Eur J Secur Res (2017) 2:57–69
DOI 10.1007/s41125-017-0012-x

ORIGINAL ARTICLE

Big Data Data Mining by the Dutch Police: Criteria for a Future Method of Investigation

S. Brinkhoff[1]

Received: 29 November 2016 / Accepted: 25 January 2017 / Published online: 10 February 2017

Abstract The term Big Data refers to the phenomenon of an ever larger and increasingly complex number of digital data and data files that keep growing in scope continuously and exponentially. It is a known fact that worldwide different intelligence agencies employ automated data analysis, known as data mining, on data and data files to understand Big Data. More and more also the Dutch police use automated data analysis of data and data files and real Big Data data mining as a method of investigation in criminal proceedings. Even though Big Data data mining can be a potentially useful and effective method of police investigation, there are some uneasy aspects associated with it. These aspects should be, but so far hardly have been, a topic of discussion in the Netherlands. In this article I will reach the conclusion that in the Netherlands, based on the worldwide discussion on mass surveillance and Big data data mining by the intelligence agencies, the time has come to also regulate Big Data data mining by the police. Regulation has to emerge through the democratic legislative process. I will formulate criteria or propositions for the implementation of this legislation in the Dutch Code of Criminal Procedure. These criteria may, in an international context, be useful for the broader debate about Big Data data mining by police agencies.

Keywords Criminal Law · Criminal Procedural Law · Big Data · Privacy

✉ S. Brinkhoff
s.brinkhoff@jur.ru.nl

[1] Criminal Law and Criminal Procedural Law, Faculty of Law, Radboud University Nijmegen, Nijmegen, The Netherlands

1 Introduction

The term *Big Data* refers to the phenomenon of an ever larger and increasingly complex number of digital data and data files that keep growing in scope continuously and exponentially.[1] It is a known fact that worldwide different intelligence agencies employ automated data analysis, known as data mining, on data and data files to understand Big Data and to subsequently target an individual citizen as, for instance, a potential terrorist. Those agencies collect data through the use of methods of mass surveillance. The Snowden affair in relation to the mass surveillance by the American National Security Agency (NSA) is a painful example thereof. More recently, it came to light that the British Government Communications Headquarters GCHQ uses the so-called "Data vacuum cleaner", known as the Karma Police. Using this Data vacuum cleaner, GCHQ intended to create a profile of *all* internet users in the world and to keep track of how they surf the internet for the purpose of data analysis. Apart from the Big Data data mining by intelligence agencies, more and more also the Dutch police uses automated data analysis of data and data files and real Big Data data mining as a method of investigation in criminal proceedings.[2] The automated data analysis by the police can be used (1) to bring up additional personalized data about an individual or individuals who were already labelled as a suspect of a criminal offence and (2) to gather personalized results about a possible suspect or group of suspects. The use of the iColumbo system is an example of Big Data data mining by the Dutch police.

Even though Big Data data mining can be a potentially useful and effective method of police investigation, there are some uneasy aspects associated with it. These aspects should be, but so far hardly have been, a topic of discussion in the Netherlands. In this article I will address these uneasy aspects. In this regard, it is first important to emphasise that the process of Big Data data mining, by definition, in most cases also affects data of innocent civilians, and therefore, interferes with the right to privacy of these civilians as protected under the first paragraph of article 8 of the European Convention on Human Rights (ECHR) and articles 7 and 8 of the Charter of fundamental rights of the European Union (the Charter). Second there are several uneasy aspects concerning the use of this method of investigation in criminal proceedings. The question, for instance, arises if there has to be a reasonable suspicion prior to the use of this method in a criminal proceeding. Furthermore, it is relevant which criteria the police is allowed to use for the automated data analysis. Can these criteria also focus on, for instance, data regarding gender, religion and political preference? The question also arises what the value is of the outcome of data mining in criminal proceedings. For example, can a civilian be arrested by the police solely on the basis of the results of automated data analysis? A final uneasy aspect is that until now it is unclear to what extent Big Data data mining by the

[1] Sietsma et al. (2002).

[2] Borking et al. (1998), Sietsma et al. (2002) and Sietsma (2006) en A.R. Lodder, N.S. van der Meulen, T.H.A. Wisman, L. Meij and C.M.M. Zwinkels, *Big Data, big consequenses. Een verkenning naar privacy en big data gebruik binnen de opsporing, vervolging en rechtspraak*, WODC-report 2014.

Dutch police is permitted, who should make the pertinent decision and how administering the supervision of this method of investigation should be shaped.

Many of the above mentioned aspects are related to the necessity for limitations to Big Data data mining by the police as a method of investigation. In this article I will reach the conclusion that in the Netherlands, based on the worldwide discussion on mass surveillance and Big Data data mining by the intelligence agencies, the time has come to also regulate Big Data data mining by the police. Regulation has to emerge through the democratic legislative process. I will formulate criteria for the implementation of this legislation in the Dutch Code of Criminal Procedure (CCP). These criteria may, in an international context, be useful for the broader debate about Big Data data mining as an investigative method by police agencies in criminal proceedings.

2 Big Data data mining and the Right to Privacy

As stated earlier, the use of Big Data data mining as an investigative method by the police may intervene with the right to privacy of (innocent) civilians. The right to privacy is protected under article 8 ECHR and article 7 and 8 of the Charter. Article 8 of the Charter focusses on the protection of personal data and states, among other things, that the processing of such data has to have a basis in law. Recently, the European Union issued specific rules for the protection of data, the General Data Protection Regulation and corresponding directive.[3] Under Article 8 ECHR an interference with the right to privacy has to be in accordance with the law, there has to be a legitimate aim for the interference and it has to be necessary in a democratic society.[4] So when the use of the investigative method of Big Data data mining by the Dutch police does in fact intervene with the right to privacy of (innocent) civilians this does not automatically mean that this method cannot be used as such. It *does* mean that this interference has to comply with the standards as set out in the Article 8 ECHR and Article 8 of the Charter.

Based on the abovementioned, the question arises when there is an actual interference with the right to privacy if the police use Big Data data mining as an investigative method in criminal proceedings. The European Court on Human Rights (ECtHR) in this regard has reiterated that private life is a broad term not susceptible to exhaustive definition. Article 8 ECHR is not limited to the protection of an "inner circle" in which the individual may live his own personal life as he chooses and to exclude therefrom entirely the outside world not encompassed within that circle. It also protects the right to establish and develop relationships with other human beings and the outside world. Private life may even include activities of a professional or business nature. There is, therefore, a zone of interaction of a person with others, even in a public context, which may fall within the scope of "private life".[5] Relevant in the context of Big Data data mining is that the ECtHR found that

[3] See the website eur-lex.europa.eu for the content of regulation 2016/679 concerning the General Data Protection and the accessory directive 2016/680.

[4] Harris et al. (2009).

[5] See for instance ECtHR December 16th 1992, appl.nr. 13710/88 (Niemietz v. Germany), ECtHR September 25th 2001, appl.nr. 44787/98 (P.G. and J.H. v. the United Kingdom).

public information, such as the open source Big Data that is available on the internet, can fall within the scope of private life were it is systematically collected and stored in files held by the authorities.[6] Because the protection of personal data is of fundamental importance to a person's right to privacy, it is also relevant that the ECtHR has consistently found the systematic collection and storing of data by security agencies on particular individuals constituted an interference with these persons' private lives, even if that data was collected in a public place or concerned exclusively the person's professional or public activities.[7] Based on the abovementioned, the assumption can be made that even if the investigative method of Big Data data mining is merely used to analyze open source-information to obtain personalized results about possible suspects this constitutes to an interference with the right to privacy which, under the second paragraph of article 8 ECHR, requires a specific basis in law. This law also has to be accessible and foreseeable.[8]

Under the ECtHR case-law, the expression "in accordance with the law" within the meaning of the second paragraph of Article 8 requires, first, that the measure should have some basis in domestic law; it also refers to the quality of the law in question, requiring it to be accessible to the person concerned, who must, moreover, be able to foresee its consequences for him.[9] The ECtHR reiterates in this connection that in the special context of secret measures of surveillance, the abovementioned requirements cannot mean that an individual should be able to foresee when the authorities are likely to resort to secret surveillance so that he can adapt his conduct accordingly. However, especially where a power vested in the executive is exercised in secret, the risks of arbitrariness are evident. The ECtHR, therefore, ruled that it is essential to have clear, detailed rules on the application of secret measures of surveillance, especially as the technology available for use is continually becoming more sophisticated. The law must be sufficiently clear in its terms to give citizens an adequate indication of the conditions and circumstances in which the authorities are empowered to resort to any measures of secret surveillance and collection of data. In addition, because of the lack of public scrutiny and the risk of abuse intrinsic to any system of secret surveillance, the following minimum safeguards should be set out in statute law to avoid abuses: the nature, scope and duration of the possible measures, the grounds required for ordering them, the authorities competent to permit, carry out and supervise them, and the kind of remedy provided by the national law.[10] Based on the abovementioned, the conclusion can be made that the case law of the ECtHR calls for a specific basis in law for the use of Big Data data mining as an investigative method in criminal proceedings.

[6] ECtHR May 4th 2000, appl.nr. 28341 (Rotaru v. Romania) and ECtHR April 28th 2003, appl.nr. 44647/98 (Peck v. the United Kingdom).

[7] ECtHR November 28th 2011, appl.nr. 30194 (Shimovolos v. Russia).

[8] Harris et al. (2009).

[9] See for instance ECtHR April 24th 1990, appl.nr. 11801/85 (Kruslin v. France).

[10] ECtHR Juli 1st 2008, appl.nr. 58243/00 (Liberty and Others v. the United Kingdom).

3 Big Data data mining by the Dutch Police

Aside from the concrete examples of data mining which will be discussed below, in general police practice shows that automated data analysis is being used as investigative method in criminal proceedings.[11] The circumstances for proceeding to use automated data analysis and even Big Data data mining in criminal proceedings are, from the perspective of the police and the Public Prosecution Service, becoming ever more favourable since they more often obtain data files containing information regarding civilians based on covenants and collaborative arrangements with other public authorities such as the Tax and Customs Administration.[12] Information freely available on the internet, the so-called open sources, obviously contributes to this more favourable climate. A situation in which the Dutch police have access to real Big Data becomes even more apparent. Such Big Data can subsequently be subjected to automated data analysis. The Articles 9, 10 and 11 of the Dutch Police Data Act (*de Wet Politiegegevens*), make it possible to subject data and data files to automated analysis.[13] The legal basis for Big Data data mining by the Dutch police is, in addition to the aforementioned provisions, mostly found in the general task-setting, and non-specific, Article 3 of the Dutch Police Act (*de Politiewet*). Article 3 states that the Dutch police is, among other things, burdened with the investigation of criminal offences. Case law of the Dutch Supreme Court (*de Hoge Raad*) shows that in criminal proceedings Article 3 of the Dutch Police Act can be used as a legal basis for investigative methods that are not specifically regulated by Dutch law, such as Big Data data mining, as long as the use of this method only leads to a limited interference with fundamental rights (such as the right to privacy) of civilians.[14]

The above described situation translates in a concrete example of the application of real Big Data data mining by the Dutch police; the use of the iColumbo system. This automated system, on the basis of certain keywords or profiles, views and analyses Big Data on the internet to get personalized results about (possible) criminal offences.[15] The iColumbo system in this respect, does not just look at actual data but also at past information.[16] iColumbo classifies the results of this data analysis in order of relevance. The rationale underlying the use of this system is to make the non-automated method of searching the internet redundant. The legal basis

[11] Y. Buruma, 'Opvragen, bewerken en kennisnemen van gegevens voor de opsporing', *DD* 2010-57 and J. Kurpershoek, 'Zeecontainers vol data doorzoeken', *Blauw* maart 2014, p. 22–25 about the cooperation between the police of Rotterdam and the expert centre Kecida of the Dutch Forensic Institute (*het Nederlands Forensisch Instituut*).

[12] See in this context the letter of the Dutch Minister of Justice dated December 13th 2007 and the attached 'Programma versterking aanpak georganiseerde misdaad', *Kamerstukken II* 2007–2008, 29 911, nr. 10.

[13] Y. Buruma, 'De informatiemaatschappij en het strafrecht', *DD* 2007-43.

[14] Hoge Raad December 19th 1995, *NJ* 1996, 249 and Hoge Raad July 1st 2014, *NJ* 2015, 114 en 115.

[15] Memorandum *Vrijheid en Veiligheid in de digitale samenleving*, *Kamerstukken II* 2013–2014, 26 643, no. 298 and Timan and Koops (2014), p. 284–290.

[16] M. Roessingh's article in "Trouw" newspaper of November 2nd 2013 entitled "*iColumbo kan meer dan hij mag'"*.

for using the iColumbo system is still unclear. Because of its general task-setting nature and lack of a specific legal provision Article 3 of the Dutch Police Act seems to be applied here.[17]

There are a number of aspects that focus the mind when looking at the application of iColumbo by the Dutch police. In the first place, it can be surmised that in the absence of an explicit legal provision, this investigative method can *even* be deployed *without* the element of a reasonable suspicion of a criminal offence. So in criminal proceedings iColumbo can be used to obtain personalized results about individual citizens who at that moment are not seen as a suspect. One can also argue that in the absence of a legal provision, no limitations have been put in place as to what *kind* of data can be looked at. Nor do there appear to be any limitations on the keywords or profiles that the Dutch police may apply. As such iColumbo can be easily deployed to gather sensitive information concerning ethnicity, religion, political beliefs or sexual orientation about individuals. Finally, and in conjunction with the above, it is remarkable that there is no formalized oversight of, or supervision over, the deployment of the system by, for example, a public prosecutor or an investigative judge.[18] In many ways, the aspects mentioned expand the potential scope of iColumbo. This is remarkable since the deployment of this system and the absence of statutory standards and frameworks can easily lead to an interference with the right to privacy of innocent civilians as protected under the first paragraph of article 8 of the ECHR. The discussed case law of the ECtHR has made that crystal clear.

Apart from the aforementioned example of real Big Data data mining, automated analysis of data has already for some time been used by the Dutch police as an investigative method in criminal proceedings. The first concrete example thereof is the use of the Automatic Number Plate Recognition system (ANPR-system). This system automatically stores the licence plates of passing cars on freeways. These data are then compared with a so-called reference file. Such a file can be filled with police data, such as files on car theft or drug trafficking. A "hit" between the licence plate and the information from the reference file may lead to the start of a criminal proceeding and the application of coercive measures, it may also be used in an ongoing criminal proceeding and it may even be used as evidence in a criminal procedure. Police practice shows that these license plate data are also saved in the case of a 'no hit'.[19] The objective then is to compare it at a later time with other police data. The legal basis for the use of the ANPR-system is found in the general task-setting Article 3 Dutch Police Act. However, legislation is being implemented, the proposed Article 126jj CCP, which specifically focuses on the use of this system for criminal proceedings.[20] What stands out with respect to this form of data mining is that, unlike the use of iColumbo, specific legislation is being created and that in this legislation the element of reasonable suspicion limits the use of this method of automated data analysis. The license plate data may in fact, under the Article 126jj

[17] Koops et al. (2012).

[18] Koops et al. (2012).

[19] Hoge Raad November 11th 2014, *NJ* 2015, 296.

[20] *Kamerstukken II* 2012–2013, 33 542, nr. 2 en 3 (MvT).

paragraph 3 under a CCP, only be compared with a police files if there is a reasonable suspicion to do so.

Another concrete example of the use of the method of data mining by the Dutch police is hidden in the work of the Financial Intelligence Unit of the Netherlands (FIU).[21] Companies and financial institutions pursuant to the Act on Prevention of Money Laundering and Financing of Terrorism (*Wet ter voorkoming van witwassen en financieren terrorisme*) are obliged to report unusual financial transactions to the FIU. The unusual transactions can, after analysing the information, be declared suspicious. In the annual review of the FIU it is stated that in the year 2014, 277.000 unusual transactions were reported.[22] Of these 277.000 unusual transactions 29.000 (about 10%) were declared suspect. The analysis of the unusual transactions can be done by automated data analysis. The second paragraph of Article 14 Act on Prevention of Money Laundering and Financing of Terrorism is the legal basis hereof. The transactions which are declared suspect may lead to the start of a criminal proceeding and the use of coercive measures and may also be used as evidence in a criminal proceeding. What focusses the mind with regard to this form of data mining by the Dutch police is that although there is a specific legal basis for the data mining by the FIU, this legislation contains no limitations. For example, the element of reasonable suspicion is not required prior to the use of this method and there is no formalized monitoring role for the public prosecutor. This is remarkable since this kind of data mining evidently intervenes with the right to privacy of non-suspected citizens on a large scale. The annual figures of FIU show crystal clear that 90% of the unusual transactions are not seen as suspicious *after* using the method of data mining by the FIU. Without substantial reason, therefore, the (financial) data of many innocent civilians are seen and analysed by the Dutch government.

In conclusion, one can state that for the existing forms of Big Data data mining by the Dutch police the legal framework is too limited, and therefore, is not in accordance with the second paragraph of Article 8 ECHR. This legal framework after all is not very specific, and therefore, it is unclear to what extent Big Data data mining is permitted, who should make the pertinent decision and how administering the supervision of this method of investigation should be shaped. The range, scope and application of automated data analysis by the Dutch police, therefore, seems limitless and the interference with the privacy of (innocent) civilians is a given fact. The annual figures of the FIU, for instance, show this in a crystal-clear fashion.

4 Big Data data mining by intelligence agencies

The intelligence agencies already use Big Data data analysis on a larger scale than the police. In recent years the nature and scope of this method of investigation has regularly led to huge public and political commotion. For instance, it came to light recently that the British intelligence service GCHQ uses the "Data vacuum cleaner", known as the Karma Police. For the purpose of doing data analysis

[21] Doorenbos (1997), Faber and van Nunen (2004) and Mout (1994), p. 968–970.

[22] See the website http://www.fiu-nederland.nl.

GCHQ, using this Data vacuum cleaner, intended to create a profile of *all* internet users in the world by keeping track of how they surf the internet. Within this framework it was furthermore revealed that GCHQ, every day, stores billions of data concerning civilians.[23] The Snowden affair also needs to be mentioned here.[24] This case concerns the revelations by Snowden with regard to automated and very large-scale gathering of information, also known as mass surveillance, by the American NSA, of foreign telecommunication providers with a view to enable data analysis. The NSA had unlimited access to internet and mobile phone communications of European civilians.[25] In 2014, in the aftermath of this case, the actions of the Dutch intelligence services AIVD and MIVD and the minister responsible also became a topic of discussion.[26] This discussion focused on the providing of meta data by these Dutch agencies to the NSA.

The Snowden case and the revelations about GCHQ have been a catalyst for the discussion on mass surveillance with a view to Big Data data mining. Central to this discussion is the tension between the protection of the right to privacy as set out in the first paragraph of Article 8 ECHR and the danger of abuse. Limiting this kind of data analysis, by formulating the requirement to proceed only when the element of a reasonable suspicion is present, has also been suggested in the discussion.[27]

The abovementioned discussion has also lead to a pending case before the ECtHR.[28] The applicants in this case allege that they are likely to have been the subject of generic surveillance by GCHQ and that there may have been an interception of material relating to their electronic communications. They contend that the resulting interference with their right to privacy under Article 8 ECHR was not "in accordance with the law". In their submission, there is no basis in domestic law for the receipt of information from foreign intelligence agencies and an absence of legislative control and safeguards in relation to the circumstances in which the GCHQ can request foreign intelligence agencies to intercept communications and to access to stored data that has been obtained by interception, and the extent to which GCHQ can use, analyze, disseminate and store data solicited and/or received from foreign intelligence agencies and the process by which such data must be destroyed. The applicants also contend that the generic interception of external communications by GCHQ, merely on the basis that such communications have been transmitted by transatlantic fibre-optic cables, is an inherently disproportionate interference with the private lives of thousands, perhaps millions, of people.

On the periphery of the discussion on mass surveillance with regard to Big Data data mining the effectiveness of collecting and analysing Big Data by intelligence

[23] Also refer to the R. Gallagher web article The Intercept van 25 September 2015 article entitled "*profiled: from radio to porn, British spies track web users' online identities*".

[24] Brown and Korff (2014), p. 243–251.

[25] Refer to various articles on this matter on the site of The New York Times (http://www.nytimes.com).

[26] Refer to *Kamerstukken II* 2013–2014, 30 977, no. 80.

[27] E.g. refer to the Omtzigt Report to the European Council on *mass surveillance* on http://www.coe.int. Also refer to the Minister of Home Affairs' reaction to this Report, *Kamerstukken II* 2014–2015, 30 977, no. 110.

[28] The case of Big Brother Watch and Others v. the United Kingdom, application nr. 58170/13.

services also came up. The New York Times reports that, based on research by the American government, the collection of Big Data and the ensuing application of automated data analysis by the NSA has hardly produced any results for tackling terrorism; previously unknown data or suspects have *not* come into view.[29]

With respect to Big Data data mining, the discussion in the Netherlands on mass surveillance received an extra boost from the legislative procedure to amend the Intelligence and Securities Act 2002 (Wiv).[30] In the draft legislative proposal, the current authority for wire/phone tapping as laid out in art. 25 Wiv is expanded to the extent that non-specific tapping of *any* kind of telecommunications should be allowed. This will open the door for even more mass surveillance, because articles 33, 34 and 35 of the draft legislative proposal do allow for a phased-in approach by: (1) non-specific gathering of any kind of telecommunications, (2) preparation, e.g. by establishing the characteristics, type and identity of the telecommunications user and (3) further processing of any kind of telecommunications, e.g. by subjecting it to an automated data analysis. These powers are subject to prior ministerial approval, supervised by the Intelligence and Security Services Supervisory Committee (CTIVD).

The draft legislative proposal also provides for valuable information with regard to Big Data data mining by the police. Paragraphs 3 and 4 of article 18 set out clear limitations to the kind of data that may be processed by the police. The starting point is that, unless unavoidable, personal data pertaining to someone's religion, race, health and sexual orientation will *not* be processed. Presently, Big Data data mining by the Dutch police does not contain such restrictions. Furthermore, the proposed article 47 of the draft legislation illustrates the powers that the Dutch intelligence agencies have in regard to Big Data data mining. It states that, among other things, this method of investigation can be applied to data from their own data files, from open sources and from data obtained from third party data files. Paragraph 2 sets out to what extent data mining is authorized. These data files can be put side by side for comparison purpose, can be searched on the basis of profiles and can be compared with a view to detecting certain patterns. Paragraph 3 of this article is interesting as it concerns data analysis on the basis of profile searching. It states that the intelligence services are *not* permitted to act purely on the basis of the results of this kind of data mining. It points out that in reference hereto, *due consideration* still has to be given as to whether the services can take action.

The example of mass surveillance in view of Big Data data mining by intelligence services already shows full scale application of this method of investigation. At the same time, it shows that its nature and scope can lead to social and political commotion. An important last conclusion, as can be seen in the draft legislative proposal for the new Wiv, is that the Dutch legislature, in this context, elaborates in more detail when and under which conditions data mining can be used and where the limitations lie. These conclusions in conjunction with the conclusions

[29] Refer to the New York Times article of 23 January 2014 entitled "*Watchdog report says NSA Program is illegal and should end*".

[30] Its origin is based on the outcome of the Dessens Committee Report concerning the current Wiv. Refer to Appendix to *Parliamentary Papers II* 2013–2014, 33820, no. 1 (Wiv evaluation: On the way to a new balance between powers and safeguards).

relating to the existing kinds of Big Data data mining by the Dutch police, give rise to further exploration of the criteria for future kinds of Big Data data mining by the police.

5 Criteria for Big Data data mining as a future method of investigation

Presently, automated data analysis and real Big Data data mining is already being applied by the Dutch police in criminal proceedings and, as discussed, the use of this investigative method does not seem to be in accordance with the second paragraph of Article 8 ECHR. There is hardly any doubt that, in the future, this method of investigation will be utilised on a larger scale and that the results will play a more important role in criminal proceedings. This will not be any different in other countries. If we link this to the observation that the existing kinds of Big Data data mining by the Dutch police hardly seem to have any statutory limitations, an ominous picture of the future emerges, a picture in which the prosecution authorities quite simply and in a variety of ways can apply data analysis to real Big Data. The clear conclusion, therefore, is that a discussion about this method of investigation is overdue. I am proposing four criteria as starting points for this discussion of present and future kinds of Big Data data mining by the Dutch police as an investigative method in criminal proceedings. The criteria are that: (1) this method of investigation requires an explicit statutory basis in the Dutch Criminal Code of Procedure, (2) this method can only be used upon the presence of the element of reasonable suspicion, (3) the public prosecutor and definitely the judge play a role when automated data analysis potentially has a serious impact on the privacy of innocent civilians and (4) the police be prevented from acting purely on the basis of the results of automated data analysis.

The first criterion is that new and specific legislation is needed for this method of investigation in criminal proceedings; creating an explicit legal provision in a democratic legislative process is paramount. Basing this method on the general task as set out in article 3 of the Police Act is insufficient, since that legislative process has not taken place. In the legislative process the nature and scope of present and future applications of Big Data data mining by the police need to be discussed *and* attention has to be paid to the regulation and monitoring of this method. There are at least two reasons why legislation has to be made. The first reason is that the method of Big Data data mining on data or data files which are collected or held by the police can be seen as in interference to the right to privacy as stated in the first paragraph of article 8 ECHR. Thus, according to the second paragraph of article 8 ECHR this interference must have a specific basis in law. The fact that Big Data data mining can violate the privacy of large groups of innocent civilians and that its scope can even include sensitive information, makes the need for a legislative process even more urgent, especially in light of this human right. This law has to be precise on the conditions under which the police may use this method. As a result it is foreseeable for civilians in what kind of situations the Dutch police may use this method. The general task setting article 3 of the Dutch Police Act has none of the above mentioned elements. The second reason why legislation has to be made by

the Dutch legislator is because of the huge public and political commotion that arose after the revelations of NSA whistle-blower Snowden. This commotion calls for a firm democratic (parliamentary) debate about the nature and scope of present and future applications of Big Data data mining by the police.

The second criterion is that this investigative method in criminal proceedings can only be used upon the presence of the element of reasonable suspicion. The element of suspicion is a condition for applying this method in criminal proceedings and defines its scope and timing; automated data analysis can only be initiated when reasonable grounds for suspicion arise based on facts and circumstances. A reason for this limitation in particular arises from the example of *mass surveillance* with a view to Big Data data mining by the intelligence agencies. This has made it obvious that collecting large quantities of data *without* the presence of the element of reasonable suspicion, for the purpose of automated data analysis, has led to considerable social resistance, because it is undesirable that large quantities of privacy-sensitive information of innocent civilians get into the hands of intelligence agencies, and therefore, the government, without a concrete reason. Such social resistance should be taken seriously and should be reflected in the application of Big Data data mining by the police. To add the limitation of the element of reasonable suspicion is also more in line with the criminal procedural system in which methods of investigation in criminal proceedings can only be used with the presence of a reasonable presumption of guilt.

The third criterion is that the law needs to outline how Big Data data mining can be applied. Without legal provisions this automated data analysis should not be extended to searches for or links with sensitive information concerning ethnicity, sexual orientation, political preference or religious beliefs. In view of the privacy protection under the first paragraph of article 8 ECHR its use should be restrained. A way to formalise this restraint can be to require prior authorisation by the public prosecutor and in some cases a judge for the most serious forms of Big Data data mining. The wording in the Dutch Wiv draft legislative proposal does formulate restraint in respect of sensitive information. The judge's role is important when, in a specific situation, the purpose of automated data analysis is to widen the scope to sensitive information, for instance by a proposed search or by profiling.

The fourth and last criterion concerns the value of Big Data data mining in the framework of investigating, prosecuting and adjudication of criminal acts. The question here is whether, in an actual case, the result justifies the use of a coercive measure or authority to investigate or whether it can even be considered in the building of evidence? It is clear that this question arises for every kind of information, but the results of Big Data data mining carry specific risks. In the first place, it is possible that the data files that are subjected to this automated data analysis are false or contain outdated information. Commencing criminal proceedings could then lead to the arrest of an innocent civilian whose home could, for instance, be wrongfully searched and the ultimate consequence could even be the sentencing of an innocent person based on the faulty outcome of automated data analysis. Furthermore, the outcome of Big Data data mining explicitly poses the question: "What value is to be attributed to an observed correlation?" For instance in an actual case the question can arise whether digitally available information

concerning frequent visits to a mosque or a synagogue combined with internet searches into the conflict in the Middle East justifies doing a house search. Or whether an additional requirement is, that the person in question has non-Dutch nationality, has a criminal record and is in possession of an airline ticket to Turkey? In short, which correlation has to be present to commence criminal proceedings? Commencing criminal proceedings based on a correlation that turns out not to show any criminal behaviour can already have the above mentioned undesirable effects. By formulating as a starting point the requirement that the actors in the criminal proceedings such as the police, the public prosecutor and the examining judge *always* cast a critical eye before the outcome of data mining can be used, these undesirable effects can be largely limited. In this assessment, they can consider that the outcome of Big Data data mining needs to be based on *actual* information leading to a *concrete* presumption of criminal behaviour. What is relevant here is based on which data files this outcome is arrived at. Casting a critical eye is not limited to the outcome of automated data analysis, but also on how the outcome is arrived at, in other words: what is behind it. Furthermore, it is preferable to have the outcome of the Big Data data mining process corroborated, as much as possible, by the outcome of other methods of investigation or by information available to the police. With regard to this criterion I refer to an element of the Wiv draft legislative proposal, as discussed earlier. There I state that in some instances, the intelligence service is not allowed to act solely on the basis of the outcome of Big Data data mining; due consideration is necessary prior to making the assessment. This must also apply to the commencement of criminal proceedings based on the outcome of Big Data data mining by the police. This requirement can be provided for through legislation, but this rule of law can also be established by case law.

6 Conclusion

In this article I have shown that already data mining is used as a method of investigation by the Dutch Police in criminal proceedings and that real Big Data data mining is on the rise. The examples of the NSA and GCHQ show a similar trend in the context of the work of security agencies in other countries. This development calls for further discussion on the nature, scope, regulation and limitations to the use of this method of investigation by the Dutch police in criminal proceedings. There has hardly been any such discussion in the Netherlands up until now. I advocate for having this discussion, not just in the Netherlands but also in other countries and that we should learn from earlier discussions regarding the powers of the intelligence agencies. It is time to start a democratic legislative process to deal with the previously mentioned issues. The criteria to feed this discussion are that this method of investigation in criminal proceedings can only be used when the element of reasonable suspicion is present, that no criminal proceedings should be initiated based solely on the outcome of automated data analysis and that the public prosecutor and certainly the judge need to play a role when the privacy of innocent civilians can be severely impacted by an automated data analysis.

References

Borking J, Artz M, van Almelo L (1998) Gouden bergen van gegevens. Over datawarehousing, datamining en privacy. Registratiekamer, Den Haag

Brown I, Korff D (2014) Foreign surveillance: law and practice in a global digital environment. Eur Hum Rights Law Rev 2014(3):243–251

Doorenbos DR (1997) Witwassen en voordeelsontneming. W.E.J Tjeenk Willink, Deventer

Faber W, van Nunen AAA (2004) Uit onverdachte bron. Evaluatie van de keten ongebruikelijke transacties. Boom Juridische uitgevers, Den Haag

Harris DJ, O'Boyle M, Bates EP, Buckley CM (2009) Harris, O'Boyle & Warbrick: law of the european convention on human rights. Oxford University Press, Oxford

Koops EJ et al (2012) Juridische scan open brononderzoek. Een analyse op hoofdlijnen van de juridische aspecten van de iRN/iColumbo-infrastructuur en HDIeF-tools. TILT, Tilburg

Mout CCh (1994) Gebruikelijk of ongebruikelijk: een bloemlezing van vragen rond de wet MOT. Advocatenblad, 1994, pp 968–970

Sietsma R (2006) Gegevensverwerking in het kader van de opsporing. Toepassing van datamining ten behoeve van de opsporingstaak: afweging tussen het opsporingsbelang en het recht op privacy. Sdu Uitgevers, Den Haag

Sietsma R, Verbeek J, van den Herik J (2002) Datamining en opsporing. Toepassing van datamining ten behoeve van de opsporingstaak: strafprocesrecht versus recht op privacy. Sdu Uitgevers, Den Haag

Timan T, Koops EJ (2014) Sociale media en surveillance: over verschuivende rollen en vervagende grenzen. Strafblad 2014(2):284–290