strengthening.

- At the level of analysis – the algorithmic heart of Big Data processes – a duty of care should be introduced that is part of an internal audit and external review procedure. Big Data projects should also be subject to a sunset clause.
- At the level of use, profiles and (semi-)automated decision-making should be regulated more tightly. Moreover, the responsibility of the data processing party for accuracy of analysis – and decisions taken on its basis – should be anchored in legislation.
- The general and security-specific oversight functions should be strengthened in terms of technological expertise, access and resources. The possibilities for judicial review should be expanded to stimulate the development of case law.

The Netherlands Scientific Council for Government Policy (WRR) provides independent advice to the Dutch government and parliament on strategic issues that are likely to have important political and societal consequences on the longer term. The Council's science-based advice can open up new perspectives and directions, change problem definitions, set new policy goals, investigate new resources for problem-solving, and support social cooperation and cohesion.

WRR-Policy Briefs reflect from an academic perspective on important societal and policy issues, may provide knowledge-based input for the policy agenda, and can suggest policy options.

This policy brief is based on the Dutch report *Big Data in een vrije en veilige samenleving* (Big Data in a Free and Secure Society), presented by WRR to Ard van der Steur, the Dutch Minister for Security and Justice, on 28 April 2016.

The original report was written by a project team chaired by Ernst Hirsch Ballin, member of the Council, and further consisting of Dennis Broeders (project coordinator), Erik Schrijvers, Bart van der Sloot, Rosamunde van Brakel and Josta de Hoog. Sacha van Schendel worked on the project as an intern.

January 2017, no. 6
Publisher: WRR
Design cover: cimon communicatie, The Hague
Layout: Textcetera, The Hague
ISSN: 2352-1392

All publications are available online at www.wrr.nl/en.

Please cite this publication as:
Broeders, D., E. Schrijvers & E. Hirsch Ballin (2017), *Big Data and Security Policies: Serving Security, Protecting Freedom*, WRR-Policy Brief no. 6, The Hague: WRR.

CONTENTS

1 THE PROMISE AND PERILS OF BIG DATA IN SECURITY POLICIES

Big Data is a catchword that promises radical change. Expectations are high when it comes to increasing sales, targeting advertising, optimising processes and generating unforeseen, unexpected and unprecedented insights. According to some, Big Data will revolutionise the way we live, work and think.[1] Governments are keen to make sure that the benefits of these new technologies will be integrated into public policies as well. In the policy domain of security – broadly interpreted as ranging from national security, via law enforcement to the combat and prevention of fraud – the number of programmes that involve large-scale data collection, linking and analyses are on the rise. Most of those are not on the scale of Big Data 'proper' yet, but the trends indicate that this may change in the coming years.

The opportunities and benefits (both potential and realised) of applying Big Data analytics in the security domain are many, including greater operational efficiency and speed, more precise risk analyses and the discovery of unexpected correlations, all of which feed into risk profiles, better targeted inspections and more efficient use of scarce resources. Big Data analyses help in reconstructing past events (immediately after an attack, for example) and are useful in monitoring developments in real time. This is of great value, for example, in traffic management, organising information and aid following a disaster or for crowd control at events. Most of all, however, there is the promise that Big Data analytics will deliver insights into the future and may provide the foundation for effective preventive policies. However, these potential gains in security will come at a price in terms of individual and collective freedoms and fundamental rights. Just as the state is responsible for the security of its citizens, it is also – and equally – tasked to protect their personal freedom.

This policy brief aims to lay the groundwork for a regulatory framework for the use of Big Data in security policies that respects and protects fundamental rights. Most crucially, this requires a shift from regulating data *collection* to regulating the *analysis* and *use* of Big Data.

2 A WORKING DEFINITION OF BIG DATA

The many Vs of Big Data

Big Data is still very much a moving target. Technological developments and new applications continue to feed into the debate about what defines Big Data and sets it apart from earlier forms of data analysis. There is no real consensus regarding its key characteristics, although most definitions of Big Data refer to the ubiquitous three Vs.[2]

[1] Mayer Schönberger and Cukier (2013); see also Greengard (2015).

[2] Laney (2001).

The first of these three stands for *Volume* (the use of large amounts of data), the second V is for *Variety* (the use of diverse data sources that are stored in diverse structures or even in an unstructured way) and the third stands for *Velocity*, or the speed of data processing (data is often analysed in real time). Over time, a number of authors have added additional Vs to this threesome, such as Veracity[3], Variability[4], Value[5] and Virtual[6]. The various definitions do not amount to a broad consensus on the issue but do demarcate the corners of Big Data as a field of study.

From definitions to a frame of reference for Big Data in public administration

In this policy brief, we will not add our own definition, but rather collect a number of important elements from the definitions of others to construct a frame of reference for the use of Big Data in the context of public administration, especially in the security domain.[7] This frame of reference is grouped around three main aspects of Big Data processes: data (collection), analysis (techniques) and the use of Big Data results (see Table 2.1).

Table 2.1 Frame of reference for Big Data

Data	• Amount of data: large amounts of data are involved. • Organisation of data: Big Data analytics can deal with both structured and unstructured data. • Variety of data: there is a combination of various data sources and data formats (text, sound, video).
Analysis	• Method of analysis: the analysis is *data-driven*, so patterns are sought in the data without pre-established hypotheses. It favours correlations over causality. • Orientation of the analysis: although Big Data analyses also give information about the past (retrospective analyses), it is particularly the analyses of the present (*real-time analyses / nowcasting*) and the future (*predictive analyses / forecasting*) that draw attention.
Use	• Decompartmentalisation of domains: data from one domain is used for decisions in another domain. • *Actionable knowledge*: conclusions at aggregated level can be applied to decisions at group or individual level (person or object).

Big Data is seen here as the interplay between these characteristics rather than as a well-defined and definable object. This leaves room to discuss the use of data analysis in public policy making that includes some of these characteristics to a degree but does not tick all the boxes. Most current policy programmes analysed in the Netherlands do not cover the full range of this frame of reference. It is often the potential to grow into full Big Data systems that makes it important to scrutinise policy initiatives now.

[3] IBM (2015); Klous (2016).
[4] Hopkins and Evelson (2011); TechAmerica Foundation (2012).
[5] Dijcks (2012); Dumbill (2013).
[6] Zikopoulos and Eaton (2011; Akerkar et al. (2015).
[7] See WRR (2016: 33-34).

3 BIG DATA, SECURITY AND FREEDOM

Security, freedom and the need for distance

The use of Big Data analytics in security policies influences both freedom and security at the individual and societal level and, therefore, touches upon the very foundations of the constitutional state. Both freedom and security are rooted in fundamental rights.

Freedom presupposes distance – a certain amount of social space between the individual and others, including supervising bodies. In the history of the modern state, distance in relation to institutions that want to observe and direct our behaviour – such as the government – has brought about an increase in personal freedom. For the government, it is only citizen behaviour in relation to the law that should count. In a free society, citizens are not judged according to *who* they are: their intentions and emotional lives have no relevance for the law. This freedom is an important dimension of their personal security.

At the same time, it is the government's essential duty to protect its citizens and increase security, precisely for the purpose of ensuring that they can live in freedom. Therefore, government will have to ensure societal and individual freedom by gathering information, being alert to dangers and working to eliminate threats to security while maintaining sufficient distance from the personal lives of its citizens. This distance distinguishes a constitutional state from a totalitarian one and determines the degree of personal security and the security of society. The government's security policy must be structured in a way that serves to protect both personal and social freedom, for if it fails to do so, this policy will undermine exactly that which it has set out to protect.

Big Data as the negation of distance

Big Data, however, constitutes an assault on the protective function of distance. The amount of information that is now available or can be accessed for surveillance, investigation and prosecution has risen sharply. Combined with cheaper and more flexible forms of data storage and computers that can carry out ever more complicated data processing tasks, this results in government bodies increasingly encroaching on the lives of citizens. This interferes with the protective function of distance and undermines people's freedom. This effect is reinforced if the knowledge that has been built up is incorrect or merely expresses a statistical probability, whereas the individual involved could be an exception. These potential effects increase if an individual is targeted on the basis of large-scale statistical data analysis as in Big Data processes. The use of Big Data in public administration, therefore, has to navigate a course that respects the protective function of distance and compensates for infringements through adequate accountability mechanisms.

4 BIG DATA AND SECURITY PRACTICE

It is not so easy to analyse how and to what extent Big Data applications already manifest themselves in the field of security. This is due to the secrecy that often shrouds security policy operations as well as the experimental nature of some applications and the understandable – though regrettable – reluctance to debate those in public. There is a well-founded fear among security agencies of being framed in a big brother context. There is some insight into the use of Big Data analytics in the security domain when it comes to combating fraud[8], the use of data analytics for the purpose of 'predictive policing'[9] and the use of data analytics by the intelligence and security services, notably through the Snowden revelations about the NSA.[10] Different government agencies, however, working in a broadly defined security field, vary widely in their legal authority to collect and process data as well as in their technical ability to deal with large-scale data analyses. Of all government agencies, for example, the Dutch tax authority has the most extensive database and operates under a legal framework that allows it to collect and combine data with many other sources.[11]

In the fight against fraud, data mining is increasingly taking a central place, both in the Netherlands[12] and abroad.[13] Using data to predict where crime is likely to be committed or even *who* is likely to commit a crime, so-called predictive policing, is on the rise in the USA and the UK [14] and is also making its way into Dutch policing practice.[15] Smart city technology is up and coming and will not only improve services to city dwellers but, with new crowd control and surveillance possibilities, will also have implications for security policy.[16] Profiling and data mining are also on the rise in the implementation of border and immigration policies. Sorting out the risky from the trusted travellers[17] and creating more 'situational awareness' for the border authorities in the Mediterranean[18], for example, has increasingly become a 'datified' activity.

On the brink of Big Data analysis

In spite of large-scale database linking and mining and more sophisticated methods of analysis, many security organisations seem to be on the brink of working with Big Data rather than in the thick of it. Their data programmes incorporate some of the characteristics of Big Data – as outlined in Table 2.1 – but not the full set. Most importantly, they often still work on the basis of investigating a known suspect, instead

[8] Olsthoorn (2016).
[9] See, for example, Rienks (2014); Van Brakel (2016). For an overview and critical analysis, see Ferguson (2017).
[10] Greenwald (2014); see also https://theintercept.com.
[11] WRR (2016).
[12] Olsthoorn (2016); WRR (2016: 52-58).
[13] O'Neil (2016).
[14] Ferguson (2017).
[15] Van Brakel (2016); Rienks (2014).
[16] Galdon Clavell (2013), (2016); Kitchin (2014a).
[17] Broeders and Hampshire (2013); Jeandesbosz (2016).
[18] Broeders and Dijstelbloem (2016).

of switching to data-driven analyses, in which data mining will tease the suspects out of the dataset – in various degrees of probability – on the basis of profiles or even mere correlations. Some organisations, such as iCOV, a Dutch anti-fraud agency, find themselves exactly on that brink.

iCOV

The Dutch anti-fraud agency iCOV (infobox Criminal and Undeclared Assets) gathers data from government agencies such as the National Police, public prosecuting authorities, the tax authority and several other law enforcement and anti-fraud agencies. iCOV produces reports on the assets and incomes of suspected individuals or groups and maps the financial networks of people or organisations. It receives its data from member organisations and stores it in a safe data warehouse. The members can request the investigation of a person or company under suspicion. The requesting partner's legal competence determines what information from the database will be included in the analysis. So far, the results have been positive. The amount of time saved in comparison with normal financial investigations is substantial. iCOV is now at the point where the expertise it has built up in previous years can be translated into building profiles that could be used to data mine its extensive databases to unearth potential fraudsters. This would amount to a shift in the direction of data-driven rather than suspect-driven analysis. The expectation is that this Big Data type of analysis will allow iCOV to track down even more fraudsters. Within iCOV, however, there is uncertainty about what is and what is not permissible with regard to data processing under current legislation.[19]

Even though the 2016 security domain is not dominated by Big Data, it stands to reason that the growth of available data and analytical possibilities will accelerate current pilots, practices and experiments and give Big Data analytics a more prominent place in security policies in the near future. If we extrapolate current developments, we can expect to see far-reaching effects on the collection, analysis and use of data in the field of security. Some of these effects can already be observed.

[19] Olsthoorn (2016: 177-200); WRR (2016: 54-56).

5 LOOKING INTO THE FUTURE: TRENDS IN BIG DATA AND SECURITY

Even though the often proclaimed Big Data revolution[20] is taking time to arrive, there are some major trends that point in the direction of a Big Data future.

1. Although the growth of available data over the last two centuries has been substantial, this is nothing compared with the current explosion in data size and variety.[21] Such data are increasingly the product of three processes: data collection may be 'directed' (the intentional capture of data on people or objects), 'automated' (data that are created as an inherent feature of a device or a system), or 'volunteered' (data that are created by voluntary use of systems, devices and platforms).[22] The amount of data from the last two categories in particular has grown exponentially with the advent of smart devices, social media and digital transactions. Most data are now 'born-digital' data.[23] A 'datafication' of everyday activities can be observed, where data is gathered virtually unnoticed, outside the control and without any meaningful permission – or even the awareness – of the individual.[24] The Internet of Things will propel this trend of datafication even further.[25] These developments will collectively add up to a qualitative leap in data collection.

2. Another development is taking place in the field of data analysis. New methods are emerging under the auspices of well-established techniques and algorithms, for example self-teaching algorithms and machine learning. The increase in the amount of data is already yielding better analysis. This is sometimes called 'the unreasonable effectiveness of data': moderately effective algorithms produce better results from very large amounts of data than better algorithms do from smaller amounts of data.[26] A key characteristic of Big Data is data-driven analysis, which is very different from the traditional statistical method[27]: the aim of data-driven analysis is not to test hypotheses but to find interesting and unexpected correlations and patterns, which may prove to be relevant for commercial purposes or for public goals such as the provision of services and security. Methods such as self-learning algorithms,

[20] Mayer-Schönberger and Cukier (2013).
[21] Kitchin (2014b: chapter 4). Greengard (2015: chapter 2).
[22] Kitchin (2014b: 87-98).
[23] Kitchin (2014b); see also Jerven (2013) for a global south context.
[24] PCAST (2014); Zuiderveen Borgesius (2015); Schneier (2015).
[25] Greengard (2015).
[26] Halevy, Norveg and Pereira (2009). This also underscores that, in data analysis, there is not much point in talking separately about the data and the method of analysis (algorithm), or as Gillespie (2014) observes: 'Algorithms are inert, meaningless machines until paired with databases upon which to function.'
[27] The scientific method formulates a hypothesis about the causality of a certain problem (A causes B) and tests this hypothesis with a data set. Big data analysis focuses on uncovering correlations (A and B correlate) without an a priori judgment on the causality between A and B. Anderson (2008) predicted that Big Data would bring about the 'end of theory' as Big Data correlations would generate accurate pictures of social reality. Others, such as Gillespie (2014) and Gitelman (2013), have underscored the limits and dangers of a theory-free interpretation of correlations.

machine learning and the 'production of correlations' are considered potential game changers, particularly if they become the standard for many future applications.

3. A third development is the increased use of predictive data analytics. Big Data can be used for historical analyses, but its real promise lies in real-time and predictive use. The growing availability of real-time data facilitates a growing number of real-time analyses. The idea of predicting the future – or, more accurately, predicting a possible future with a certain degree of probability – is the underlying rationale for many commercial and public programmes. Predictive analyses can take different forms: they can be used to help people make the right choice (consequential), identify our preferences (preferential) or restrict options (pre-emptive).[28] In the field of security, Big Data analyses are mainly used for pre-emptive surveillance and investigation and rarely, up to now, to make sure people are better equipped to deal with possible risks and threats, which would be a consequential approach.

The nature and origins of data that are available for security purposes, therefore, are changing. Public and private data are getting mixed. Relatively hard data (financial data and all kinds of registries) can be linked to softer, more social data. The wealth of data also renders the difference between personal and non-personal data potentially meaningless as it is now relatively easy to 'construct' a person on the basis of a limited set of data points that do not directly reference a person.[29] This means there is a limit to anonymisation and pseudonymisation methods and, more importantly, that the legal difference between different types of (personal) data and the level of protection they are awarded is being hollowed out.

Private data collections are already starting to play a bigger role in security analyses, supplementing data from government sources. The police are analysing social media to interact with the public and to gather intelligence (SOCMINT).[30] The Dutch tax authorities are using private data such as parking and transport details from private organisations.[31] Security and intelligence agencies have far-reaching authority to gather, share and commandeer data. Data exchange and linking are taking off. After all, it is often not the data itself that is valuable, but data linking, above all the linking of large amounts of data. Data collection and exchange for security purposes will be undergoing significant changes in the coming years.

In the Netherlands, a growing number of government agencies will wish to join existing and yet-to-be-created partnerships and cooperation agreements that exchange data and have it analysed.[32] This may involve data exchanges with private parties,

[28] See Kerr and Earle (2013) for this distinction.

[29] See, for example, the 'Unique in the crowd' research project (De Montjoye et al. 2013).

[30] See Bartlett et al. (2013); Omand et al. (2012).

[31] See https://decorrespondent.nl/1766/vergeet-de-politiestaat-welkom-in-de-belastingstaat/54315096-f35e98af.

[32] See WRR (2016: chapter 3); Olsthoorn (2016).

commandeering and requesting data, as well as purchasing data on the private market.[33] As ever larger and more diverse databases are being used, Big Data on people who are not under any suspicion are increasingly being collected and analysed.

Security organisations will increasingly make use of information-driven methods to inform and implement policies. This shift is fuelled by growing public and political concerns about security and new technological possibilities, austerity measures and the desire to work more efficiently.[34] Moreover, Big Data analytics will likely be used in the more lightly regulated parts of security policies, such as surveillance, public order and preliminary investigations, to guide the use of scarce resources and achieve more targeted and efficient checks and investigations. The legal framework covering such activities, however, is underdeveloped compared with criminal law, where a reasonable suspicion of guilt is a precondition for processing data and evidence has to stand up in a court of law.

6 BENEFITS OF BIG DATA

There are many benefits (potential and realised) of applying Big Data analytics in the security domain. Governments have traditionally gathered and owned a great deal of personal data, which can now be used for Big Data analyses. On top of this, government agencies working in the security domain are authorised to request data from third parties, provided that this falls within their remit. They have many opportunities, therefore, to work with Big Data, and for good reason, because Big Data can make a positive contribution to the field of security. The availability of a lot more data and the refinement of analysis techniques obviously offer opportunities for improving security policy, provided that the use of data and analysis does not itself become a security risk.

Operational efficiency

Big Data can contribute towards greater operational efficiency. There are profits to be made, particularly in organisations that are active in collecting and analysing data and information.[35] Analyses that used to take days, weeks or months can be completed in a few minutes, hours or days with the help of Big Data analysis techniques.[36] Facts can also be brought to light that would otherwise have remained needles in haystacks, simply because of the wealth of historical information.[37]

[33] In the Netherlands, the government is currently working on a general framework law for data exchange in the domain of anti-fraud policies. See Werkgroep Verkenning kaderwet gegevensuitwisseling (Working group to explore 2014 framework law on data exchange) (2014).

[34] Galdon Clavell (2016); Prins et al. (2011); Prins, Broeders and Griffioen (2012).

[35] OECD (2014: 19).

[36] See, for example, the results that the Dutch anti-fraud agency iCOV has booked (see text box on page 9).

[37] Schneier (2015: 35-45).

More precise risk analyses

Big Data makes it possible to carry out more precise risk analyses due to the larger size and greater diversity of the databases used. Methods of Big Data analysis also focus on 'discovering' unexpected connections, which can be worked into risk profiles. This may result in better targeted inspections and more efficient use of scarce resources such as ambulances and police surveillance.

Reconstructions for criminal investigations

Big Data analyses can help in reconstructing past events (immediately after attacks, for example) and can aid in criminal investigations. Uncovering unexpected connections can be of great use in various criminal cases, especially those that take place in a data-rich environment and that have specific and repeated patterns.

Real time analysis and crowd control

They can also be useful in following ongoing developments in real time. This can be of great value, for example, following a disaster or for crowd control at events, when it is important to get a clear picture of the situation on the ground and to do so in real time, so that services can offer help or intervene in dangerous situations.

Predictions

Big Data analyses can even be used to make predictions. In the field of security, there are high expectations when it comes to predicting the time and place of an increased crime risk and even of identifying future perpetrators and victims.[38] Such knowledge allows preventive action to be taken and to warn individuals and organisations about potential risks. Furthermore, predictions can increase the chances of apprehending criminals by providing more insight into their behaviour. Whether such methods can be applied, however, depends on the availability of sufficient information about events in combination with clear and repeated patterns of (criminal) behaviour and threats in sufficiently large numbers. Only then will predictive analyses be of use in such events.[39]

7 LIMITATIONS OF BIG DATA

Despite some claims to the contrary, Big Data is not a miracle cure. Big Data solutions are not equally applicable or appropriate to all security problems and, like all instruments, they have not only strengths but also inherent shortcomings.

[38] See, for example, Perry et al. (2013), Willems and Doeleman (2014) and Ferguson (2017) on predictive policing. See, for example, Bennett Moses and Chan (2014) and Harcourt (2007) on the use of predictive data analyses in the decision of judges and parole boards on whether or not to grant parole to detainees. For an overall view of the use of new technologies by police agencies worldwide and their experiences, obstacles and results, see Custers and Vergouw (2015).

[39] Ratcliffe (2010); Schneier (2015); O'Neil (2016).

Lacking the right data

The *right* data is not always available – even in a digital world. Sometimes the data are simply not there, sometimes there are problems with retention periods and sometimes different data platforms prove not to be interoperable. The quality of data is not a given either, as data can be outdated, corrupted, biased or even manipulated. These weaknesses can work their way into the data sets used and undermine efforts to enhance security through Big Data analyses.

The wrong tool for the job

Pattern recognition lies at the heart of Big Data, and not all threats, security issues and types of crime show patterns that can be analysed in a meaningful way. Data mining and profiling are an ineffective method of preventing terrorist attacks in the sense of looking for the needle (the lone wolf terrorist) in the haystack (bulk data) to prevent an attack before it happens. Pattern recognition works best for offences that show a fixed and repeated pattern. Because every terrorist attack is unique and the number of attacks is very low, it is virtually impossible to make a good profile: the percentage of errors – particularly false positives – will be far too high under these circumstances. Data-mining is simply the wrong tool for the job here.[40] Other offences – financial fraud, for example – may be better suited because there are more repeating patterns and methods of operation and far more case materials for profile building. The cost of false alarms, moreover, is often low in these cases, contrary to the lengthy investigations required by data-based terrorism-detection systems.

Working with probabilities and margins of error

By definition, Big Data analyses are based on historical data and data patterns, which can only offer a partial and probabilistic picture of the future. This means that those working with the results of those analyses should treat them as indications of possible outcomes rather than as straightforward results. One must be sensitive to the risk that people who have improved their lives continue to be classified as belonging to particular high-risk groups because they were registered as such at one point. In the field of security, these kinds of limitations should be given serious attention. If correlation is taken to be causality, and probabilities are treated as certainties, this may easily cause firm conclusions to be drawn, particularly in the area of surveillance and preliminary investigations, which are not bound by the rules of evidence guiding a court case. A forced entry by the police is of an entirely different order from an erroneous book recommendation by Amazon.com. This is why the outcomes of Big Data analyses in criminal law can never be more than one – potentially important – aid to investigation, and they can never push aside strict evidence requirements.

Any analysis based on statistical probabilities also produces both false positives and false negatives. False positive results criminalise innocent people, and false negative results allow security risks to continue unnoticed. A reduction in the number of false negatives usually means that the number of false positives increases, and the opposite is

[40] Schneier (2015).

also true. In short: results have to be weighed up carefully, both *within* Big Data analyses and when choosing *between* the use of Big Data solutions and other means.

8 RISKS OF BIG DATA

The application of Big Data in the security domain also comes with a number of risks that, if not properly addressed, can outweigh the benefits and may erode public support for Big Data solutions. Some of these risks may result from not addressing some of the limitations outlined above, and others are the result of policy choices and overreach of Big Data methods.

Bias, discrimination and data determinism

Big Data analyses may reinforce social stratification by reproducing and reinforcing the bias that is present in every dataset: data are often not 'given' but 'taken' or are extracted through observations, computations, experiments and record keeping.[41] As such, data are 'inherently partial, selective and representative', and the criteria used in their capture can distort the results of data analyses.[42] If uncorrected, the bias that characterises every dataset to a greater or lesser extent may, in time, translate into discrimination and unfair treatment of particular groups in society.[43] When used on a large scale, the results of Big Data analyses may well feed on each other, magnifying social and economic inequalities.[44] In the most extreme case, Big Data methods may result in data determinism, which means that individuals are judged on the basis of probabilistic knowledge (correlations and inferences) of what they might do, rather than what they actually have done. This is at odds with the presumption of innocence that is a cornerstone of criminal law.

Damage to individual and collective privacy

Big Data is at odds with individual privacy rights as it requires the collection and analysis of data on large numbers of people who are not in any way implicated as suspects. In addition to the damage to individual privacy rights, it can also affect privacy as a collective expression of freedom. Individual privacy rights are only legally triggered by the principle of *individual harm*, which is not something that often happens in the case of Big Data. The fact that your data are part of a massive data analysis often fails to meet the threshold of individual harm.[45] Privacy as an expression of collective harm, or conceived as damage to the fundamental right itself due to a large number of individual privacy violations, is hardly recognised by law but seems a more fitting risk in the current age of Big Data practices.[46] If we take risk to be defined as

[41] Borgman (2007) in Kitchin (2014b:2); see also Gitelman (2013).

[42] Kitchin (2014b: 3).

[43] Zarsky (2016: 126-127). Gandy (2009).

[44] O'Neil (2016).

[45] Take, for example, the Snowden revelations about mass surveillance. These were considered to be highly disturbing by many but barely amounted to individual harm for any individuals.

[46] See Van der Sloot (2016); Taylor, Van der Sloot and Floridi (2017).

‘impact x probability’, then incursions into the right to privacy have a small impact but a very high probability – as they happen every day to virtually all people. The risk of a nuclear bomb combines very low probability with high potential impact, which we take very seriously indeed. If we follow the calculus, however, the privacy risk may be just as big, but we tend ignore it because the risk is so distributed.

Function creep

Big Data solutions are susceptible to *function creep*.[47] One might almost say that Big Data – with its emphasis on the value of secondary use of data – requires function creep in principle. *Function creep* is a source of concern in the field of security due to differences in the legal authority of various agencies to collect data and the far-reaching consequences in the everyday lives of citizens that may result from actions taken on the basis of Big Data analyses. In the domain of security, there is also a trickledown effect in which hardware and software originally designed for security and intelligence agencies finds its way to lower-level security organisations, such as law enforcement and surveillance.[48]

Chilling effects

The large-scale collection, storage and analysis of data by government bodies, including intelligence and security services, and the loss of anonymity on the Internet, can give people the feeling that their privacy and freedom of expression are in danger. This can undermine civil liberties and lead to *chilling effects*, that is, cause people to modify their behaviour and restrict their own freedom because they know they are being monitored. Some maintain that chilling effects, rather than loss of privacy, are the real cost of the NSA activities exposed by Snowden.[49] The damaging effects are greatest for people and organisations that matter to the functioning of democracy, such as journalists, writers, whistleblowers, NGOs and lawyers.[50]

Transparency paradox

Big Data can lead to a transparency paradox: citizens become increasingly transparent to government, while the profiles, algorithms and methods used by government organisations are hardly transparent or comprehensible to citizens.[51] The result is a shift in the balance of power between state and citizen in favour of the former. The secret nature of activities in the field of security reinforces this transparency paradox.[52] Now that large numbers of citizens are increasingly coming under the spotlight – citizens who are linked to profiles and may become subject to decision-making based on these profiles (the needles) as well as those who are not (the haystack) – this will increasingly cause friction.[53]

[47] Lyon (2007: 52) describes function creep as the expansion of the use of a system or token to include other functions not originally envisioned by their promotors.

[48] Završnik (2013).

[49] See Walt (2013).

[50] For empirical evidence on chilling effects caused by online surveillance, see Penney (2016).

[51] Richards and King (2013); Schneier (2015).

[52] Broeders (2016: 299).

[53] Hildebrandt (2016); De Hert and Lammerant (2016).

9 A MIXED LEGAL FRAMEWORK FOR BIG DATA AND SECURITY

Big Data is here to stay. Sooner or later its development will take off, also in the field of security. It is essential, therefore, to manage the use of Big Data effectively. Big Data analyses have the potential to make a valuable contribution to the security and freedom of society, but for this to happen, they must be made on a solid legal basis, covering the risks presented by Big Data and including measures for dealing with them or compensating for them.

The current rules and regulations

The European tradition of data protection and privacy regulation is predicated on fundamental rights, which is different from the American tradition, which is 'characterized by a sectoral approach to legislating modern data privacy issues (including through self-regulation) as well as the adoption of information privacy law at the State level'.[54] Data protection regulation in Europe is based on the EU Data Protection Directive, which has been replaced by the General Data Protection Regulation[55] (GDPR), which will enter into force on 25 May 2018. Under the Directive, EU member states were responsible for translating the EU regulations into national law – the Personal Data Protection Act in the Netherlands. The new GDPR will be directly applicable without requiring any national translation. The GDPR will not be applicable to the police and justice sector, whose work will be regulated by national legislation to be based on the new EU Police and Criminal Justice Data Protection Directive.[56]

Fundamental rights and security exceptions

The regulation of data protection and privacy is founded on fundamental rights that are enshrined in treaties such as the European Convention on Human Rights (ECHR) and the Charter of Fundamental Rights of the European Union. Article 8 of the ECHR and articles 7 and 8 of the Charter[57] lay down the fundamental rights of privacy and data protection. Art. 8.1 of the ECHR, on the 'Right to respect for private and family life', or the right to privacy, reads: 'Everyone has the right to respect for his private and family life, his home and his correspondence.' However, these international agreements make an exception for matters relating to public safety and (national) security, where, on the basis of national law, the responsible agencies usually enjoy more authority to conduct operations that 'violate' privacy and data protection. Article 8(2) of the ECHR states

[54] Van Hoboken (2016: 242).

[55] Regulation (EU) 2016/679, adopted on 27 April 2016.

[56] Directive (EU) 2016/680 of 27 April 2016, replacing Framework Decision 2008/977/JHA, entering into force on 6 May 2018.

[57] Charter of Fundamental Rights of the European Union, Art. 7 on 'Respect for private and family life' reads: 'Everyone has the right to respect for his or her private and family life, home and communications.' Art. 8 on 'Protection of personal data' reads: '(1) Everyone has the right to the protection of personal data concerning him or her.(2) Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified. (3) Compliance with these rules shall be subject to control by an independent authority.'

that: ‘There shall be no interference by a public authority with the exercise of this right *except* such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic wellbeing of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others’ [emphasis added].

A mixed legal framework for security and law enforcement

Although security and public order are exempt from the general data protection and privacy regulations, they are not unregulated in this respect. The activities of security and intelligence agencies and law enforcement are guided by specific laws and regulations (see Table 9.1 for the Dutch situation), and in some cases oversight is entrusted to a party other than the national Data Protection Authority. In the case of the Netherlands, for example, the activities of the military and general intelligence services are regulated by the Intelligence and Security Agencies Act 2002 (Wiv 2002) – which is currently being revised – and oversight is entrusted to the Review Committee on the Intelligence and Security Services (CTIVD)[58].

Table 9.1 Legal frameworks

<table>
<tr><th colspan="3">Constitutional framework

International Covenant on Civil and Political Rights (ICCPR)
European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR)
Charter of Fundamental Rights of the European Union

Constitution of the Netherlands</th></tr>
<tr><th>Police and judiciary</th><th>Intelligence and security services</th><th>Government (other agencies)</th></tr>
<tr><td>Code of Criminal Procedure (WvSv)
Judicial Data and Criminal Records Act (Wjsg)
Police Data Act (Wpg)
Special Investigative Services Act (Wet BOD)
Police and Criminal Justice Data Protection Directive (EU) 2016/680</td><td>Intelligence and Security Services Act (Wiv 2002)</td><td>Personal Data Protection Act (Wbp)
Data Protection Directive (95/46/EC)
General Data Protection Regulation (EU) 2016/679</td></tr>
</table>

Law enforcement and public prosecuting authorities operate under their own legal framework, often with specific laws regulating the collection, exchange and use of data within the police organisation and the wider law enforcement community. Other government agencies, such as the tax agency and partnerships of agencies cooperating in anti-fraud data projects, are usually covered by the general data protection legislation, although specific laws may lend considerable authority to organisations to collect, commandeer and analyse data. The Dutch tax authority, for example, has extensive powers to collect and analyse data that far outstrip those of most other government agencies. Law enforcement and other agencies working in a broadly

[58] In Dutch: Commissie van Toezicht op de Inlichtingen- en Veiligheidsdiensten, see http://english.ctivd.nl.

defined security domain are usually subject to oversight by the national DPA, such as the Dutch Personal Data Authority.[59]

Because security and intelligence agencies, law enforcement and public prosecuting authorities, and government agencies cooperating in anti-fraud projects operate under different legal frameworks, there is no one-size-fits-all solution for dealing with Big Data within the field of security. Nevertheless, these frameworks have a common denominator: their main goal is to regulate data collection. There are good practical and legal reasons why this should change.

10 REGULATION: FROM DATA COLLECTION TO DATA ANALYSIS AND USE

The inherent tension between Big Data and current data protection law

In its ideal form, Big Data is based on the principle of unfocused data collection, as well as on linking and reusing data collected for other purposes and by other parties. Secondary use of data and the idea that more data leads to more accurate – as well as unexpected – insights are core ingredients of the promise of Big Data. The current European and national rules and regulations, however, are mainly concerned with the initial data collection phase and are built on legal principles, some of which are at loggerheads with the ideal form of Big Data analysis.

Big Data puts pressure on important legal principles such as purpose limitation and data minimisation, which are strongly connected with the data collection phase. Purpose limitation stipulates that data may only be collected and stored if there is a clear purpose for data processing. With large amounts of data being available, it is increasingly common to collect data first and only sort it into usable and unusable data afterwards. Moreover, the general trend is to combine data from various sources, build profiles and mine the resulting databases.

The principle of data minimisation requires that as little data as possible is collected, and in any case no more than is necessary in order to achieve the specific goal that has been formulated. Moreover, these data must be deleted once the aim has been achieved. The Big Data logic that 'more is better' and the idea that the value of Big Data analysis lies in the secondary use of data, therefore, puts pressure on core legal principles and also highlights the limits of the current focus of data protection law on the data collection phase as the main target for regulation.

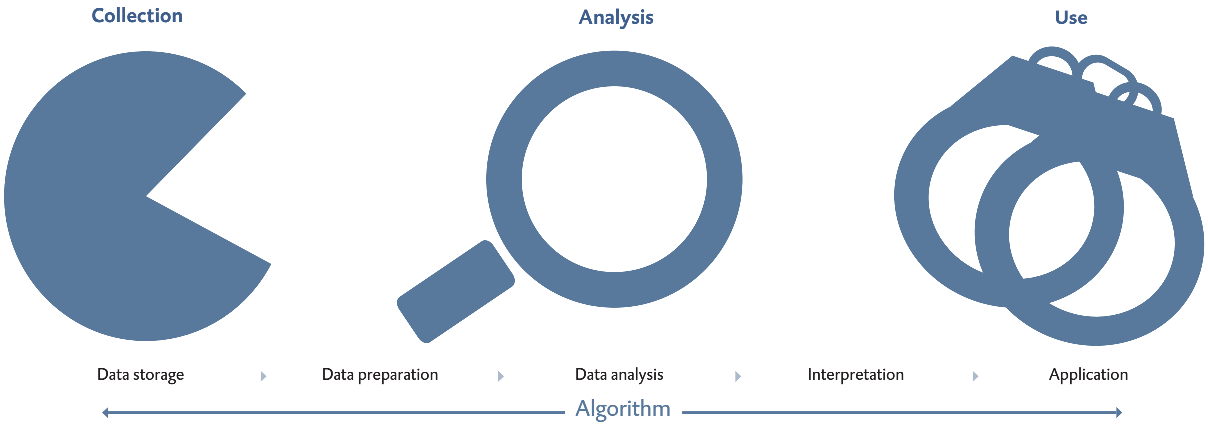
Looking beyond data collection: regulating data analysis and use

Existing regulations on collecting data and those that will enter into force shortly – such as the GDPR – continue to have an important function in the era of Big Data. For example, data must be legally obtained by those that process it, which may require extra

[59] In Dutch: Autoriteit Persoonsgegevens. See https://autoriteitpersoonsgegevens.nl/en.

effort in the era of Big Data. Given the pressure on the current legal framework, however, there is reason to look beyond regulating the data collection phase. Current academic and public debates, for example, entertain the notion of shifting the emphasis from regulating data collection to its use[60], and there are publications that underline the need for 'algorithmic accountability'[61], indicating a need to scrutinise the data analysis phase. Therefore, we have divided the processes of Big Data analytics into three main phases: *collection*, *analysis* and *use*.

Figure 10.1 Big Data process in stages



Considering the current legal framework, it would seem obvious to regulate data collection more strictly and to enforce purpose limitation in particular. However, this would mean adding more black letter law to an already densely regulated phase of Big Data processes and would also largely nip the promise of Big Data in the bud. Moreover, it remains to be seen to what extent the legal principle of purpose limitation will hold its own given the high pressures in commercial practice and in public policy to share data and conduct Big Data analyses. Some scholars consider data minimisation to be out of touch with reality[62] or argue that 'social trends and technological developments (such as Big Data and the Internet of Things) mean that we will have to abandon the principle of purpose limitation as a separate criterion.'[63] It is worthwhile, therefore, to pursue an approach in which the emphasis shifts from regulating data collection to regulating data analysis and use. There is more to be achieved in these later stages of Big Data processes than by intensifying the regulation of data collection.

[60] Van Hoboken (2016); Ohm (2010); Koops (2013); Van der Sloot (2016).

[61] See Diakopoulos (2013), (2016). From a research perspective see Kitchin (2017). For a more technical approach to this question see Kroll et al. (2017).

[62] Koops (2014: 8), 'The Data Protection Directive has done little to prevent the development of massive databases or the advent of the Big Data era, and it is folly to think that the GDPR will fare better in preventing "unnecessary" data processing. Who in his right mind can look at the world out there and claim that a principle of data minimisation exists?'

[63] Moerel and Prins (2016: 2).

11 REGULATING ANALYSIS: LOOKING INTO THE BLACK BOX

In Big Data processes, the important choices are made in the phase of the analysis: selecting the algorithms, data sources and categorisation, assigning weight to various data, et cetera. It is in this phase of Big Data processes – the algorithmic heart – that the various risks that we outlined earlier materialise.

Duty of care

In the current legal regime, the analysis phase has remained relatively unregulated, and algorithmic accountability is by and large lacking. To address this, quality criteria should be made more explicit.

To increase organisational awareness and to create more accountability, a legal and explicit duty of care should be introduced for government organisations using Big Data analysis in the domain of security.

It is impossible to prescribe in advance precisely what conditions must be met in the analysis phase, as each case is different. However, there are some general requirements for data quality, methodological soundness and insight into the algorithms used that constitute a base line:

1. Government organisations must ensure that their data is up-to-date and that they are aware of and correct for the bias contained in their datasets. They need a strategy to mitigate such bias. This obligation also covers data obtained from third parties.
2. Data used must be obtained legitimately from third parties, who, in their turn, must have obtained them legitimately.[64]
3. The algorithms and methodology must be sound and must meet the scientific criteria for good (statistical) research.
4. Algorithms and methodological choices must be open to review by the competent oversight authority.[65] This may prove problematic in the case of commercial algorithms which the supplier considers to be proprietary trade secrets.[66] Nonetheless, research results, profiles and correlations must be open to oversight:

[64] This is especially important in the field of the Intelligence and Security agencies when they obtain data from foreign counterparts that operate under different legal regimes and have different authorities.

[65] Algorithmic accountability does not always and under all circumstances require full transparency of the source code. What is needed has to be assessed on a case-by-case basis (Diakopoulos 2016: 58; see also Kroll et al. 2017).

[66] See, for example, the case of Mr. Loomis who was sentenced in a Wisconsin court on the basis of a proprietary algorithm (a so-called Compas assessment) that labeled him a 'high risk' to the community. Mr. Loomis could not get access to the algorithm, which was considered a trade secret, and now challenges his sentence on the basis that it is unclear on what grounds he received this label. See http://www.nytimes.com/2016/06/23/us/backlash-in-wisconsin-against-using-data-to-foretell-defendants-futures.html?_r=0.

the data-processing party must be able to show clearly how they arrived at particular results.

The various aspects of duty of care – a series of quality criteria – are discussed during monitoring of the analysis process and ex post scrutiny by the oversight authority. Responsibility for data quality and methodological soundness remains with the data-processing party at all times.

External reviews and audits

In view of the great importance of the data-analysis phase, which is in fact the core of Big Data processes, external regulation should also be strengthened on this point.

Big Data projects in the field of security must be made subject to external review by the oversight authority, which should particularly monitor choices made with respect to data and methods of analysis. In their review, the oversight authority should also check whether duty of care principles have been complied with.

These reviews may be combined with internal audits that are already in place. In the Netherlands, for example, Section 33 of the law Pertaining to Police Data[67] requires that internal audits are sent to the *Autoriteit persoonsgegevens*, the Dutch DPA. A similar process could be used. The audits can be done annually for large government data processing projects, in particular those in the field of security, in view of their potential consequences for individuals and groups of citizens. The report that is sent to the DPA must enable the oversight authority to gain a clear picture of the sources, data and methods employed.

The same obligation should apply to the intelligence services. In the Dutch context, the body that is reported to is the Review Committee on the Intelligence and Security Services (CTIVD). The technical and statistical capacity and expertise of both the AP and the CTIVD need to be strengthened in order for them to conduct meaningful oversight in a world of Big Data. The AP and the CTIVD then report to the Lower House of Parliament. In the case of the CTIVD, where national security and secrecy play an important role, a new process is needed to determine how it should report to Parliament. Preferably it should report to the Parliamentary standing committees on the Interior (regarding the work of the General Intelligence and Security Service) and Defence (regarding the work of the Military Intelligence and Security Service), which are more transparent to the general public; only where necessary should it report to the Committee for the Intelligence and Security Services (CIVD), which deliberates behind closed doors on the explicit understanding that everything discussed there will not go any further. In their reports to Parliament, the oversight authorities can provide a more detailed opinion regarding possible regulation or setting new boundaries.

[67] Art. 33 WPG.

Sunset clauses

When setting up new Big Data projects, the planning should include a date for their evaluation. This is important given the potential positive and negative impact of data-driven applications, especially in the field of security. Another reason for integrating a moment of evaluation at the outset is the fact that governments often allow large ICT projects to continue to run, even if it is realistic and perhaps better to end them.[68] The internal political and policy dynamic often keeps the engine running, regardless of other considerations, and projects become entrenched.

Large data-processing projects in government, particularly by the police, intelligence and security services, inspection bodies, the tax authority and anti-crime and anti-fraud cooperation bodies must be subject to a sunset clause of three to five years.

A three to five year period gives projects sufficient time to develop and prove their worth, but it is also short enough to be able to intervene at an early stage. In evaluating projects, there are three specific assessment points: firstly, it should be assessed whether there is still any need for the project: circumstances may change, after all. Secondly, it should be assessed whether the data-processing process was effective: did Big Data achieve the aims that the project set out to achieve? Points one and two can only be assessed meaningfully if realistic and measurable goals were formulated at the beginning of the project. The evaluation must clarify whether these goals have been achieved fully, partially or not at all, and to what extent the results can be attributed to the Big Data analysis. If it is found that these goals have not been achieved, or only to a very limited degree, the project will be stopped. If there are limited positive results, a plan to revise the project will be needed. Thirdly, the evaluation must include a cost-benefit analysis, which must explicitly include a proportionality and subsidiarity test with respect to the effects on personal freedom and security. These fundamental rights must be considered explicitly: what are the concrete benefits that the infringement of these rights of citizens has delivered? This evaluation can be modelled on what are called Surveillance Impact Assessments (SIA). These consist of four connected elements: the impact on individual privacy; the impact on individual relations, positions and freedoms; the impact on groups and categories; and the broader impact on society and the political system.[69] This evaluation, therefore, is considerably more comprehensive than the conventional Privacy Impact Assessment (PIA), which is required by the upcoming Regulation.

A report should be compiled on these three points and sent to the competent authority, which can then report to parliament. This way, accountability is divided into stages, with a *trusted* external party in the form of an oversight authority first of all, followed by a public report to parliament. As reports to oversight authorities are not made public, there is room for greater transparency and detail. The subsequent public report

[68] Prins et al. (2011); Prins, Broeders and Griffioen (2012).

[69] Raab and Wright (2012); see also Bennett and Bayley (2016: 216-219).

that is sent to parliament does not reveal the 'tricks of the trade', i.e. the specific methods used, but it is written with knowledge and understanding of them.

12 REGULATING USE: BIG CONSEQUENCES MEAN BIG RESPONSIBILITY

Big Data analysis should result in actionable knowledge.[70] At some time, some person or persons will be confronted with the results of an analysis in the real world: the tax authorities may investigate, or the police may knock on someone's door. The real life consequences – which may be especially felt when it is about security considerations – merits a very thorough scrutiny of how Big Data analyses contribute to decision-making processes and their practical use.

Bounding profiles

In the use of data analyses in Big Data processes, i.e. the consequences of decisions made on the basis of analyses, *profiling* stands out as an important issue. The power of Big Data analyses lies mainly in detecting structural patterns at the aggregate level. When these general insights are applied to real situations and specific individuals and groups, there is always a mismatch because a profile is always over-inclusive as well as under-inclusive. Benchmarks need be developed to determine admissible margins of error when working with profiles to determine action 'in the real world'. These benchmarks should be linked to both to the importance of the service or organisation for security and to the impact on individual and collective fundamental rights.

The use of profiling requires more detailed rules on admissible margins of error.

No (semi-)automatic decision-making

Profiles are increasingly influential in making choices and decisions. There is a tendency to follow profiles and patterns fairly uncritically and to regard computer analyses as quasi objective. Current rules relating to automated individual decisions in the EDPD (Art. 15) and in the upcoming GDPR are generally considered to be weak.[71] It is up to individual states and the oversight regime to ensure that automated decision-making is banned and remains so, as 'computer says no' can never be allowed to be the end of an exchange between government and citizen. Those responsible must also be more alert in responding to semi-automatic decision-making, in which a human being formally makes the decision but does not or dare not deviate from the digital advice obtained.[72]

[70] 'Analytics indicates the analysis of so-called "raw" data in search of patterns, and the consequent transformation of these results into the kind of knowledge decision makers need to optimally orientate their course of action, also known as "actionable" knowledge' (Gandy (2012) in Degli Esposti (2014: 211-212)).

[71] Bygrave (2001); Hildebrandt (2009); Savin (2013).

[72] See, for example, Bovens and Zouridis (2002).

The existing ban on automated decision-making should be strictly enforced, and government agencies should be more alert to semi-automatic decision-making.

Own the data, own the consequences

Care must be taken that data analyses and profiles do not lead to an actual reversal of the burden of proof. This is not really a factor in criminal law, where there are strict rules for evidence to be admissible in court, but it may play a role in various forms of surveillance, preliminary investigations, network analyses, enforcement and anti-fraud work. With increasing data exchange and inter-agency cooperation, the analytics in the government's back offices will gain in importance. The danger is that individual citizens, instead of the government, will have to prove that they have been wrongly associated with a profile when a dispute arises about decisions based on data analyses. Because Big Data processes in the field of security shift the balance of power even further 'in favour' of government, citizens need to gain a better grasp of the decisions that affect them as well as strengthen their position vis-à-vis the government.

The principle that responsibility for the accuracy of Big Data processes remains with the data-processing party at all times must be anchored in legislation. The party that *acts* on the basis of an analysis is required – to be able – to show what a decision is based on and what factors and considerations were taken into account.

13 WHO WATCHES THE WATCHERS: REINFORCING OVERSIGHT AND STRENGTHENING TRANSPARENCY AND JUDICIAL REVIEW

The use of Big Data in the security domain requires intensified oversight. An effective and confidence-inspiring oversight regime, in its turn, requires a higher degree of data processing transparency. In this, transparency is not an aim in itself but serves the interests of *accountability*. Citizens and organisations must also have opportunities to discuss the accuracy and proportionality of decisions based on data analyses and made by government institutions and, if necessary, to have them assessed by the courts.

Big Data, Big Oversight?

Current oversight of data processing leaves a lot to be desired, even more so in view of the ongoing rapid digitisation of government and use of data analysis. DPAs and the various forms of oversight on security and intelligence agencies do not appear to be properly equipped to face the challenges of the Big Data era in terms of powers, expertise and financial resources.[73] In the Netherlands, many parties, including the

[73] FRA (2015).

oversight committee CTIVD itself[74], believe that the planned expansion of the powers of the Dutch general and military intelligence agencies requires oversight capacity and technical expertise to be significantly expanded at all levels. [75] Although the powers and resources of national DPAs will in principle be increased through the entry into force of the GDPR, it is up to the national legislatures to allocate the corresponding financial resources, powers and capacity. As indicated above, oversight of the analysis phase will be of paramount importance.

If possibilities for collecting and analysing data increase, independent oversight should be strengthened. For regulating the intelligence services, it would be appropriate, in view of the need to protect fundamental rights, to introduce an overriding power/ the possibility of passing judgments regarding lawfulness.[76]

Transparency

Big Data also requires greater transparency in the government's data-processing activities. There is still a lot to be gained on this point, as data processing is a 'black box' in many cases. In addition, data subjects are not so quick to invoke their right to information because they often simply do not know that their data is being collected. Given the sensitive nature of the work of law enforcement and national security agencies, there cannot be full transparency, which is precluded by the danger of 'gaming the system' and the need to protect sensitive and classified information. Nevertheless, it is possible to work with a layered system of transparency, as was suggested above in relation to reports by the police and the intelligence services to the oversight competent authorities, which, in their turn, report to parliament.

Big Data: for the people, known by the people?

It would also be desirable to give citizens more insight into the frequency of data collection, the reasons why it is done, and, if possible, what effect complex data analyses have. In the field of security, some organisations have the legal right to keep all or parts of their operations secret from data subjects and the general public. However, the growing amount of data that the government can obtain under existing secrecy provisions is out of step with the data processing transparency that is required. There are indications, moreover, that the agencies involved in national security are prone to a culture of 'overclassification'.[77] Furthermore, better information meets a democratic need; a well-informed discussion of the use of Big Data solutions requires a better understanding of the use of data by government organisations in the area of security policy.

[74] CTIVD (2015: 11).

[75] Eskens, Van Dalen and Van Eijk (2015). See also the internet consultations that were collected during an earlier phase of the revision process of the Dutch Intelligence and Security Services Act (Wiv 2002). Many parties that contributed to the Internet consultation on this bill called for a strengthening of external regulation of intelligence and security services, which is seen as a weak point.
See https://www.internetconsultatie.nl/wiv.

[76] See, for example, FRA (2015); Loof et al. (2015); and Eskens, Van Dalen and Van Eijk (2015).

[77] Curtin (2011: 18-19); Schneier (2015: 99).

Data processing transparency must be enhanced, and a better balance must be achieved between the secrecy requirement and the interests of openness as regards Big Data applications that affect fundamental freedoms.

Greater transparency is needed in at least two areas. A growing number of organisations in the field of security are involved in Big Data applications, above all in the area of fighting fraud. There is a lack of good regulation in this area. Although a lot of relevant information about data processing in partnerships is contained in publically available covenants and decisions, this is not very accessible. Citizens should not have to be detectives to find the relevant information. There should be greater openness when organisations intend to work with Big Data applications, for example, by requiring them to draw up a policy plan detailing what Big Data applications they use, what public policy they pursue, at what cost and what results they expect from the application.

At the accountability level, too, more is possible than is currently being done, for example in annual reporting. Compared to the Netherlands, some European countries are practising a significantly higher degree of openness about their intelligence techniques and operations, without this hindering the work of their intelligence services in any noteworthy way. This is being done in Belgium, for example, with the aim of being able to conduct an informed discussion about how the intelligence and security services work, what their powers are and how they are monitored and regulated.[78]

Judicial review

The imbalance of power between citizens and the government in relation to data-processing capacity and techniques is expected to grow in the era of Big Data. It is important, therefore, to strengthen the citizens' position. This will happen in part by granting the oversight authorities greater powers to monitor and control activities and by increasing data processing transparency. It will also be achieved by ensuring, in the matter of the accuracy of Big Data processes, that the burden of proof rests firmly with the data-processing parties. However, it is also important to give citizens themselves a stronger voice in scrutinising and monitoring Big Data applications. Citizens can give voice to their interests either directly or through interest groups.

In the Netherlands, however, the right of complaint is strongly linked to the notion of individual harm, and possibilities for collective proceedings are very limited.[79] This leaves citizens but few possibilities to question decision-making based on Big Data processes if they cannot produce evidence of joint personal disadvantage. The Dutch constitutional order lacks an independent judicial review of legislation if personal damage has not been demonstrated, and this has caused the focus of the judicial review

[78] See www.comiteri.be/images/pdf/Jaarverslagen/Activiteitenverslag_2014.pdf. See pp. 70-77 for the statistics on intelligence operations.

[79] Van der Sloot (2016).

process to shift to the European Court of Human Rights and the European Court of Justice. Citizens who are concerned about the social effects of Big Data applications should have greater access to national options for judicial review of legislation and Big Data policy initiatives must be improved.

Because many of the large data-processing projects extend beyond the individual, it is important not only to concentrate on individual rights but also to strengthen and consolidate the position of NGOs and citizens' rights organisations in legal procedures. This does not mean that the courts should open their doors to every class action lawsuit, but there should be selective admission of cases that address collective concerns and contribute to the development of case law in this important and relatively undeveloped area.[80]

14 SERVING SECURITY, PROTECTING FREEDOM

Big Data has a lot to offer for surveillance, investigation and prevention in the field of security. However, Big Data processes can also have a significant impact on citizens, even if they are innocent and not suspected of anything. The application of Big Data, therefore, must be accompanied by additional measures to protect fundamental rights. Only under this condition can Big Data make a substantial contribution to security and freedom.

[80] Zwenne and Schmidt (2016).

REFERENCES

Akerkar, R., G. Lovoll, S. Grumbach, A. Faravelon, R. Finn and K. Wadhwa (2015) 'Understanding and mapping Big Data', deliverable 1.1 byte project, available at: http://byte-project.eu/wp-content/uploads/2016/03/BYTE-D1.1-FINAL-post-Y1-review.compressed-1.pdf.

Anderson, C. (2008) 'The end of theory: The data deluge makes the scientific method obsolete', *Wired Magazine* 16.07, available at: http://www.uvm.edu/~cmplxsys/wordpress/wp-content/uploads/reading-group/pdfs/2008/anderson2008.pdf.

Bartlett, J., C. Miller, J. Crump and L. Middleton (2013) *Policing in an information age*, London: Demos.

Bennett, C. and R. Bayley (2016) 'Privacy Protection in the Era of 'Big Data': Regulatory Challenges and Social Assessments', pp. 205-227 in B. van der Sloot, D. Broeders and E. Schrijvers (eds.) *Exploring the Boundaries of Big Data*, WRR-Verkenning 32, Amsterdam: Amsterdam University Press.

Bennett Moses, L. and J. Chan (2014) 'Using Big Data for legal and law enforcement decisions: Testing the new tools', *University of New South Wales Law Journal* 37, 2: 643-678.

Bovens, M. and S. Zouridis (2002) 'From street-level to system-level bureaucracies: How information and communication technology is transforming administrative discretion and constitutional control', *Public Administration Review* 62, 2: 174–84.

Brakel, R. van (2016) 'Pre-emptive Big Data Surveillance and its (Dis)empowering Consequences: The case of Predictive Policing', pp. 117-141 in B. van der Sloot, D. Broeders and E. Schrijvers (eds.) *Exploring the Boundaries of Big Data*, WRR-Verkenning 32, Amsterdam: Amsterdam University Press.

Broeders, D. (2016) 'The secret in the Information Society', *Philosophy and Technology* 29, 3: 293–305.

Broeders, D. and J. Hampshire (2013) 'Dreaming of seamless borders: ICTs and the preemptive governance of mobility in Europe', *Journal of Ethnic and Migration Studies* 39, 8: 1201-1218.

Broeders, D. and H. Dijstelbloem (2016) 'The datafication of mobility and migration management: The mediating state and its consequences', pp. 242-260 in I. van der Ploeg and J. Pridmore (eds.), *Digitizing identities*, London: Routledge.

Bygrave, L.A. (2001) 'Automated profiling: Minding the machine: Article 15 of the EC Data Protection Directive and Automated Profiling', *Computer Law and Security Review* 17, 1: 17-24.

CTIVD (2015) *Jaarverslag 2014 - 2015*, Den Haag: Commissie van Toezicht betreffende de Inlichtingen- en Veiligheidsdiensten.

Curtin, D. (2011) *Top secret Europe*, inaugural lecture University of Amsterdam, available at: http://dare.uva.nl/document/2/103309.

Custers, B. and B. Vergouw (2015) 'Promising policing technologies: Experiences, obstacles and police needs regarding law enforcement technologies', *Computer Law and Security Review* 31, 4: 518-526.

Degli Esposti, S. (2014) 'When Big Data meets dataveillance: The hidden side of analytics', *Surveillance and Society* 12, 2: 209-225, available at: http://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/analytics/analytic.

Diakopoulos, N. (2013) *Algorithmic accountability reporting: On the investigation of black boxes*, Tow Center for Digital Journalism, available at: http://cjrarchive.org/img/posts/tow-content/uploads/2014/02/78524_Tow-Center-Report-WEB-1.pdf.

Diakopoulos, N. (2016) 'Accountability in algorithmic decision making. A view from computational journalism', *Communications of the ACM* 59, 2: 56-62.

Dijcks, J. (2012) *Oracle: Big Data for the enterprise*, Oracle White Paper, available at: www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf.

Dumbill, E. (2013) 'Making sense of big data editorial', *Big Data* 1, 1: 1-2, available at: http://online.liebertpub.com/doi/abs/10.1089/big.2012.1503.

Eskens, S., O. van Dalen and N. van Eijk (2015) *Ten standards for oversight and transparency of national intelligence services*, Amsterdam: IVIR, available at: http://www.ivir.nl/publicaties/download/1591.pdf.

Ferguson, A. (2017) 'Policing predictive policing', *Washington University Law Review* 94 (forthcoming), available at: https://papers.ssrn. com/sol3/papers.cfm?abstract_id=2765525.

FRA (2015) 'Surveillance by intelligence services: Fundamental rights safeguards, and remedies in the EU. Mapping Member States' legal frameworks', available at: http://fra.europa.eu/en/publication/2015/surveillance-intelligence-services.

Galdon Clavell, G. (2013) '(Not so) smart cities? The drivers, impact and risks of surveillance-enabled smart environments', *Science and Public Policy 40,* 6: 717–723.

Galdon Clavell, G. (2016) 'Policing, Big Data and The Commodification of Security', pp. 89-115 in B. van der Sloot, D. Broeders and E. Schrijvers (eds.) *Exploring the Boundaries of Big Data*, WRR-Verkenning 32, Amsterdam: Amsterdam University Press.

Gandy, O. jr. (2009) *Coming to terms with chance: Engaging rational discrimination and cumulative disadvantage*, Farnham and Burlington: Ashgate.

Gillespie, T. (2014) 'The relevance of algorithms', pp. 167-194 in T. Gillespie, P. Boczkowski and K. Foot (eds.) *Media technologies: Essays on communication, materiality, and society*, Cambridge (MA): MIT Press.

Gitelman, L. (ed.) (2013) *Raw Data is an Oxymoron*, Cambridge (MA): MIT Press.

Greengard, S. (2015) *The Internet of Things*, Cambridge (MA): MIT Press.

Greenwald, G. (2014) *No Place to Hide. Edward Snowden, the NSA and the US Surveillance State*, New York: Metropolitan Books.

Halevy, A., P. Norvig, and F. Pereira (2009) 'The unreasonable effectiveness of data', *IEEE Intelligent Systems* March/April 2009: 8-12.

Harcourt, B. (2007) *Against prediction: Profiling, policing and punishing in an actuarial age*, Chicago: Chicago University Press.

Hert, P. de, and H. Lammerant (2016) 'Predictive Profiling and its Legal Limits: Effectiveness Gone Forever?', pp. 145-173 in B. van der Sloot, D. Broeders and E. Schrijvers (eds.) *Exploring the Boundaries of Big Data*, WRR-Verkenning 32, Amsterdam: Amsterdam University Press.

Hildebrandt, M. (2009) 'Who is profiling who? Invisible visibility', pp. 239-252 in S. Gutwirth, Y. Poullet, P. de Hert, J. Nouwt and C. de Terwangne (eds.) *Reinventing data protection?*, Dordrecht: Springer.

Hildebrandt, M. (2016) 'Data gestuurde intelligentie in het strafrecht', pp. 137-240 in E.M.L. Moerel, J.E.J. Prins, M. Hildebrandt, T.F.E Tjong Tjin Tai, G-J. Zwenne en A.H.J. Schmidt, *Homo Digitalis*, Handelingen Nederlandse Juristen-Vereniging 146e jaargang/ 2016-I, Wolters Kluwer, available at: http://njv.nl/wp-content/uploads/2011/04/Preadviezen-NJV-2016.pdf.

Hoboken, J. van (2016) 'From Collection to Use in Privacy Regulation? A Forward-Looking Comparison of European and U.S. Frameworks for Personal Data Processing', pp. 231-259 in B. van der Sloot, D. Broeders and E. Schrijvers (eds.) *Exploring the Boundaries of Big Data*, WRR-Verkenning 32, Amsterdam: Amsterdam University Press.

Hopkins, B. and B. Evelson (2011) 'Expand Your Digital Horizon With Big Data', Forrester, available at: http://www.asterdata.com/newsletter-images/30-04-2012/resources/forrester_expand_your_digital_horiz.pdf.

IBM (2015) *What is Big Data?*, available at: http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html.

Jerven, M. (2013) *Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It*, Ithaca: Cornell University Press.

Jeandesboz, J. (2016) 'Smartening border security in the European Union: An associational inquiry', *Security Dialogue* 47, 4: 292-309.

Kerr, I. and J. Earle (2013) 'Prediction, preemption, presumption: How Big Data threatens big picture privacy', *Stanford Law Review Online* 66, 65: 65-72, available at: https://www.stanfordlawreview.org/online/privacy-and-big-data-prediction-preemption-presumption.

Kitchin, R. (2017) 'Thinking critically about and researching algorithms', *Information, Communication & Society* 20, 1: 14-29.

Kitchin, R. (2014a) 'The real-time city? Big data and smart urbanism', *GeoJournal* 79, 1: 1-14.

Kitchin, R. (2014b) *The data revolution: Big Data, open data, data infrastructures and their consequences*, London: Sage.

Klous, S. (2016) 'Sustainable Harvesting of the Big Data Potential', pp. 27-47 in B. van der Sloot, D. Broeders and E. Schrijvers (eds.) *Exploring the Boundaries of Big Data*, WRR-Verkenning 32, Amsterdam: Amsterdam University Press.

Koops, B.J. (2013) 'On decision transparency, or how to enhance data protection after the computational turn', pp. 196-220 in M. Hildebrandt and K. de Vries (eds.) *Privacy, due process and the computational turn*, Abingdon: Routledge.

Koops, B.J. (2014) 'The Trouble with European Data Protection Law', *International Data Privacy Law* 4, 4: 250–261.

Kroll, J., J. Huey, S. Borocas, E. Felten, J. Reidenberg, D. Robinson and H. Yu (2017) 'Accountable Algorithms', *University of Pennsylvania Law Review* 165 (forthcoming 2017).

Laney, D. (2001) '3D data management: Controlling data volume, velocity and variety', available at: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

Loof, J., J. Uzman, T. Barkhuysen, A. Buyse, J. Gerards and R. Lawson (2015) 'Het mensenrechtenkader voor het Nederlandse stelsel van toezicht op de inlichtingen- en veiligheidsdiensten', available at: http://www.ctivd.nl/documenten/publicaties/2015/08/26/rapport-universiteit-leiden.

Lyon, D. (2007) *Surveillance Studies: An Overview*, Cambridge: Polity Press.

Mayer-Schönberger, V. and K. Cukier (2013) *Big Data. A revolution that will transform how we live, work and think*, London: John Murray Publishers.

Moerel, L. and C. Prins (2016) *Privacy for the Homo Digitalis: Proposal for a New Regulatory Framework for Data Protection in the Light of Big Data and the Internet of Things* (May 25, 2016), available at: https://ssrn.com/abstract=2784123.

Montjoye, Y. de, C. Hidalgo, M. Verleysen and V. Blondel (2013) 'Unique in the crowd: The privacy bounds of human mobility', *Scientific Reports* 3, available at: http://dx.doi.org/10.1038/srep01376.

OECD (2014) *Data-driven innovation: Big Data for growth and well-being*, Paris: OECD Publishing.

Ohm, P. (2010) 'Broken promises of privacy: Responding to the surprising failure of anonymization', *UCLA Law Review* 57: 1701-1777.

Olsthoorn, P. (2016) *Big Data voor fraudebestrijding*, WRR Working Paper 21, Den Haag: WRR.

Omand, D., J. Bartlett and C. Miller (2012) *#Intelligence*, London; Demos.

O'Neil, C. (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishers.

PCAST (2014) *Big Data and privacy: A technological perspective,* Report to the President, available at: https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf.

Penney, J. (2016) 'Chilling Effects: Online Surveillance and Wikipedia Use', *Berkeley Technology Law Journal* 31, 1: 117- 182. Available at: https://ssrn.com/abstract=2769645.

Perry, W., B. McInnis, C. Price, S. Smith and J. Hollywood (2013) *Predictive policing: The role of crime forecasting in law enforcement operations*, Santa Monica: Rand.

Prins, C., D. Broeders, H. Griffioen, A. Keizer and E. Keymolen (2011) *iGovernment*, WRR Report 86, Amsterdam: Amsterdam University Press.

Prins, C., D. Broeders and H. Griffioen (2012) 'iGovernment: a new perspective on the future of government digitisation', *Computer Law & Security Review, 28,* 3: 273–282.

Raab, C. and D. Wright (2012) 'Surveillance: Extending the Limits of Privacy Impact Assessment', pp. 363-383 in D. Wright and P. De Hert (eds.) *Privacy Impact Assessments*, Dordrecht: Springer.

Ratcliffe, J. (2010) 'Crime mapping: Spatial and temporal challenges', pp. 5-24 in A. Piquero and D. Weisburd (eds.) *Handbook of quantitative criminology*, New York: Springer.

Richards, N. and H. King (2013) 'Three paradoxes of Big Data', *Stanford Law Review Online* 41, available at: http://ssrn.com/abstract=2325537.

Rienks, R. (2014) *Predictive policing. Kansen voor een veiligere toekomst*, Apeldoorn: Politieacademie.

Savin, A. (2013) 'Profiling and automated decision making in the present and new EU data protection frameworks', pp. 249-270 in P. Arnt Nielsen, P. Koerver Schmidt and K. Dyppel Weber (eds.) *Erhvervsretlige emner*, Kopenhagen: Juridisk Institut CBS.

Schneier, B. (2015) *Data and Goliath. The hidden battles to collect your data and control your world,* New York: W.W. Norton & Company.

Sloot, B. van der (2016) 'The Individual in the Big Data Era: Moving towards an Agentbased Privacy Paradigm', pp. 177- 203 in B. van der Sloot, D. Broeders and E. Schrijvers (eds.) *Exploring the Boundaries of Big Data*, WRR-Verkenning 32, Amsterdam: Amsterdam University Press.

Taylor, L., L. Floridi and B. van der Sloot (eds.) (2017) *Group privacy. The challenges of new data technologies*, New York: Springer.

TechAmerica Foundation (2012) *Demystifying Big Data*, available at: www1.unece.org/stat/platform/download/attachments/80053387/ Demistyfying%20Big%20Data.pdf?version=1&modificationDate=13742235 53898&api=v2.

Walt, S.M. (2013) 'The real threat behind the NSA surveillance programs', Foreign Policy, available at: http://foreignpolicy.com/2013/06/10/the-real-threat-behind-the-nsa-surveillance-programs.

Werkgroep Verkenning kaderwet gegevensuitwisseling (2014) *Kennis delen geeft kracht. Naar een betere èn zorgvuldigere gegevensuitwisseling in samenwerkingsverbanden*, Den Haag, available at: http://njb.nl/Uploads/2015/1/blg-442395.pdf.

Willems, D. and R. Doeleman (2014) 'Predictive policing: Wens of werkelijkheid?', *Het Tijdschrift voor de Politie* 76, 4/5: 39-42.

WRR (2016) *Big Data in een vrije en veilige samenleving*, WRR Report 95, Amsterdam: Amsterdam University Press.

Zarsky, T. (2016) 'The trouble with algorithmic decisions. An analytic road map to examine efficiency and fairness in automated and opaque decision making', *Science, Technology, & Human Values* 41, 1: 118-132.

Završnik, A. (2013) 'Blurring the line between law enforcement and intelligence: Sharpening the gaze of surveillance?', *Journal of Contemporary European Research* 9, 1: 181-202.

Zikopoulos, P. and C. Eaton (2011) *Understanding Big Data: Analytics for enterprise class Hadoop and streaming data*, McGraw Hill Professional.

Zuiderveen Borgesius, F. (2015) *Improving privacy protection in the area of behavioural targeting*, Alphen aan den Rijn: Kluwer Law International.

Zwenne, G-J. and A. Schmidt (2016) 'Wordt de homo digitalis bestuursrechtelijk beschermd?', pp. 307-385 in E.M.L. Moerel, J.E.J. Prins, M. Hildebrandt, T.F.E. Tjong Tjin Tai, G-J. Zwenne en A.H.J. Schmidt, *Homo Digitalis*, Handelingen Nederlandse Juristen-Vereniging 146e jaargang/2016-I, Wolters Kluwer, available at: http://njv.nl/wp-content/uploads/2011/04/Preadviezen-NJV-2016.pdf.

WRR

THE NETHERLANDS SCIENTIFIC COUNCIL FOR GOVERNMENT POLICY