## Protecting individuals against the negative impact of big data

*The potential and limitations of the privacy and data protection law approach*

Oostveen, M.A.A.

**Publication date**
2018
**Document Version**
Other version
**License**
Other

# ENGLISH SUMMARY

## CHAPTER 1: INTRODUCTION

This thesis is about the protection of individuals against the negative impact that big data may have on their private lives. Many positive and promising developments result from big data, but the massive collection and use of data also raise a host of issues. At the centre of this thesis is the position and protection of the individual whose personal data are used in big data, or who experiences the negative consequences of the application of big data.

In the European Union (EU), the rights to privacy and to data protection are the focal points in the discussion on the negative sides of big data. Privacy and data protection are perceived as being the primary rights at risk, as well as the (legal) solution to the problems that big data creates. This thesis researches the privacy and data protection approach for the protection of individuals. It focuses on the question what the potential and the limitations of the EU legal framework on privacy and data protection are with respect to protecting individuals' rights and freedoms against the negative impact of big data.

Chapter one introduces the subject and research questions. It discusses the limitation to the EU, the exclusion of the criminal law domain, and the selection of individual rights and freedoms based on the European Convention on Human Rights (ECHR) and the EU Charter of Fundamental Rights (Charter), elucidated in Chapter 2. In addition it contains a concise conceptual analysis of these individual rights and freedoms and of the definition of big data, as well as the methodological justification of the research.

## CHAPTER 2: BIG DATA

Chapter 2 discusses the concept of big data and the resulting possible negative consequences for the rights and freedoms of the individual. The meaning of the term "*big data"* depends on the context in which it is used. It is commonly assumed that the term was first used in 2011, linked to three factors that refer to data, being *volume, velocity,* and *variety.* These factors were introduced as characteristics of new possibilities for data management and analysis, that were considered a result of technological developments that increased these possibilities and made them cheaper. However, over the years the term has been used in various new ways, ranging from a marketing term that refers to the mere analysis of a large amount of data and is used as a means to attract attention and generate income, to a generic reference to a socio-technological phenomenon with far-reaching consequences for the way in which we regard knowledge and societal developments.

Considering the diversity in interpretations, a further delineation of big data is required. After all, when the content of the term and its consequences are not clear, it is impossible to determine how it is regulated and how it should be regulated. Therefore a process-oriented logic is applied, which facilitates a normative and legal analysis of big data. The common denominator of the different interpretations that are in use is that in all cases big data can be seen as a process in which data are collected and analysed, with the aim of subsequently applying the results of this analysis. In practice, it is obviously a complex and iterative process, but in general these three phases of acquisition, analysis, and application can be distinguished in big data projects. This three-phase model makes it possible to conduct a normative and legal analysis of big data. In Chapter 2 the different phases are elucidated and illustrated using three practical examples of big data: credit in the financial services industry, biobanks in medical research, and online personalisation. In brief, the phases consist of the following steps. In the acquisition phase, data are collected. They can be acquired from the individual herself, but they can also be bought from data brokers, or for example created through combining different data sets. Thereafter follows the analysis of the data. The analysis is automated, using specialised software programs that are continuously being developed. From this analysis flows information in different forms, for example as knowledge, models, or predictions. In the application phase, this information is used as a basis for decisions. These can be general decisions, for example about building infrastructures or prescribing new (combinations of) drugs for specific diseases. But often it concerns decisions that are targeted at the individual, such as the rejection of a credit application or the adjustment of the prices of flights on a website. For the normative and legal analysis, it is important to give thought to the fact that the people whose data are gathered and serve as input for analysis, and the people to whom the results of analysis are applied, are two different groups. Depending on the big data project at stake they can overlap, where data from the application phase lead to new input. But in general, the acquisition and analysis are separated from the application; a model is developed using the data of people from the first group, and can subsequently be applied to an unlimited new group of people.

After the explanation of the three-phase model, follows an analysis of the possible negative consequences on individual rights and freedoms that can result from each phase. Notwithstanding the positive influence that big data can have on the lives of individuals and society as a whole, which is not denied in this thesis, given the research question most attention is paid to the negative aspects of big data. The conclusion is that every phase can negatively influence individual rights and freedoms in its own way, because different actions take place in each of the phases. In the acquisition phase, the risks primarily lie in the large scale collection and combination of (personal) data. This has negative effects on the rights to privacy and to data protection, but personal autonomy and freedom of expression are at stake as well. The knowledge that one's behaviour is monitored and that data on behaviour is tracked and saved for future use can influence people's behaviour, the choices they make, and the information that they gather and produce. In the analysis phase the effects seem more limited, because often data will be processed in a form that makes it impossible to trace them back to identifiable individuals, as will be explained in more detail below. But the analysis phase is the source of many of the negative effects that crystallise

in the application phase, such as discrimination. As a rule, nether the people whose data are the source of analysis, nor the individuals that experience the negative consequences in the application phase, can exert any influence over what happens in the analysis phase. Essentially, there can be said to be a negative effect on personal autonomy: there is not much influence over what is done with data, or which data and models underpin a decision. In the application phase the full range of negative effects may materialise, that is, negative impact on the rights to personal autonomy, privacy and data protection, non-discrimination, and freedom of speech. Personal autonomy is under pressure, because of the possibilities to personalise, persuade, coerce, nudge, and manipulate that big data creates. It must be borne in mind that it is not about *high-impact* decisions only; small but cumulative instances of personalisation can also have a lasting influence on the individual's free choice, development, and identity. Given the increasing knowledge and possibilities, and the scale at which it takes place, particularly in the digital environment, it is important to also heed this cumulative personalisation, as well as the long-term consequences that it can have for individuals' futures. With respect to privacy and data protection in the application phase, roughly the same problems occur as in the acquisition phase. In addition, the application of big data can result in self-standing interferences with the personal sphere, independent of the processing of personal data. Personalisation can also negatively affect the free gathering of information and ideas. And finally, the application of big data can lead to (direct or indirect) discrimination.

### CHAPTER 3: THE FUNDAMENTAL RIGHTS TO PRIVACY AND TO DATA PROTECTION

Central to Chapter 3 is the scope of the fundamental rights to privacy and to data protection in the EU, to determine their normative content with regard to big data. The chapter examines what the fundamental rights involve in the context of big data, and what they demand of the protection that secondary EU law offers. The chapter focuses on the stand-alone value of these rights, as well as on the protective function that they fulfil with respect to other individual rights and freedoms, which is referred to as the *enabling function* of the rights to privacy and to data protection in this thesis.

Article 8 ECHR and Articles 7 and 8 CFREU are compared, being part of the most important fundamental rights instruments in the EU. Complaints about violations of the ECHR and CFREU can be brought before the ECtHR and the CJEU. The ECtHR has, in its long tradition of case law on Article 8 ECHR, to a large extent shaped the content of the right to (informational) privacy in the EU. The interpretation of the rights to privacy and to data protection by the CFREU is strongly influenced by the Strasbourg case law, mainly because the right to privacy in the Charter is a *corresponding right* in the sense of Article 52 (3) CFREU that has to offer minimum protection in accordance with Article 8 ECHR. The Charter adds a separate right to the protection of personal data to the EU fundamental rights level.

The interpretation of these fundamental rights in the light of big data is difficult for multiple reasons. The judgments of both Courts are casuistic and strongly embedded in the context of the cases at hand, which usually deal with interferences (through actions or neglecting to act) by states. Concrete cases on big data and explicit references to the enabling function are uncommon. On the other hand, through relinquishing the procedural context and instead looking at the normative content of the rights, in combination with the interpretative doctrines of the Courts such as the *living instrument* and *practical and effective* doctrines of the ECtHR, meaning can be given to the scope of the rights and to the level of protection that is required in secondary legislation.

The successive analysis of Articles 8 ECHR and 7 and 8 CFREU in the third chapter leads to a relatively coherent image of the normative concepts of the rights to privacy and data protection with respect to big data as interpreted by the ECtHR and the CJEU. The right to privacy encompasses personal data processing in all phases, for which the interpretation is linked to Convention 108 and includes metadata and location data. However, personal data processing does not automatically lead to an interference; it depends on circumstances such as sensitivity of the data, reasonable expectation of privacy, and the scope of the processing. In any case there are minimum requirements that need to be met to protect the individual. These requirements include limitations on storage and retention times, and obligations to take measures to protect the personal data, amongst others to prevent unauthorised access and abuse. Independent of personal data processing, the right to privacy is at stake when there is an interference with the home or correspondence, or in the sphere of sexual activities, social life, personal relationships, or personal, moral, or physical identity. This can be the case when the application of big data results in a decision that interferes with these interests. Given the diversity in possible applications of big data, it is impossible to give a detailed comprehensive overview of such cases.

The scope of the normative concept of privacy with respect to personal data seems very broad, but the right to data protection is still of added value. It adds, amongst others, specific requirements related to the processing of personal data. Article 8 CFREU is more precisely formulated than Articles 8 ECHR and 7 CFREU; it contains detailed requirements that the processing of personal data is based on consent or another legitimate ground, rights on access to data and rectification, and the necessity of having independent oversight on compliance.

The enabling function of privacy and data protection is acknowledged in the literature, but there are not many explicit references to be found in the case law of either Court that discuss the rights to privacy and to data protection. Personal autonomy is regarded by the ECtHR as part of the right to privacy, and with that the facilitating function of privacy for personal autonomy is inherent in the normative concept of the right to privacy. However, because an interference with personal autonomy has never been explicitly established by the ECtHR, its exact scope is undetermined. Regarding freedom of expression, the CJEU has made the connection in its case law and has thereby acknowledged the facilitating function, but this enabling function is not formulated as a duty of the right to privacy or data protection. The limited attention for the enabling function is likely due to the fundamental rights tradition in the EU: the rights to privacy and to data protection have stand-alone value, and other rights and

freedoms are protected by their respective fundamental rights and freedoms, which are generally analysed separately when they are also part of a case. These conclusions on the normative scope of the rights to privacy and to data protection reappear in Chapter 4, where they are compared to the conclusions on the scope of EU data protection law, to assess its duty to protect individual rights and freedoms.

### CHAPTER 4: DATA PROTECTION LAW

Chapter 4 analyses to what extent current EU data protection law, in the guise of the GDPR, protects individual rights and freedoms, can protect them, and is required to protect them. To answer these questions, this chapter is divided into four parts, which discuss whether EU data protection law applies to big data, how its substantive norms function in this context, how it is enforced, and whether the normative concepts of the rights to privacy and to data protection demand a different implementation of secondary legislation.

The first part of the chapter analyses when the regulatory framework applies within the big data process. Decisive for this question is the definition of "*personal data"* and the associated concept of "*identifiability".* The material scope of EU data protection law depends on whether personal data are processed, and therefore on the criteria that make up the concept of "*personal data*". In the context of big data, the crucial criterion is "*identified or identifiable",* which refers to whether data can be traced back to natural persons. On the basis of this criterion, four types of data can be distinguished: directly identifiable data, indirectly identifiable data, de-identified data, and non-personal data. The first two types are within the scope of the GDPR, and are here referred to as *identifiable data.* The other two types are not within the scope of the law, and are referred to as *non-identifiable* data. On the basis of this division, multiple processing scenarios are possible. When only non-identifiable data are processed in a given big data phase, the regulation does not apply, and accordingly there is no protective effect emanating from EU data protection law. Whether and in which phase identifiable data are processed depends on the specific big data project. There are projects in which personal data are processed in all three phases, which makes the regulation applicable without reservations. But there are also projects that do not collect and use personal data in the acquisition and analysis phases, but that can nevertheless have an impact on the individual in the application phase. An example is when a government allows for the building of a chemical plant in a specific location on the basis of big data analysis of non-identifiable data, and the health of people living in that area is at stake due to pollution or chemical waste. Generally, it will be the case that personal data processing often, but not necessarily, occurs in the acquisition phase and (to a limited extent) in the application phase, but not in the analysis phase. After all, there are strong incentives for entities engaging in big data to process data in a non-identifiable manner. In big data (analysis) there is generally no interest in specific individuals: general patterns are what matters. There is usually no necessity to keep data in an identifiable state. De-identification ensures that the GDPR is no longer applicable. If data protection rules are not applicable, principles such as purpose limitation, data minimisation, and

the necessity of having a legitimate processing ground, do not apply. This increases the possibilities for big data, while decreasing the risk with respect to, inter alia, compliance, which explains this strong incentive. The conclusion is that EU data protection law does not apply to (parts of) the big data process in some cases, and consequently big data partially avoids protection through the GDPR. This constitutes the first lacuna in the protection of individual rights and freedoms.

The second part of Chapter 4 analyses how the substantive norms of the GDPR function in the context of big data, if EU data protection law is applicable. This analysis focuses on rules that are of relevance in big data, given the actions in the different phases and the possible negative effects in every phase, with additional attention for the potential of innovations in data protection law resulting from the GDPR, like the right to data portability. The analysis of the potential and the limitations of substantive data protection law is divided into rules that should make data processing *transparent* for the individual, rules that aim to give the individual *control* over the processing, and rules that intend to map out or regulate specific *risks.* The duties to inform individuals about the aims and the way in which data are processed, and the logic and consequences of automated individual decisions (including profiling), constitute the rules that shape transparency. These rules address the acquisition phase, and the application phase in particular when there is automated individual decision-making. Control rights that are of importance are the requirement for a legitimate ground for processing and the accompanying consent requirement, the right not to be subjected to automated individual decision-making, the right to erasure (right to be forgotten), and the right to data portability. Together the rules on transparency and control create the necessary preconditions for the individual's informational self-determination: enabling awareness about, and insight into, personal data processing and being able to influence it. This informational self-determination safeguards personal autonomy and the right to data protection, and can (indirectly) mitigate discrimination and have protective effects on the general rights to privacy and to freedom of expression. However, in the context of big data this protection does not reach its full potential. One of the reasons for this is that the scope of the rights is limited due to their criteria for application. For example, the automated decision-making rules are limited to *high-impact* decisions. Profiling and personalisation can alter people's choices and lives in small but cumulative ways, which is not addressed by these rules. Moreover, accurate predictions do not always require much personal data from the person to whom big data is applied. Although designed with big data in mind, these provisions are to a certain extent already outpaced by technological developments. Additionally, in practice it is often difficult to meet the criteria that exist with respect to informing the individual and acquiring her consent, for example because future applications and consequences are unknown or insufficiently clear at the moment of personal data acquisition. But the most important reason is that individual rights and control do not function well in an environment of online data collection in a data-driven economy, and application of big data. Given the complexity and ubiquitousness of digital personal data processing, the management of online privacy and data protection has become an almost impossible task for the individual. Moreover, the supply of personal data frequently functions as a requirement and counter performance for the provision of (online) services, whereas the (negative)

consequences of personal data processing are often of an abstract nature, or only substantiate in the long run or within the context of cumulative processing instances or decisions. For individuals, it may seem as if they are not directly affected by the processing of their personal data. These are additional reasons to assume that adequate protection of individual rights and freedoms in the context of big data cannot be achieved through transparency and control by the individual alone.

The rules of the GDPR that are directed at mapping or regulating specific risks primarily address the entities that process personal data or are responsible for the processing. As such, the rules largely avoid the drawbacks associated with the aforementioned rules of the informational self-determination sphere. Important in the context of big data are the rules that (conditionally) prohibit the processing of special categories of data, the principles of purpose limitation and data minimisation, the obligations regarding data protection impact assessments, and the rules on implementing data protection by design and default. The limited prohibition on the processing of personal data that are perceived sensitive fulfils an important function in the protection of the rights to data protection and to privacy. It can also function as a means to mitigate discrimination, through limiting the processing of data related to grounds for discrimination such as ethnicity, sex, or religion. Yet it is of no avail in cases of indirect discrimination, which is a genuine danger in big data because certain (combinations) of data can serve as *proxies* for special categories of personal data. Data protection impact assessments and data protection by design and default demand, in brief, that big data entities map their processing activities and the associated risks, and take (technical) measures to try to safeguard individuals' data protection rights. The obligation to assess risks and possible safeguards in order to protect the rights of individuals prior to the processing of their personal data in all likelihood has a positive influence on the protection of the individual. But there are doubts about the practical implementation and enforcement of such rules, and as such about the protective potential of these provisions. The same holds for the provisions on purpose limitation and data minimisation. In conclusion, these rules seem to impose important limits on the processing of personal data on the one hand, but appear irreconcilable with big data, which can ultimately lead to enforcement issues. The GDPR improves enforcement possibilities, amongst others through substantially higher fines. However, due to inter alia the ubiquitousness of personal data processing and the continuing friction between law and practice, enforcement of data protection law will remain a problem in big data to some degree.

The last part of Chapter 4 explains the lacunae that exist in protection of individual rights and freedoms in big data when analysing the GDPR. It matches these lacunae with the normative concepts of privacy and data protection from Chapter 3, to draw conclusions on the potential and limitations of the data protection law approach to issues in big data, and assess to what extent data protection law *should* address these effects.

The lacunae in protection are divided into three types, according to the causes that lie at their foundation: whether data protection law is applicable, and if it is not, whether data processing is the core of the problem or not. In the first type of lacuna, the EU privacy and data protection law framework applies, but it offers insufficient protection.

This cause is at the root of most of the acquisition phase problems. It can also be at stake in the application phase, for example if people are targeted on the basis of personal data, but the automated decision-making rules do not apply because the decisions are not high-impact, or there is human intervention in the process. There are multiple reasons underlying this lacuna: as explained above, control and transparency do not function well in the context of the data-driven economy in general, and small but cumulative big data applications in particular. Second, the criteria that apply to the GDPR's rights, for example those on automated decision-making as explained above, do not yield broad protection in big data.

The second type of lacuna exists when data protection does not apply, yet data processing is at the heart of the issue. This frequently occurs in the analysis phase, as most personal data will have been de-identified before they enter analysis, and these data are therefore outside the scope of data protection law, but the risks lie in the processing of the data. In such cases, the means to exert influence over the analysis through data protection law are virtually non-existent, and there is no task for data protection law as no personal data are processed. In the third type, the framework is not applicable, and data processing is *not* the essence of the problem. This situation occurs in the application phase where, in addition to targeted personal data-based decisions, non-targeted decisions are made. These decisions do not rely on personal data for application, but can affect individuals' lives nonetheless, as explained above.

On the basis of the summary of how data protection law functions in the context of big data and the taxonomy of lacunae in protection, conclusions can be drawn about the normative concepts of the rights to privacy and to data protection, and how EU data protection law corresponds with them. As elucidated above, in the insufficient protection lacuna, data protection law applies but is not entirely up to the challenge of big data. EU data protection law does, however, seem to meet the level of protection that is required under the normative concepts of the rights to privacy and data protection law. In effect, the fundamental rights level demands a measure of informational self-determination, i.e. through the rights bestowed upon the individual under Article 8 (2) CFREU, and obligations regarding safeguards to protect personal data to be taken by those processing the data. Data protection law contains such rights and provisions, even though there are issues with their scope of application and protection in the context of big data.

A significant number of the big data problems, particularly (part of) those occurring in the analysis and application phases, does not fall within the scope of the normative concepts of privacy and data protection. If this is the case, there is no place for an enabling effect. Accordingly, these are stand-alone problems. For type 2 and 3 lacunae, particularly if data processing is not the core of the issue, there is no clear link between a problem and the rights to privacy and to data protection. In such cases, the biggest problems seem to reside in information asymmetries and power imbalances that data-driven decision-making generates, and the effects of certain decisions that may harm individuals. This can even be the case if the data protection law applies. An example is problems in the acquisition phase, where asymmetrical relationships involving, for example, network effects can be regarded as an important

cause. Therefore, it is key to acknowledge that not every problem of big data constitutes a privacy or data protection law problem. Data protection law should not be regarded as a *panacea* in the context of big data. The problems need to be assessed and addressed based on their individual characteristics. This observation strongly influences the preferred approach of big data as maintained in this thesis.

### CHAPTERS 5 AND 6: CONCLUSIONS AND FUTURE

Big data requires a combined approach. Privacy and data protection are of great importance, but big data does not concern the rights to privacy and data protection alone. Big data is a complex process that consists of different actions, summarised here as phases, that can have (negative) consequences in different ways, for equally different individual rights and freedoms. For a large part a combined approach simply requires that the big data process and its negative effects are not reduced to privacy and data protection problems, but instead are evaluated on their own characteristics. For example, when a person has a weak position vis-à-vis a company that provides an important service, and the service is made conditional on collecting her personal data, she is not only a data subject, but also a consumer. This does not always require precise coordination, because it concerns separate areas of law and enforcement issues do not necessarily overlap. Researchers, policy makers, and enforcement agencies should assess issues with an open mind; cases or problems should not be categorised as privacy and data protection problems as soon as personal data are at stake. Fortunately, such ideas are gaining ground in the literature and amongst policy makers and enforcement authorities, at EU level and at the national level.

In addition to explaining the necessity of a combined approach to big data, Chapter 5 discusses possible legal alternatives that might be part of such an approach. These alternatives are selected on the basis of the problems identified in Chapter 2 and the issues and lacunae found in Chapter 4, on the basis of literature on problems and solutions in the context of (elements of) big data. The overview of alternatives aspires to be a well-founded inspiration for further research; due to the scope of this research it is impossible to make a definitive judgement on the broader social and legal desirability of the different alternatives.

The first avenue for further research is amending data protection law. Data protection law can be altered in different ways to increase its protective potential and make it match better with recent technological developments. Creating property rights in personal data and broadening the material scope of data protection law to data that does not meet the criteria of personal data are discussed, but not considered to be valuable solutions to current problems. What should receive more attention is the role of other areas of law, and their potential to protect individuals in big data. Consumer law, competition law, and non-discrimination law are all areas that already regulate parts of the big data process, and could play a more significant role in the neutralisation of big data's negative impact. Furthermore, from the literature a number of new solutions specifically aimed at big data can be distilled, such as initiatives on *algorithmic transparency* (gaining insights into the operation of algorithms

used in big data), *big data ethics,* and sector-specific regulation. The latter may be of particular value in sectors that are distinct in terms of context, problems, or regulation, such as the financial and medical sectors.

Yet in themselves neither of these different alternatives provides the ultimate solution that would suffice for the overall protection of individual rights and freedoms. The general conclusion, expounded in Chapter 6, is that an approach solely based on privacy and data protection does not suffice, and neither does an approach that is solely based on one of the other alternatives. Because of the variety in problems, in phases in which problems occur, and in individual rights and freedoms that are affected by these problems, only a combined approach can offer sufficient protection against the negative impact of big data on the rights and freedoms of individuals.