The Information / Guarantees Balance - Protecting informational privacy interests within the European data protection framework

Emile Antoine Ennosuke Douilhet

Bournemouth University

The goal of this thesis is to study the notion of informational privacy, and how it is protected both by the right to privacy and the right to data protection. The age of Big Data has brought with it new ways of creating and processing data, which has led to challenges to the way that individuals are protected from the use of that data. This thesis shows that a new approach is emerging in European data protection law, which is based on protecting informational privacy specifically instead of the existing approaches which have been developed so far. We attempt to study and develop this new approach, and propose a framework based on it and how this might handle the challenges to informational privacy brought about by the age of Big Data.

This new approach is based on two core concepts. First, that there is a difference between “data” and “Information”, the second being the combination of “data” and human agency into a new entity, which includes not just data but also human intelligence, biases and imperfections. We assert that the absence of distinction between the two in EU data protection has led to the notion of “personal data” and “identifiability” as defined in the GDPR having challenges in handling data in the information age.

The second core concept is what we describe as “Guarantees”. The new approach being developed in data protection law attempts to restrict the processing of personal data by implementing measures which limit what data controllers can do when processing data. These measures all attempt not to prevent all possibility of harmful consequences occurring, but only to ensure that reasonable guarantees are in place to make these consequences unlikely. We thus call those measures “Guarantees”, and build a taxonomy of such Guarantees based on Lawrence Lessig’s four modalities: legal norms, social norms, market forces, and architecture.

As such, this thesis asserts that data protection is moving towards an approach having the protection of informational privacy as its core goal, protecting it through a balance between Information and Guarantees binding that Information.

Table of Contents

Introduction

A. Introduction to the Thesis Topic

I. Greater Context: Privacy, Data Protection and Informational Privacy

The Rights to Privacy and to Data Protection are some of the newest human rights in European law, protected under Article 7 and 8 of the Charter of Fundamental Rights of the European Union[1]. So new in fact, that protection of personal data as a distinct right has only been introduced with the European 1995 Data Protection Directive[2]. As data gathering becomes ubiquitous, control over it becomes an issue that no longer just matters to privacy, but to autonomy, democracy, and free speech as well[3]. This has led to a large academic body of work which has attempted to understand each right separately, and how they work together[4].

The Right to Data Protection is a young right, only a handful of decades old[5]. It is a right created to solve a specific problem - the creation of new technologies with capabilities that had never existed before[6]. The need for data protection is all the more important because of the arrival of the Big Data age. Big Data is the phenomenon by which various entities (corporations, governments, and increasingly even individuals[7]) have the ability to process a very large Volume of data, from a vast Variety of sources, with a high Velocity of processing[8]. This profound transformation in how knowledge is created and shared has had a similarly profound impact on what is known about individuals, and by

[1] Charter of Fundamental Rights of the European Union, 26 October 2012, 2012/C 326/02, available at: http://www.refworld.org/docid/3ae6b3b70.html [accessed 3 May 2018]
[2] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processingTry of personal data and on the free movement of such data, Off. J.L. 281 (Nov. 23, 1995) ("Data Protection Directive")
[3] Sobolewski, M., Mazur, J., & Paliński, M. (2017). GDPR: A Step Towards a User-centric Internet?. *Intereconomics, 52(4), 207-213.*
[4] Burkert, H. (2000). Privacy-Data Protection, *in Governance of Global Networks in the Light of Different Local Values, edited by Engel, C. and Keller, K.* (pp 43-70).
[5] Bennett, C. (1992). Regulating privacy: Data Protection and Public Policy in Europe and in the United States, *Cornell University Press*
[6] Burkert (n.4)
[7] Swan, M. (2013). The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data*, *1*(2), 85-99.
[8] Tene, O., & Polonetsky, J. (2012). Big data for all: Privacy and user control in the age of analytics. *Nw. J. Tech. & Intell. Prop.*, *11*, xxvii.

whom, creating new challenges for privacy and data protection. Data protection has a controversial relationship, however, with the right to privacy. The newly introduced General Data Protection Regulation[9] has carefully pruned references to privacy, but has not developed a fully autonomous concept of data protection, furthering the idea that the two rights are interwoven in a core way[10]. This leads to questions about the foundations of data protection and the notion of ''personal data''. Why is protecting data important? What is the link between data protection and privacy? Is privacy the core interest that data protection is trying to protect, or are there other interests to protect? If so, how far can data protection stretch? Over time, some patterns started to appear[11] - some common forces that seemed to be found in cases involving both privacy and data protection[12]. As we will show, these forces highlight the overlap of privacy and data protection to protect a specific interest: informational privacy.

II. Methodology

This thesis is concerned with these forces, and how the rights to privacy and data protection have been challenged by the advent of the Big Data age. The methodology of this work is doctrinal, studying the development of the right to privacy and data protection, combining an analysis of the legislation and case law surrounding privacy and data protection, with a particular focus on the General Data Protection Regulation, and a study of the literature developed around theories of privacy and data protection. Doctrinal research is defined as ''a detailed and highly technical commentary upon, and systematic exposition of, the context of legal doctrine[13]''.

The ''black letter law'' methodology is the primary research strategy of this thesis, analysing the legal rules to understand the meanings and implications of these rules, as well as the principles which underpin them. Using a comparative method of various theories of privacy and data protection, this thesis attempts to understand the foundations of those rights, and how the Big Data age affects the validity and

[9] Reg (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Dir 95/46/EC (GDPR) 2016

[10] Burkert (n.4)

[11] Banisar, D., & Davies, S. (1999). Global trends in privacy protection: An international survey of privacy, data protection, and surveillance laws and developments. *J. Marshall J. Computer & Info. L.*, *18*, 1.

[12] Burkert (n.4)

[13] Salter, M. and Mason, J. (2007), Writing Law Dissertations: An Introduction and Guide to the Conduct of Legal Research *Pearson 31.*

effectiveness of these theories. Using a doctrinal methodology for this work is necessary, since these rights are primarily based on legal instruments and cases, and the legal doctrines surrounding them. This means that a large part of the thesis will be devoted to fundamental doctrinal research, that is, to studying the underlying assumptions behind those rights.

Apart from legal doctrinal research, this thesis looks beyond the scope of the law itself and analyses the historical, cultural, sociological, and especially technological factors which have moulded these rights. This interdisciplinary element is vital, as this thesis seeks to understand what factors have created the law as it is today, and the arrival of Big Data has been one such powerful factor. These interdisciplinary factors are nevertheless limited so as not to broaden the scope of the thesis beyond what is necessary, and to keep it firmly anchored in doctrinal research, with a critical and qualitative analysis of legal materials in order to support its hypothesis.

III. Research Overview

The General Data Protection Regulation defined personal data as ‘’any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly’’[14]. According to this definition, the scope of the notion of personal data, and hence the scope of application of data protection law as such, essentially depends on how “identifiability” is construed. To keep pace with technological development, courts and data protection authorities have construed “identifiability” in an increasingly broad sense, to include any data which on its own is not identifiable, but may potentially be linked with other points of data in order to re-identify an individual[15]. Because ‘’identifiability’’ is based on the ‘’means likely to be used’’ to re-identify the data, and those ‘’means’’ keep getting more affordable and widespread, the notion of personal data is rapidly ballooning into a state where a huge, unenforceable regulatory burden weighs on every company. Additionally, it is possible to violate the societal interests that data protection is meant to protect without actually violating data protection law[16]. In short - the current data protection rules, though

[14] GDPR, Article 4(1)

[15] Clauβ, S., Kesdogan, D., & Kölsch, T. (2005, November). Privacy enhancing identity management: protection against re-identification and profiling. In *Proceedings of the 2005 workshop on Digital identity management* (pp. 84-93). ACM.

[16] Schreurs, W., Hildebrandt, M., Kindt, E., & Vanfleteren, M. (2008). Cogitas, ergo sum. The role of data protection law and non-discrimination law in group profiling in the private sector. In *Profiling the European citizen* (pp. 241-270). Springer, Dordrecht.

functional in a majority of cases, are increasingly at risk of both over-regulating non-problematic practices, while under-regulating practices that encroach upon fundamental rights and interests of individuals.

In the Patrick Breyer v Germany case decided on the 19th of October 2016, the Court of Justice of the European Union was faced with the question of whether dynamic IP addresses were ‘’personal data’’. In that case, the CJEU took a relative approach to the ‘’identifiability’’ criteria, which means the same piece of information can be personal or non-personal data, depending on who is holding it and whether they can identify the person using it. That goes as far as the idea that since a website operator has ‘’legal means’’ of obtaining access to the information held by an ISP to identify an individual, this is personal data. The number of factors which might make data ‘’personal’’ are so varied, so subjective, that it may change from day to day, from person to person[17]. This does not lend itself well to a binary approach.

The starting point of this thesis is the limitations of the notion of “personal data” as defined by the legislation and as expanded by jurisprudence: that as shifts in technology increasingly change how information is created and shared, the traditional conceptual framework for understanding privacy and data protection is failing, and a new framework is taking shape in order to address the previous one’s limitations. The goal of this thesis is to analyse and understand these patterns, and use them to articulate this new conceptual framework and how it might be able to address the challenges of the Information Age.

A possible conceptual framework is the ‘’risk-based approach’’, an approach based not on whether the data is personal, but on the impact on the individual if a risk manifests[18]. The risk-based approach is pushed heavily by the GDPR as a parallel system to the rules-based system surrounding consent, and was considered a more pragmatic, realistic, adaptable system[19]. However, it eventually became apparent that taxonomies of harm related to the topic were impossible to establish conclusively, principally because the interests that data protection seeks to cover are very broad, and the GDPR expressly avoids defining them, instead trying to base risk on the data and its processing itself. This would be a fine approach (and one which inspires the approach developed in this thesis), but it is held back by the fact that parallel to this ‘’data protection risk’’, guidance

[17] Schwartz, P.., and Solove, J. (2011). The PII problem: Privacy and a new concept of personally identifiable information. *NYUL rev.* 86 (2011): 1814.
[18] Maldoff, G. (2017). The risk-based approach in the GDPR: interpretation and implications. *IAPP https://iapp. org/media/pdf/resource_center/GDPR_Study_Maldoff. pdf*
[19] Ibid

related to the GDPR soon became linked back to privacy, and from there to the quagmire that is ''privacy harm''[20], a notion that is particularly elusive and controversial[21].

This has meant that one cannot base the framework either on the ''identifiability'' of the data itself, or the actual ''harm'' its mishandling could cause. Additionally, the rights to privacy and data protection, though they are held as separate rights, are not fully independent from one another. Despite the GDPR not mentionning the word ''privacy'' (unlike its predecessor, the Data Protection Directive)[22]/[23]. Because of that, our thesis asserts that that the overlap between the two notions – dubbed ''informational privacy'', a term based on existing taxonomies of privacy and data protection[24] - cannot be neatly split apart, but must be taken together. The notion of ''informational privacy'', was best described in ''A typology of privacy'' by Bert-Jaap Koops and others[25] as '' a broader concept, encompassing information/data/facts about persons or their communications.''[26], a broad notion which is not based on particular interests but overlaps many. This thesis argues that data protection should focus on informational privacy, and any particular measures aimed at particular interests should lay these out explicitly, instead of attempting to stretch the notion of personal data to include an ever-increasing number of informational interests[27]. The notion of ''informational privacy'' which will be developed in this thesis is based, specifically, on an expansion of the scope of the right to data protection in order to overcome the limitations of the ''personal data'' approach.

The GDPR itself is moving away from the traditional ''personal data'' approach: most of its new developments use new tools in order to protect individuals[28], and increasingly shift focus away from a strict set of rules and towards a risk-based, pragmatic

[20] Calo, R. (2011). The boundaries of privacy harm. *Ind. LJ, 86, 1131.*
[21] Maldoff (n.18)
[22] The word ''privacy'' appears 13 times in the Data Protection Directive, usually in the form of the statement ''protect the fundamental rights and freedoms of natural persons, and in particular their right to privacy'', indicating the prevalence of the right to privacy in data protection. Comparatively, the word privacy appears not once in the text of the GDPR.
[23] Rouvroy, A. (2008). Privacy, data protection, and the unprecedented challenges of ambient intelligence. *Studies in ethics, law, and technology*, *2*(1).
[24] Koops, B. J., Newell, B. C., Timan, T., Skorvanek, I., Chokrevski, T., & Galic, M. (2016). A typology of privacy. *U. Pa. J. Int'l L.*, *38*, 483.
[25] Ibid
[26] Ibid
[27] Case C-434/16, Nowak v Data Protection Commissioner, (2017) ECLI:EU:C:2017:994
[28] Goodman, B., & Flaxman, S. (2016, June). EU regulations on algorithmic decision-making and a "right to explanation". In *ICML workshop on human interpretability in machine learning (WHI 2016), New York, NY. http://arxiv. org/abs/1606.08813 v1*.

approach[29]. A new way of understanding informational privacy is developing, and the focus of this thesis is to understand and map this new conceptual framework.

In order to do so, this thesis builds a framework based on Lawrence Lessig's ''Four Modalities'', and particularly inspired by Helen Nissenbaum's ''Contextual Integrity''. This framework starts with a basic assertion inspired by Lessig's work: that when a change occurs in how information is obtained, some forces will enter into effect to restore an appropriate balance[30]. For example, new technological tools, which allow gathering of data previously impossible to obtain, may get regulated by a new legal instrument to restore balance. Based on that, we have used the ''Four Modalities'' to map out these forces – called ''Guarantees'' – and to assert that as long as these forces are balanced, ''informational privacy'' is preserved. The rest of this thesis is spent proving this assertion, and laying out a way to identify the various underlying forces and use them to assess the state of the Balance in different situations. We propose that this ''Balance'' approach is the direction that EU data protection is headed towards, and that the GDPR is the first step on that path.

One of the primary goals in creating this framework is to avoid the pitfalls of other conceptions, the greatest of which is the difficulty of defining what interests should be protected. The reason why the GDPR does not mention privacy is to dissociate itself from that right, in order to avoid having to define what privacy is (which, as the first part of this thesis will show, is the source of much academic debate[31]). The GDPR attempts to make data protection a wholly separate right, not dependant on any further rights or interests and with its own value. However, as we will show, this attempt has not been wholly successful, with the recognition of ''risk'' which goes beyond risks to data protection and to various ''harms'' covering a huge variety of rights and interests. Inspired by Helen Nissenbaum's contextual integrity, the conception I develop in this thesis looks only at whether the various forces are balanced. Differently from Nissenbaum's conception, the focus is on the Information itself, and not the context in which it is processed. Context is relevant, but only in so far as it imparts some Guarantees to the controller of the Information. Through conceptualizing informational privacy as a balance of Guarantees, we attempt to bypass the limitations of both the harms-based framework (and its difficult to define ''harms'') and the personal data framework (and its limited scope). The GDPR's conception of ''risk'' is moving towards an interest-neutral

[29] Butin, D., & Le Métayer, D. (2014, May). Log analysis for data protection accountability. In *International Symposium on Formal Methods* (pp. 163-178). Springer, Cham.
[30] Koops (n.24)
[31] Ibid

conception, in line with the rest of the GDPR, defining risk not based on particular harms, but on certain processing activities instead[32]. As data protection moves further from specific interests and how they may be harmed, we propose a framework based on a balance of Guarantees, which would be completely separate from those interests and harms, as the way forward for data protection.

B. Thesis Structure

I. Research Questions

Before going into detail as to how we will develop this issue, we will go over the core research questions which drive this approach.

This thesis is built on two core assertions, demonstrated, analysed and evidenced throughout it. The first is that the concept of ''personal Information'' should replace the one of ''personal data'' in the EU data protection regulatory system, with the main difference being the removal of the ''identification'' element of ''personal data'' and a renewed focus on the ''relating to'' element of the definition. Data can be transformed and interpreted in a way that makes linking it back to an identifiable data subject impossible. However, the ''relating to'' element of personal data includes not just whether the data is ''about'' the person, but also whether the data is processed in a way which either intends to affect the individual, or results in them being affected[33]. The second is that the concept of a ''Balance of Guarantees'' should be used to regulate the creation, use, and dissemination of this Information, accounting for forces beyond those officially recognized so far by EU regulation.

The primary research question we will try to answer is thus: ''Is the EU protection of personal information shifting from an ''identifiability''-focused framework to one based on a balance of Guarantees for those holding the means to create personal Information, and is that framework able to address the limitations of other conceptions as well as the challenges brought on by the age of Big Data?''

[32] Article 29 Working Party (2017), Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679, WP 248, Adopted on 4 October 2017

[33] Article 29 Working Party (2007), Opinion 4/2007 on the Concept of Personal Data, WP136, Adopted on 20 June 2007

This question requires some unpacking. This thesis will develop and analyse each piece of it, but first starts by building a foundation. What are ''privacy'' and ''data protection''? How has the legislation surrounding these two rights evolved, and what interests are they trying to convey? Answering that, as well as understanding what ''informational privacy'' is, is paramount to understanding the framework developed in this thesis..

We then study the development of forces which have challenged these rights in a fundamental way: new technological tools which together constitute ''Big Data''. How do these new tools challenge the existing framework protecting informational privacy? Why is the arrival of ''Big Data'' such an important transformation? Is it an altogether new phenomenon, or simply a bigger version of something that came before? Answering those questions by studying this new force is vital, as the hypothesis of this work is that informational privacy should be protected by a balance of forces: under that hypothesis, Big Data is the greater shift in these forces in history.

Because it is possible for pieces of data which, alone, are unable to identify an individual, to be combined in a way which makes it possible to learn vast amounts of information about them, the ''identifiability''-centric approach which underlies much of the GDPR is increasingly unable to handle the new ways by which knowledge can be generated. However, new tools in the GDPR[34] as well as some opinions by data protection authorities[35] seem to be moving away from that ''identifiability''-focus criteria. We propose a shift away from ''identifiability'', as well as away from focusing on the data itself, and instead focusing on what information can be learned from that data based on the means and motivations of those holding it. That is what we will call ''personal Information'' throughout this thesis (or ''Information'' when referring to the greater concept of knowledge obtained from data, regardless of whether it relates to an individual).

This leads to the question as to how we will develop the concept of ''personal Information''. How can personal Information be linked back to the interests of individuals in the same way that personal data is? How can Information be regulated without the binary approach of ''identifiability'' adopted by the GDPR?

[34] See GDPR, Article 22 – the right to automated decision-making focuses on decisions which may affect the individual, not on the data used to do so.
[35] Article 29 Working Party (2007), Opinion 4/2007 on the Concept of Personal Data, WP136, Adopted on 20 June 2007

II. Chapter Summaries

The answer developed in this work (and supported by evidence from EU privacy and data protection legislation) is to move away from a strict, binary, rules-based system, towards one based on forces, that is, based on the means and motivations by which various parties can create personal Information. These forces, dubbed ''Guarantees'', are modelled after Lawrence Lessig's Four Modalities[36]. This leads to questions of how to assess what constitutes sufficient Guarantees, and whether using Guarantees instead of other approaches – such as the harms-based approach – is more effective in protecting informational privacy.

The analysis will be divided into six chapters.

In Chapter One, we analyse privacy and the way in which that notion has changed over history, laying out the groundwork for the balance created later in the thesis by showing that there is a very clear pattern of changes to how personal information flows through a society becoming balanced over time, sometimes by legislative intervention, but sometimes by other forces. This thesis suggests that we are simply at another one of these paradigm shifts, and the only question is whether the forces we can put in place can be enough to offset the change, or if things have truly changed, for good. We then analyse the notion of ''privacy'' itself, its evolution as a right, and in particular the way it has developed in European law. This serve a double purpose - to show in more detail how these approaches were developed as reactions to changes in Information control, and to understand better how approaches to privacy work and how successful they have been. This also allows us to frame our analysis better by defining what is meant by ''Informational privacy'', allowing a development of this term without having to take on the impossible task of defining privacy and without having to take on all the varied undefined interests to which data protection can extend.

The second Chapter addresses data protection more specifically. We highlight the fact that the idea of data protection as a separate right makes it possible to avoid the limitations of ''privacy'' , and that the conception developed in this thesis is simply an expansion of that idea. Data protection has intentionally distanced itself from specific interests such as ''privacy'' or ''consumer welfare'' which are hard or even impossible to define, instead focusing on protecting ''personal data'', in and of itself, based on

[36] Lessig, L. (1999), Code and Other Laws of Cyberspace, *Basic Books*

principles of care and responsible gathering, handling, and securing of that data[37]. This transition is difficult, and not yet completed. In order to ensure that the GDPR is successful, adaptable and enforced, it was built in an interestingly pragmatic way. The core of the GDPR attempts to focus only on the data itself, without any of the context of what type of data it is and how it may impact various interests.

In some situations, however, the GDPR cannot ignore the interests which go beyond data protection alone, and has had to develop tools in order to take them into account, such as rights against automated decision-making[38], which cite non-discrimination as one of the interests which may be harmed by that practice[39]. We show that these gaps that the GDPR has filled with new measures are also gaps that have been created by the Big Data age, such as profiling, and that there is evidence that data protection is moving towards a framework which is able to emancipate itself from hard-to-define interests, the way data protection does, while still taking them into account. We propose that the framework developed in this work fulfils those criteria, and that it is the way along which data protection is already heading.

Having set the stage in terms of privacy and data protection, the third Chapter highlights the technological developments that have created the current challenges. Understanding the power and scope of these technologies is necessary in order to understand the difference between ‘’data’’ and ‘’Information’’: we are arriving at an age where it is not just the data we hold, but how we can analyse, use, and interpret that data, that has the most significant impact. One cannot understand the need for a new approach without understanding the technologies which threaten the existing one. Terms like profiling, Internet of Things, cookies or Big Data need to be laid out and analysed to see the impact they have on the way Information flows in our society.

We discuss the notion of ‘’Information’’, which we argue should be used as a conceptual framework to interpret the concept of ‘’personal data’’. In a metaphor we call ‘’the Sculptor’s Work’’, we show how the use and processing of data is less like the ‘’mining’’ of data, and more similar to a sculptor creating something new from a block of marble. The same marble can create an infinity of sculptures – depending on the sculptor. In the same way, the same data can create different Information based on a wide number of

[37] The most evident way to highlight this transformation is the GDPR not mentioning privacy a single time, Interestingly, other interests, such as discrimination, are mentioned, which show how intent data protection is intent to show itself separate to various interests, but especially privacy.
[38] GDPR, Article 22
[39] GDPR, Article 22(3)

factors, and these factors are getting exponentially more influential over time. This has a number of implications, but most important amongst them is that many factors affect what can be learned from data, which is not taken into account by the current conception of data protection. We then build on that core argument to show that a conception which allows for an ''Information'' perspective is already developing. Evidence, from intellectual property law as well as earlier developments on privacy and data protection law, demonstrate this. If it is possible to create something new from raw data with its own legal status, then ''Information'' as a concept can exist separately from ''data''. Finally, we develop ''personal Information'', as opposed to ''personal data'', and show that the transition towards ''Information'' is already showing signs of being adopted.

The fourth Chapter analyses the foundations of European data protection law, finding its limitations, and the source of these limitations. In that part, we develop the argument that the Big Data era is challenging some fundamental distinctions in EU privacy and data protection law. We display this through the study of the ''means likely to be used'' test developed to assert whether data ''identifies'' an individual, and show that the GDPR recognises, through certain of its provisions, the limitation of this conception. We then expand the analysis to examine two core pillars of EU data protection law and how they show the limitations of that law: anonymisation and consent. The ''personal data'' conception and the focus on ''consent'' as the way to legitimise data processing are central to how the GDPR works, and their weaknesses highlight the need for a solution.

After identifying the weaknesses in two core pillars of data protection, the Chapter does the same with a core pillar of privacy: the public/private distinction. In the same way that data protection is strongly divided between data that is personal and data that is not, privacy has a separation between the public and the private space, which is quickly eroding in the information age. This analysis is grounded on the idea of the private/public distinction being obsolete, and needing to be replaced with a ''transparency/opacity'' distinction, which takes into account the new ways in which personal Information is created and disseminated.

Chapter five analyses the Guarantees, the balancing forces which limit how those who control the means to create Information can act. We start out by analysing the literature which backs up the idea of Guarantees: the works of Lawrence Lessig and Helen Nissenbaum. We also study the common ways this problem has been answered, the limitation to those approaches (especially in the case of ''privacy harm'') but also how they can inform the framework developed in this thesis.

This leads to the explanations of what the Guarantees are. Explaining this notion requires some development, because it comes from a very specific mindset, and those Guarantees have a very wide scope, based on Lawrence Lessig's taxonomy. We start by studying technological Guarantees – or ''architectural'' as Lessig put it – by analysing how technology and the ability we have to use it to shape our environment influences the creation of Information, but also its control. We then study market Guarantees, which are not taken into account by the GDPR: the ''identifiability'' test, based on ''means likely to be used'' to identify the data subject, only looks at the technological capacity of the data controller, and not whether they have sufficient economic motivation to use that capacity to identify them. Because of that, and the increasing expansion of both the notion of ''personal data'' and the technological means for identification, controllers with no incentive to identify individuals or obtain particularly sensitive data about them have to comply with the full weight of the GDPR, just like controllers who have that incentive. For businesses far removed from data analytics, this can be an unexpected and pointless burden. Market Guarantees are also important as a way to show that the price of technology is as much a barrier as the existence of the technology itself: in other words, if the technology is too expensive for the value of the Information it produces, then that technology will not be used.

We then examine the social Guarantees, which also have an important impact, especially when it comes to changing behaviours over time. Many rules of privacy are unwritten, and become legal rules only where technological developments make it impossible to rely on social rules alone. Additionally, social rules evolve over time, and the current conflict over the future of privacy – whether society is accepting a world without it by mass participation in social media – may change irreversibly the balance of Guarantees. One of the hypotheses of this thesis is that society tends not to change fundamentally its views on the need for privacy, which is why we argue that it is possible to set a ''default'' level for the balance. We argue that the claim society is fundamentally changing its value of privacy may be mistaken, and that the balance will reassert itself, a view which early signs are showing is likely to be happening. Finally, we go over the role of legal Guarantees, and how they are just here to fill in gaps where the natural state of the balance cannot reassert itself naturally. As such, this concludes that the only role of any law dealing with informational privacy only needs to work to re-establish this balance, and so should be conceptualized accordingly. The GDPR does so in a number of ways, but is still incomplete.

Finally, in Chapter six, we close this thesis by establishing the last steps of the Information/Guarantees balance. First, the argument for the ''default'' state for the

balance is made, based on historical precedent. This argument is based on the pattern identified in the beginning of the thesis: that over time, the balance has been uneven multiple times which has meant that measures were taken to rebalance it. As such, we identify a reasonably even balance as any point at which no measures were taken for a significant amount of time. The last step is an overview of various measures which might help balance the Guarantees. It includes many measures which have already been proposed - but put in the new light of the Information/Guarantees balance. In particular, it shows the successes of the European approach in addressing the Balance, in particular by focusing on the purpose limitation principle.

At the end of this thesis, we intend to have proven that separating informational privacy completely from specific interests and harms is necessary in order to avoid the pitfalls of other conceptions. Seeing it as a balance of forces which need to be set right allows a multidisciplinary, adaptive and effective approach, because of its ability to use various Guarantees. Basing this balance not on ‘’personal data’’ but on ‘’personal Information’’, a concept which sets aside the ‘’identifiability’’ test, makes it possible to address the challenges of Big Data by being technology neutral and not being limited by the data itself, but looking at the whole picture instead. Meanwhile, using the Guarantees makes it possible to bypass the binary test of ‘’identifiability’’ while still protecting individuals. As we have shown, this is the direction along which EU privacy and data protection law is already heading – though whether that transition can be fully made considering how core the ‘’identifiability’’ test is to the existing framework is still not a certainty.

Chapter 1. An Introduction to Informational Privacy

Introduction

Multiple conceptions of privacy, frameworks, classifications and taxonomies have been developed to understand better the right to privacy[40]. Some have attempted to see privacy as being defined by certain risks[41], some relying on a torts-based approach to address potential abuses[42], with several taxonomies of the ''harm'' that may result from a breach of privacy having been attempted[43]. Some theorists have suggested new rights, from identity[44] to procedural data due process[45]. Some see the answer to adapting privacy to the Big Data age as lying in new technologies[46], while some see it in stronger legislation[47]. With the birth of the Information Age, this is only getting more complex as privacy concerns multiply and obtain new dimensions and interests.

Instead of dealing directly with this - seemingly unanswerable - question, we propose a limited approach which bypasses this question altogether, by not stating an absolute definition of privacy and instead looking at the role of privacy as a force put into place to make up for an imbalance of forces - a force of "control". This conception of privacy -

[40] Danezis, G., & Gürses, S. (2010). A critical review of 10 years of Privacy Technology. *Surveill. Cult. A Glob. Surveill. Soc.*, 1–16.
[41] Centre for Information Policy Leadership. (2014). *A Risk-based Approach to Privacy: Improving Effectiveness in Practice*, 19 June 2014
[42] McKay, C. (2015). Next Generation Privacy: The Internet of Things, Data Exhaust, and Reforming Regulation by Risk of Harm. *Groningen Journal of International Law*, *2*, 30.
[43] See "Calo, M. R. (2011). The Boundaries of Privacy Harm. *Indiana Law Journal*, *86*, 1131.", and "Solove, D. (2006) "A taxonomy of privacy". *University of Pennsylvania Law Review*, 477-564"
[44] De Hert, P. (2007). A right to identity to face the Internet of Things ?, *Strasbourg: Unesco* 1–21..

[45] Crawford, K., & Schultz, J. (2014). Big Data and Due Process - Toward a Framework To Redress Predictive Privacy Harms. *BCL Rev.*, *55*(1), 93–128.
[46] Danezis (n.40)
[47] Hornung, G. (2012). A General Data Protection Regulation for Europe? Light and shade in the Commission's draft of 25 January 2012. *SCRIPTed*, *9*(1), 64–81. doi:10.2966/scrip.090112.64

known as ‘’informational privacy’’ - has a limited scope based on where data protection and privacy overlap.

Our aim is to show that attempts to define and conceptualize privacy may not be taking into account certain underlying forces surrounding it. By looking at the evolution of the technological landscape, it would indeed seem that technology is ever-changing, and that each paradigm shift is so unpredictable and so unlike what we've seen before that no conception of privacy could be "future-proof". For example, in "Seven Types of Privacy"[48], Rachel L. Finn, David Wright, and Michael Friedewald build on a previous approach by Roger Clarke, arguing for the replacement of Clarke's taxonomy of "four categories of privacy" by expanding and updating it to seven categories to adapt to the modern world[49].

The authors' thesis in this article (which we will examine in greater depth later) is that privacy issues become more complex over time, and privacy conceptions need to adapt to them as they appear[50]. And while that is correct, our argument is that, though there is change and development, there are also recurrent patterns within these privacy issues which can be identified. Our goal is to highlight these patterns, and the different forces of which they consist, in order to build a conception of privacy which has a limited, defined scope that can allow for a framework which can protect it without losing itself in a myriad of other interests.

The conception we will be using is thus based on a limited conception of the European conception of privacy, one which overlaps to a great extent with data protection. Since, as we will see, the European conception of privacy is quite extensive and its boundaries are ill-defined, we will be using a more specific understanding of privacy, which we qualify as "informational privacy", and purposefully removing from this analysis the many different other interests that privacy can cover.

Our conception of "informational privacy" aims at showing that there exist some underlying patterns involved in the creation and control of personal information, that have been approached before in many ways, but never identified within their own framework. We will then highlight some later attempts, in particular Helen Nissenbaum's contextual

[48] Finn, R., Wright, D. and Friedewald, M (2013). Seven Types of Privacy, *Dordrecht European Data Protection: Coming of Age*
[49] Ibid
[50] Ibid

integrity[51], and what makes them successful by showing that they touch on the "balance" framework that we will be attempting to demonstrate. By studying the history of attempts at defining "privacy", one can identify the need to adapt to new technologies and the challenges they pose. Instead of redefining privacy every time a new technology appears, we argue we should instead look at the underlying forces driving these changes, and build a framework that takes them into account, as these forces will exist no matter the time or technological level, only in different forms.

We will show the inception of the idea that the "balance" is upset, by looking at how privacy is unique in the sphere of Human Rights in that it is an altogether young right, which only exists because of technological developments having upset a system which previously did not need legal intervention to be balanced. We will also show that this essential character of privacy is extremely significant in understanding its role: why it was needed then, why it is needed now, and what form it needs to take in order to fulfill that role.

We will finally link this model to our working conception of privacy, based on European privacy law. Through this analysis, we will show the role of data protection in this conception, as a right that was created in part to respond to an upset in the balance. Because of the right to data protection being a consequence of this upset, it has been the field of law which has been given the task of solving the problems brought about by this imbalance. As such, data protection is a field which will be used throughout most of this thesis, because the concept of "privacy" and the concept of "data protection" overlap to a great extent. In other words, this thesis will limit itself to a specific interest, one which lies in the area where privacy and data protection overlap – ''informational privacy''. This is especially relevant because, though data protection was originally created to deal specifically with informational privacy, it is expanding to protect other interests as well[52].

Through this development, we will show how a vision of privacy as a balance of shifting forces can give new insight into this particular area of privacy.

Before elaborating on the vision of privacy as a balance of control, we will first establish the currents in conceptions of privacy over time and how they show hints of this balancing approach. This will show a trend in conceptions of privacy being developed almost

[51] Barth, A., Datta, A., Mitchell, J. C., & Nissenbaum, H. (2006). Privacy and contextual integrity: Framework and applications. *Proceedings - IEEE Symposium on Security and Privacy*, *2006*, 184–198. doi:10.1109/SP.2006.32

[52] Burkert (n.4)

entirely to balance out new technological developments, and will highlight the preponderant role that data protection has taken as a tool for privacy.

A. The Evolution of Privacy: Identifying Recurring Patterns

Privacy, and tools to ensure that privacy, is not a new idea. The opposition of society and solitude goes back to Greek philosophy[53]. There have always been mechanisms[54], legal, social, or technological, meant to separate private and public life. Throughout the evolution of the concept of privacy, we can see a clear causal link between changes in the ability to obtain information by new or different means, and new developments in the landscape of privacy law - and later, data protection.

This is true of the development of the "home" as a sanctuary of privacy coinciding with the development of cities where traditional social norms linked to community and family had disappeared[55]. For most of history, communities were close knit, society was "small" in the sense that one's social sphere was limited to neighbors, coworkers and family[56]. As such protecting one's privacy was not a legal matter so much as a social one - a small social circle means easy control over who knows what. But cities led to the erosion of this sense of community, and with it came a demand for privacy. The intimacy of a necessarily small social circle, which had until then been the major protection for privacy, was no longer there, and the "home" became the new focus for privacy protection[57]. Even today, American privacy law sees a violation of the home as the utmost invasion of privacy[58]. The imbalance created by the move to cities was rectified by the rise of the home as a private space: whereas one's control over privacy was once due to the close community, it was now due to protection afforded by the home, and the sociolegal recognition of that home[59].

---

[53] De Hert, P., & Gutwirth, S. (2006). Privacy, data protection and law enforcement. Opacity of the individual and transparency of power. *Privacy and the criminal law*, 61-104.
[54] Holvast, J. (2008) History of privacy. *Springer Berlin Heidelberg.*
[55] McKeon, M. (2009) The Secret History of Domesticity: Public, private, and the division of knowledge. *JHU Press*
[56] Ibid
[57] Koops, B. J., & Leenes, R. (2005). Code and the slow erosion of privacy. *Mich. Telecomm. & Tech. L. Rev.*, *12*, 115.
[58] Ibid
[59] Baumer, D. L., Earp, J. B., & Poindexter, J. C. (2004). Internet privacy law: a comparison between the United States and the European Union. *Computers & Security*, *23*(5), 400-412.

This is also true of the increasing public dissemination of newspapers concerning notable individuals at the end of the 19th Century[60], and the invention of the instantaneous photograph, being directly responsible for Warren and Brandeis writing their famous essay, "The Right to Privacy"[61].

This last example deserves a moment of pause, because it is the first development which can be directly compared to the challenges we see today in terms of data collection and analysis, and the first peek into the world of data protection. Looking at the situation before and after in terms of balance will illustrate the fact that this balance can always be found, whether one is looking at newspapers in the 1890s or Google in 2016.

Before technological developments allowing newspapers to disseminate information easily and allowing instantaneous photographs, there existed certain practical protections to privacy. The aforementioned sacrality of the home and of personal correspondence (Legal). The social norms protecting one's personal affairs from excessive scrutiny (Social). The technological challenge of getting pictures of individuals without their consent or distributing to the general public newspapers containing the information (Technology). But finally, and possibly most relevant in this case, is the fact that newspapers are businesses pursuing profit, and that it was unprofitable to gather this information before these innovations allowed for juicy, easily disseminated articles (Market).

The development of instantaneous photography and easy dissemination of media upset this balance by removing some technological limitations. The first, instantaneous photographs, was a change in what information could be collected – photos of individuals without their consent. But the second, notably, had nothing to do with the collection of information, but with its dissemination. The privacy invasion was not just that newspapers were able to obtain such information: it was that the public at large was able to obtain it too. This shows how changes in information sharing can have an impact on privacy regardless of whether there is a change in information creation. This establishes an important rule: restricting data dissemination can be as important as restricting data collection.

---

[60] Hixson, R. (1987), Privacy in a Public Society. Human Rights in Conflict . *New York, Oxford: Oxford University Press*

[61] Solove, D. J. (2006). A Brief History of Information Privacy Law. *Proskauer on Privacy: A Guide to Privacy and Data Security Law in the Information Age*, 1–46. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=914271

However, the dissemination of media had repercussions on other practical concerns, such as economic profit. Privacy invasion is only interesting to newspapers if they obtain a profit from it, and as such the fact that privacy invasion was not worth it from an economic perspective disappeared, leading to a greater dissemination of personal information. Because of these changing factors, the balance which existed before these innovations was upset. And this is what prompted Warren and Brandeis to work towards establishing a right to privacy. It is an example of how when the balance is disturbed by the weakening of an existing protection, attempts are made to redress it, in the case of Warren and Brandeis by establishing the Right to Privacy.

Finally, this causal link is also true for the creation, seventy years after Warren and Brandeis' article in the United States, of a taxonomy of privacy harms by William Prosser[62], the content of which will be studied in further detail below. As the technological landscape continued to evolve, new privacy issues arose, and eventually a clearer, more robust taxonomy was required to keep up with the expanding sphere of privacy invasion.

This same pattern, a change in the environment leading to new protections being created to re-establish the balance, is found at every level, in every country, in every time period. However, by looking deeper at various developments, we will see that this pattern has not been identified as such by most attempts to reform privacy law. This has led to some attempts not answering the root of the problem - restoring the balance. In particular, this is the case of the European data protection, which has set itself certain goals which, while they act on some forces affecting the balance, do so sometimes inadvertently and imperfectly, which means that certain cracks remain open.

Seeing current privacy issues through the lens of that balance can allow us to understand better the privacy challenges we face, and the solutions we need. In order to study this, we will analyse some major doctrinal approaches to privacy, and how these have evolved over time. Through this analysis, the pattern of new technologies disrupting the way privacy is challenged will be explored, and how various types of approaches have dealt with this pattern - or failed to.

[62] Ibid

B. Doctrinal Approaches to Privacy: Rectifying Imbalances as they Appear

We have seen that there exists a pattern of balancing forces at every stage of the development of the right to privacy. Every great step forward made by privacy law has been done in direct response to a - usually sudden, and usually technology-driven - imbalance of these forces. We will now study the main proposals made to strengthen the right to privacy, identify what imbalance led to their creation, whether they rectify it, and why. We start by looking at privacy conceptions in the US, then moving on to European ones. This is due to the fact that American approaches offer a clear case study of successive developments trying to capture the elusive concept of privacy while being challenged by successive technological developments.

I. The Torts-Based Approach

Before studying the European conception we will look at a common approach which was developed first in the United States, and now is the major legal regime governing American privacy law: the torts-based approach to privacy. The reason for this cross-Atlantic analysis is that some useful lessons can be learned from the torts-based approach. Because it needs to adapt to new harms, it is an approach that always tries to catch up, which is getting more and more difficult with the rapid development of privacy-challenging technologies. Nevertheless, it does highlight a recurrent pattern: that a ''privacy harm'' becomes codified in law not when that harm becomes a possibility, but when that harm becomes so prevalent that legal measures have to be taken to prevent it. As we will show, there are underlying patterns behind these harms that, if identified, could allow for a more reactive approach.

1. Warren and Brandeis' Right to Privacy

As mentioned above, Warren and Brandeis' "Right to Privacy" was written as a response to new technological developments, namely instant photography and the commercialization of tabloids[63]. These developments allowed an invasion of privacy which had never been technologically possible before, and in response the authors

[63] Ibid

characterized a number of privacy torts, new harms which, while they had always laid under the surface, suddenly became major concerns[64].

The idea of privacy torts began with Warren and Brandeis' "Right to Privacy", but only started to take shape as a legally-binding approach in the early 20th century[65]. The first statute on the topic dates from 1903[66], and was followed by a decision of the Georgia Supreme Court in 1905 which established a tort for privacy invasions in Pavesich v. New England Life Insurance Co.[67]. The facts of the case were that a life insurance advertisement was using the image of the plaintiff without their consent[68]. The defendant obtained the negative of an existing picture of the plaintiff, then using this picture for his commercial purposes. The Court's reasoning in the case mentioned: "One who desires to live a life of partial seclusion has a right to choose the times, places, and manner in which and at which he will submit himself to the public gaze.[69]"

When put under the lens of a "balance" between technological developments endangering privacy and forces put into place to contain the privacy intrusions created by these developments, this affirmation gains new meaning. This case would not have been possible but for the existence of photography, and the ability to obtain previously-unobtainable information. This goes back to the instant photographs which prompted Warren and Brandeis to develop their Right to Privacy, and the same reasoning driving the judge's decision to establish privacy invasions as a tort. The principle laid down by the judge in that decision went beyond the small scope of the use of photographic negatives, but the fact that it was prompted by this technological innovation is no accident. By creating a larger principle, the judge was attempting to establish a protection that would extend to all cases in which privacy was threatened - whether or not technology was a factor. As we will see, this trend continues.

[64] Ibid
[65] Ibid
[66] Ibid
[67] Pavesich v. New England Life Ins. Co., 50 S.E. 68 (Ga. 1905).
[68] Solove (n.61)
[69] Ibid

2. William Prosser's Four Privacy Harms

The next major innovation came with William Prosser's taxonomy of privacy harms in 1960, which recognized four torts: intrusion upon seclusion, public disclosure of private facts, false light, and appropriation.

Intrusion upon seclusion, as Prosser defines it, provides that: "One who intentionally intrudes, physically or otherwise, upon the solitude or seclusion of another or his private affairs or concerns, is subject to liability to the other for invasion of his privacy, if the intrusion would be highly offensive to a reasonable person"[70]. It covers in its scope offenses such as clandestine entry or photography into one's home. It extends to privacy offenses linked indirectly to new technologies, such as overbearing surveillance of an individual while in public, which shows how new technologies create new challenges when they enter the mainstream or are used by powerful entities like the government.

Public disclosure of private facts provides that: "One who gives publicity to a matter concerning the private life of another is subject to liability to the other for invasion of his privacy, if the matter publicized is of a kind that (a) would be highly offensive to a reasonable person, and (b) is not of legitimate concern to the public.[71]" This once again looks back to the roots of technology-driven privacy law, with Warren and Brandeis creating the Right to Privacy partly because of the recent availability of easily-disseminated newspapers. For example, a case based on this tort was Smith v. Daily Mail[72], in which the publication of information on juvenile offenders was seen as a breach of privacy. The obtaining of this information by the newspaper was not the problem - the problem was the consequent mass dissemination of that information, which would not have been possible a century earlier. The addition of a "public interest" exception is notable, as it highlights the balance of interests inherent in considerations of privacy.

False light, having as a tort the act of spreading false information about an individual, was not directly applicable to our purposes back when Prosser theorized it, but also takes on major significance with the technological developments of the modern age: as increasing amounts of information get processed and used to make decisions, it becomes increasingly likely that individuals will be portrayed in a wrongful light due to inaccurate or obsolete data or human bias leading to inaccurate conclusions.

---

[70] Restatement of the Law, Second, Torts, 652 Copyright (c) 1977, The American Law Institute

[71] Ibid

[72] Smith v. Daily Mail, 443 U.S. 97 (1979).

Finally, Appropriation pertains to: “One who appropriates to his own use or benefit the name or likeness of another is subject to liability to the other for invasion of his privacy”[73]. This once more is increased by technical means previously unavailable, such as camera pictures. One’s likeness was more difficult (though possible) to use for profit before the invention of the camera.

Interestingly, however, we also have with these torts the first example of where an approach that was meant to solve a problem - the problem of new ways of obtaining information - ended up not establishing exactly which problem it aimed to solve and moreover, some claim, ended up losing its way[74]. Critics of this approach to the right to privacy point out that the original torts ended up being diluted, dealing with things only tangentially related to the right to privacy, such as most cases of defamation being ruled under the doctrine of privacy[75]. We propose that part of the reason for this phenomenon is that without a clear understanding of the underlying forces impacting the ebb and flow of information, there is no clear directing line for the development of ways to rebalance those forces. This is the danger we aim to prevent with this thesis, a danger which is likely to materialise in EU law, as we will show.

As we can see, this tort-based approach was mainly built around preventing abuses which would have been impossible or impractical to commit in times past. The same trend continues with Daniel Solove’s subsequent taxonomy, which attempted to set the right to privacy back along the right tracks.

3. Daniel Solove’s Taxonomy

A more modern taxonomy was proposed more recently by Daniel Solove in 2006[76], directly referencing Prosser’s harms and the need for a new taxonomy. Solove’s argument for why that updated taxonomy was needed was mainly the complexification of the legal aspects of privacy[77]. The American vision of privacy is of particular interest to our purposes, because it is spread out over a large variety of case law, statutes - both

[73] Restatement of the Law, Second, Torts, 652 Copyright (c) 1977, The American Law Institute
[74] Kalven Jr, H. (1966).Privacy in tort law--were Warren and Brandeis wrong *Law & Contemp. Probs.* 31 : 326.
[75] Ibid
[76] Solove (n.61)
[77] Ibid

at federal and state-level - and constitutional law. This is the essence of the approach we criticise in this thesis: attempting to solve privacy issues as they appear in various contexts, without a way to adapt to emerging issues, which leads to a situation in which the values to be defended are hard to identify.

Solove attempted to solve this problem with a core taxonomy of privacy harms[78]. As we will see, this approach is focused on the activities that may be problematic for privacy. Our interest is not just in activities harmful to privacy, but also why these activities are considered harmful. However, Solove's approach highlights clearly the direct relationship between a rise in technological forces and the impact on privacy, continuing the long, uninterrupted thread running since Warren and Brandeis' essay.

Solove adds to the mostly "dignitary" harms identified by Warren and Brandeis and codified by Prosser[79] what he calls "architectural" harms, which are assimilated to environmental harms - harms not directly detremental to the person but harming the climate in which their privacy is contained.

Solove's taxonomy regroups four subdivisions of harmful activities: information collection, information processing, information dissemination, and invasion. These activities are arranged around the individual and the impact these activities have on his/her life. Importantly, none of these groups of activities were found in the time of Prosser, but the new technical means which have appeared with the age of Big Data – collecting, processing and disseminating information – have allowed these harmful activities to occur.

This pattern can be found in Solove's exposition of his taxonomy, as historical perspectives for some of the listed harms are provided in his analysis.

a. The Issue of Surveillance

A foremost example in the "information collection" category is the practice of surveillance. Eavesdropping and peeping have been offences for centuries past – the simplicity of the means involved compared with modern surveillance techniques is an illustration of how far technology has come – while laws restricting surveillance by various means of communication were developed as these means appeared. Examples include California

[78] Ibid
[79] Ibid

prohibiting the interception of telegraph communications in 1862, then telephone wiretapping in 1905, with later on federal statutes on the limits of wiretapping during the 20th century[80].

Surveillance, and more specifically surveillance in public, poses an interesting problem for our purposes. In the 1983 case of United States v. Karo[81], the defendant was tracked while driving by the police following a tracker hidden in a can of chloroform. The Court argued that the surveillance "amounted principally to the following of an automobile on public streets and highways"[82], and argued that there was no reasonable expectation of privacy in a public space[83]. The US uses a strict distinction between "public space" and "private space"[84], where the public space holds no particular expectation of privacy[85]. Our argument is that the reason this conception of privacy took hold is due to the fact that it used to be utterly impractical to gather significant private information about individuals in public, while the "home" was the core of privacy. Just as there is no law for the theft of large buildings, there was no law for invasions of privacy in public. However, practical and technological advances have made it so that the American "reasonable expectation of privacy" has disappeared in the modern world[86].

This blurring of the line between "public" and "private" due to technological advances illustrates the difficulties inherent in seeing privacy in such a light. Dealing with this problem using these same frameworks cannot fully handle the problem as we see the rise of forces which had previously been considered inexistent or impossible. This will happen again and again as new technological developments take place, and conceptions focused on the technology itself, or the torts that come from it, will not be able to keep up with those developments.

b. Aggregation

Out of Solove's categories, it is in "information processing" where the most obvious examples of the "technology-driven privacy law" phenomenon are evident. The rise of the first computers in the 1960s led to significant concerns, which led to legislative

[80] Ibid
[81] 460 U.S. 276, 277 (1983).
[82] Solove (n.61)
[83] Ibid
[84] Whitman, J. (2014). The Two Western Cultures of Privacy: Dignity Versus Liberty, *Yale Law Journal*
[85] Reidenberg, J. R. (2014). Privacy in public. *69 University of Miami Law Review 141*
[86] Ibid

scrutiny, with the aggregation of data into one database being the main source of those concerns[87]. The fact that existing information can be collected and aggregated into much more powerful information is the real issue with the aforementioned practice of surveillance: following someone's many daily innocuous acts can lead to information that observing any one act could not.

This has always been true, of course – it has always been possible to link two pieces of data together to create new information, but the extent of that ability was so limited that it was not a force which needed to be reined in by legislation. That changed in the Information Age, with privacy law and data protection law both being spurred to further development due to this change. Once again, it is not a whole new practice which led to an entirely new law: it is the natural development of existing practices.

As Solove shows, aggregation can be seen as a violation of privacy: In *United States Department of Justice v. Reporters Committee for Freedom of the Press*, the disclosure of profiles made from the compilation of various criminal records was held to be an invasion of privacy even where every one of those individual criminal records had already been disclosed. This shows that privacy is not only about data and its disclosure, but also about what information is constructed from that data. Nevertheless, the controversy over treating aggregated information as a separate entity[88] shows that consensus is not set on this matter.

c. Information Dissemination

This brings us to the third of Solove's categories, which includes the most direct "harms", such as breach of confidentiality, blackmail, appropriation (recalling to Prosser's harms) or distortion[89].

Importantly, this is also the category which was changed least by the rise of the Information Age: blackmail has always existed, alongside slander or public humiliation through the disclosure of private facts. This stands in an interesting contrast with the former two categories that Solove identified: unlike information collection and information processing, information dissemination is not a category which has stood untouched by

[87] Solove, D. (2006) A taxonomy of privacy. *University of Pennsylvania Law Review*, 477-564
[88] Ibid
[89] Ibid

legislation until very recently. Indeed, even ancient Romans, for example, had rules about nudity and its public exposure[90].

An interesting example when it comes to the obsoleteness of the public/private distinction is the case of *Daily Times Democrat v. Graham*, in which a woman had her underwear exposed while at a county fair[91], with a picture then taken at that moment and put on the front page of a newspaper. In this case, the newspaper argued that since the picture was taken in a public space, there was no expectation of privacy, which takes us back to Warren and Brandeis' reason for creating their Right to Privacy: the rise of instant photography and mass-disseminated media. This case would not exist except for these technological developments, showing once again the shifting role of privacy. But it also shows tjat whether some information is public or private changes as technological capabilities change, and in the modern age where any information can quickly become very public, the distinction between public and private becomes blurred. More recent examples are easily identified, from the recent case of a woman in Korea whose behavior in a train - refusal to pick up her dog's excrement - led to a severe barrage of personal attacks[92] or the case of Google Street View allowing private movements of individuals to be visible to the whole world[93].

In a world where capturing and disclosing information which has an impact on one's personal life can happen to anyone, anytime, the idea of a "public" sphere is relative. The public sphere only exists until someone turns on a smartphone.

d. Invasion

Invasion is the last group of harms which Solove identified, and includes intrusion upon the individual's private space - which in the American conception is principally the "home"[94]. The presence of one in a private context is different from the other types of harm according to Solove as it includes an element of physical interruption. Invasion highlights that privacy has dimensions stretching beyond simply the creation and

[90] Solove (n.61)
[91] 162 So. 2d 474, 476 (Ala. 1964).
[92] Houghton, D. and Joinson, A. (2010). Privacy, social network sites, and social relations. *Journal of Technology in Human Services* 28.1-2
[93] Segall, J. (2009), Google Street View: Walking the line of privacy-intrusion upon seclusion and publicity given to private facts in the digital age. *Pitt. J. Tech. L. & Pol'y* 10
[94] Solove (n.87)

dissemination of information[95]. This is where the scope of this thesis limits itself to ''informational privacy'', as exploring other dimensions of privacy has been attempted many times with no fully satisfying conclusions[96].

Looking at this harms-based approach to privacy, we can identify some common threads which reveal what the informational dimension of privacy entails. It means identifying ways to prevent information from being acquired (control over information collection), and finding ways to ensure that once information is acquired those controlling it know limits (control over information dissemination). In each case, it can be summed up as clash of forces: forces attempting to gather and use information, against tools attempting to provide control over those forces. The constant emergence of new harms (as shown by the evolution of various taxonomies over time) and the difficulties of quantifying the "harm" in privacy invasions means that the transposition of that privacy conception into law is always catching up with technology[97].

II. Privacy as an Interest: Roger Clarke's Taxonomy

Some definitions of privacy attempt to avoid the difficulties of conceptualising privacy by not referring to it as a "right" (a loaded legal notion), but instead treating it as an "interest". We will go over the groundbreaking conception of Roger Clarke on the matter, and further developments which have stemmed from his analysis, to illustrate how seeing visions of privacy as an interest is consistent with a vision of privacy as a clash between data gathering/usage, and tools to control them.

An important contribution to the scholarship of privacy comes from Roger Clarke, who defines privacy as "the interest that individuals have in sustaining a 'personal space', free from interference by other people and organisations.[98]". This is based on multiple "dimensions", of which Clarke identifies five. We will analyze them, showing the underlying forces expressed through them.

The first dimension is "privacy of the person", or "bodily privacy", which involves the actual human body, such as blood transfusion without consent. The second includes

[95] Zhu, B. (2014). Traditional Tort for a Modern Threat: Applying Intrusion upon Seclusion to Dataveillance Observations*, NYUL Rev. 89 : 2381.*
[96] Koops et al (n.24)
[97] Calo (n.20)
[98] Clarke, R. (2006) What's privacy. *Australian Law Reform Commission Workshop*. Vol. 28.

“privacy of personal behaviour”, such as religious or political habits. The third is “privacy of personal communications”, relating to interception of communications through technological means. Then, “privacy of personal data” includes concerns over making sure personal data is not available to other entities (both public and private) and empowers individuals to have control over such data. Finally, a fifth dimensions, ‘’privacy of personal experience’’, encompassing the everyday experiences that were, until the age of Big Data, never recorded[99].

Just like before with the harms-based approach to privacy, further developments have extended and expanded these categories to adapt to technological developments. The paper “Seven Types of Privacy” by Rachel L. Finn, David Wright, and Michael Friedewald expands the original four categories into seven. Seeing the differences will show the same pattern of increasing privacy-protective forces developed to address new information-gathering and information-using practices.

The first category not found in Clarke’s analysis is “privacy of thoughts and feelings”, which involves tracking behavior to draw conclusions about a person’s thoughts and opinions; the article recognises this aspect of privacy “may be coming under threat as a direct result of new and emerging technologies”[100]. The pattern we have previously established remains true even when seeing privacy as an interest: new interests develop for privacy to be protected, as new possibilities appear. Once again, shifts in the same underlying forces - information gathering and use - prompt a revaluation of existing categories.

The second added category is “privacy of location and space”, which involves being tracked, both in public (monitoring and tracking) and in private (the home). Our previous comments on surveillance in public are similarly applicable: the problem of privacy in public was not prominent enough at the time Roger Clarke was writing in 1997, while the 2013 article identifying this type of privacy follows the rise of surveillance technology: it acknowledges that “This categorisation of privacy was [...] not as obviously under threat when Clarke was writing in 1997, however, this has changed with technological advances”[101].

[99] Gutwirth, S., Leenes, R., De Hert, P., & Poullet, Y. (2013). European data protection: Coming of age. *European Data Protection: Coming of Age*, 1–440. doi:10.1007/978-94-007-5170-5
[100] Ibid
[101] Ibid

The last one is “privacy of association”, which relates to freedom of religion or assembly, which complements perceived gaps in Clarke’s typology when it comes to freedom of religion or assembly - which while an important element, is not in the scope of ‘’informational privacy’’ and shows the necessity of this limited scope. Without it, this thesis would have to expand as far as freedom of expression and freedom of religion.

As we can see by looking at the way US law has attempted to adapt its list of torts to new technological developments, trying to keep up with the various ways in which information can be created or spread by amending and adding to a taxonomy of torts is difficult. Instead, we propose to base the interest of informational privacy on a balance between the ability of data controllers to create and disseminate information, and the guarantees against them doing so. Nevertheless, these taxonomies bring something very important to the scholarly discussion: they take the shifting, subjective views surrounding privacy and lay them out as a comprehensive guide.

III. Helen Nissenbaum’s Contextual Integrity: A New Way of Thinking about Privacy

After looking at taxonomies of privacy and identifying the patterns of new dimensions of privacy appearing and requiring revisions of these taxonomies, we will now analyze another type of privacy conceptualisation which focuses on some of the underlying forces we have identified: Helen Nissenbaum’s contextual integrity.

In Helen Nissenbaum’s theorized framework of “contextual integrity”[102], privacy consists of appropriate flows of information: instead of focusing on the data itself, the focus being on whether the flow is correct for the context in which it is shared[103]. As such it is possible to identify the appropriateness of the flow of information without focusing on the data itself, but also on relevant outside factors[104].

An example is the doctor–patient relationship: in such a context, individuals confide very personal information to their doctor, but the social and legal context binds the doctor from sharing that information, while a similar obligation does not bind the patient, if the doctor were to confide similarly personal information to them[105].

[102] Nissenbaum, H. (2004). Privacy as contextual integrity. *Washington law review* 79.1.
[103] Ibid
[104] Ibid
[105] Ibid

"Contextual integrity" allows for separating the privacy concerns from the data itself, by focusing instead on appropriate information flows, with a breach occurring where those flows are not appropriate. This vision of privacy as flows of information is effective when taking into account the changing landscape of the personal data processing field, as it is adaptable to technological changes due to always retaining the same elements: a sender, a receiver, a message, and a referent of the message. However, this definition, in and of itself, does not posit a justificatory framework as to where contextual integrity needs to be maintained or not. Nissenbaum has addressed it by looking at US law and identifying three principles of privacy law arising from it - "(1) limiting surveillance of citizens and use of information about them by agents of government, (2) restricting access to sensitive, personal, or private information, and (3) curtailing intrusions into places deemed private or personal."[106].

In that way, contextual integrity creates a basis on which it can be applied. We build on this theory in this work by focusing on the ''informational privacy'' interest, instead of on the wider notion of ''privacy''. This limited scope will allow us to set aside the various other interests covered under the umbrella of privacy and focus on a particular interest which, as we will argue, is key to addressing the challenges of the Information Age.

IV. Understanding ''Informational Privacy'': An Interest Linked to Data Protection and Privacy

The scope of this work is limited to one specific interest, ''informational privacy''. As we have shown so far, no conception of privacy has fully managed to encompass all of the various interests of ''privacy'', and as we will show in the next chapter of this thesis the same is true for ''data protection''. Instead, we will focus on this particular interest. This is done for a number of reasons: first, as we will show, informational privacy is an interest that has been heavily disrupted in the age of Big Data, which has led to other interests being also impacted. As such, protecting informational privacy protects those interests as well. Second, informational privacy sees an overlap between privacy and data protection law, allowing for a study of both these rights through the lens of a single interest. Third, informational privacy, as we will show, is a relatively unique interest in that it is based on a balance of ''availability'' of information and not on any specific ''harms'' that may arise from its breach. Because of all those factors, we argue that

[106] Ibid

protecting this interest, using a balance between the means of obtaining information and the forces which prevent that information from being obtained, is vital in handling the challenges brought about by the age of Big Data.

This idea of ''informational privacy'' as an interest is far from new[107] and has been described and defined in a number of ways[108]. This interest is so central that some have claimed that it is synonymous to privacy as a whole[109]. Nevertheless, as we will see, its definition has remained muddled[110].

There have been multiple attempts to map out the various interests covered under the right to privacy[111]. The one we will make most reference to in this thesis will be the typology of privacy elaborated by Bert-Jaap Koops and a team of scholars[112]. According to this typology, the various interests covered under the umbrella of ''privacy'' are arranged under two axes - one going from the ''personal space'' to the ''public zone'', and one going from placing the emphasis on the ability to be left alone to placing it on the freedom towards self-development[113].

Though not aimed at being a perfect and all-encompassing classification, it contains eight primary ideals of privacy. Bodily privacy (interest in the privacy of one's physical body), spatial privacy (interest in having a private space), proprietary privacy (image and reputation), intellectual privacy (privacy of thought and mind), decisional privacy (the ability to make intimate decisions), associational privacy (interests in being free to choose who one's close associates are), and behavioural privacy (privacy while conducting publicly visible activities). This typology highlights how the European approach to privacy is broad and encompasses a large number of interests. Privacy includes the right to respect for one's business premises [114] or the right to have one's

[107] Floridi, L. (2006). Four challenges for a theory of informational privacy. *Ethics and Information technology*, *8*(3), 109-119.
[108] Tavani, H. T. (2008). Informational privacy: Concepts, theories, and controversies. *The handbook of information and computer ethics*, 131-164.
[109] Parent, W.A. (1983).Privacy, morality and the Law. *Philosophy & Public Affairs*, *12*(4), 269-288.
[110] Cohen, J. E. (1999). Examined lives: Informational privacy and the subject as object. *Stan. L. Rev.*, *52*, 1373.
[111] Tavani (n.108)
[112] Koops et al (n.24)
[113] Ibid
[114] ECtHR, Société Colas Est and others v. France, Application no. 37971/97, Judgment of 16 April 2002.

gender identity rectified in public records[115], thus touching on a wide scope of rights and interests.

This typology makes informational privacy a separate interest altogether, an overarching aspect of each type, ''typified by the interest in preventing information about oneself from being collected and in controlling information about oneself that others have legitimate access to.''[116] Since each element of privacy contains some element of restricting information about oneself, every privacy aspect contains informational privacy. This also means that if a technological innovation impacts informational privacy every aspect of privacy is also impacted. As we will show, the biggest transformation of the Big Data era is how information is created and shared, and the profound impact on privacy is the result of this transformation impacting primarily informational privacy.

One theory of informational privacy has been called the ''restricted access theory'', which claims that one has informational privacy when one is able to limit or restrict others from access to their personal information[117]. This approach was developed in the US, and is based on ''zones'' where there is a distinction between ''public'' and ''private'' spaces (a notion we will approach in Chapter 4, showing its limitations). Another is based on ''control'', the ability to control who has access to information about us and the ability to create and maintain various relationships[118]. This echoes the typology we previously analysed, showing that informational privacy is based on information control, and through it touches on every part of privacy.

One of these theories includes the aforementioned ''privacy as contextual integrity'' by Helen Nissenbaum, based on appropriate flows of information. Beyond this work, another ''benchmark'' approach is made by Luciano Floridi: an ''Ontological Interpretation'' of informational privacy, based on the forces which limit the availability of personal information, arguing that privacy is ''nothing less than the defence of the personal integrity of a packet of information''[119], arguing that a breach of one's ''informational sphere'' is a breach of one's personal identity. The limitation of Floridi's approach is conflating informational privacy with other types of privacy - linked to identity

[115] See cases such as Rees v. UK (ECtHR, Rees v. UK, Application no. 9532/81, Judgment of 25 October 1986), Cossey v. UK (ECtHR, Cossey v. UK, Application no. 10843/84, Judgment of 27 September 1990, Series A, No. 184), B v. France (ECtHR, B v. France, Application no. 13343/87, Judgment of 25 March 1992) or Goodwin v. UK (ECtHR, Christine Goodwin v. the United Kingdom, Application No. 28957/95, Judgment of 11 July 2002) for examples of such cases.
[116] Koops et al (n.24)
[117] Tavani (n.108)
[118] Ibid
[119] Floridi (n.107)

or autonomy - which brings in a group of interests that are difficult to untangle[120]. Nevertheless, it allows oneto view information as a balance of forces, which is an idea that inspires this thesis.

Informational privacy is based on the ability to create and disseminate information about individuals. Information being obtained about individuals by large entities is not new - from William the Conqueror creating a ''Domesday Book'' including data about every one of his subjects[121] to census record kept by governments as far back as the Roman era[122]. When that ability is too widespread, it impacts other interests of privacy: for example, instant photography at the time of Warren and Brandeis alongside the rise of tabloids had an impact on associational privacy (information on who one spent one's time with was easier to record and disseminate), as well as proprietary privacy (information on what one owns and prefers is also easier to obtain and share), amongst others. In fact, associational privacy, as an interest, did not change. As a societal value, being able to choose freely with whom one associates without the judgment or knowledge of others was as important before instant photography as it was after it, or as it is now in the 21st century. As we will see in Chapter 3, this is true of many interests of privacy, which shows that not only is informational privacy the one interest which overlaps all others, but that fact also makes it a core source of the challenges both privacy and data protection are facing.

Because the creation and sharing of information is easier than it has ever been, various entities - companies, individuals, governments - have been able to obtain vast amounts of information about individuals in an ever-expanding fashion. Meanwhile, means to limit what these entities can do with that information have not changed significantly. This has led to an imbalance, between the information that these entities are able to obtain and share, and the guarantees against them being able to do so. As we have seen throughout this chapter, and as we will develop further throughout this thesis, there is a pattern showing that when the balance between the information that can be created and the guarantees on that information becomes imbalanced, new guarantees will appear to restore the balance. This, we will show, is the core of what informational privacy is: the existence of sufficient guarantees to limit the information about individuals which can be created and shared.

[120] Tavani (n.108)
[121] Ibid
[122] Shank, R. (1986). Privacy: History, legal, social, and ethical aspects. In Library Trends 35 (1) Summer 1986: Privacy, Secrecy, and National Information Policy: 7-18

We will organize these guarantees alongside a taxonomy developed by Lawrence Lessig[123], whose theory identified four types of forces which affect behaviour: Law, Social Norms, Market and Architecture (in this case, “Architecture” being the technological environment). The use of this taxonomy is based on two main reasons: first, to allow an examination of those various forces and how they interact (for example, some technological developments like the easily-disseminated tabloids of Warren and Brandeis do not allow any new information-gathering, but made it cheaper to spread that information, thus incentivizing a breach of informational privacy by making it profitable), and second to allow a framework in which every situation which involves informational privacy, in any context or time period, can be examined using a common benchmark.

Out of the control mechanisms which are able to control information flows, the one we have focused the most on in this analysis is the legal mechanism. The GDPR, for example, uses legal mechanisms to deal with specific practices. For example, it includes special provisions for data profiling[124], targeting a specific practice with restrictions, in this case decisions based solely on automated processing.

But these forces can also be technological - for example, automatically compiling customers’ purchases into one database and automatically creating full profiles from dozens of different data sources are both “automated processing operations”, but one has less technological capability for information creation. As such, one needs fewer legal control measures to be balanced.

These controls can also be social. Relying on social trends is always uncertain, as they can change over time, however social norms play an important role of regulating behavior parallel to legal ones. There is no law against revealing someone’s secret, but social norms ensure that it is possible to share information with a reasonable guarantee that it will not be revealed to those outside your confidence.

Finally, these controls can also be economic, that is, Market-oriented. Paparazzi are more interested in the lives of celebrities than regular people because information about their private life is more valuable, while data only is gathered and analyzed in a privacy-intrusive way when it became a valuable commodity. By increasing or reducing incentives to gather this information (including making it more or less difficult to obtain or process), the information can be protected. Controlling these incentives is as important as stemming it through legal or social means.

[123] Lessig (n.36)
[124] GDPR, Article 22

The framework developed in this thesis takes from Helen Nissenbaum's contextual integrity, Lessig's's taxonomy, Floridi's ontological interpretation of informational privacy, and Bert-Jaap Koops' typology of privacy to formulate informational privacy thus:

''Informational privacy is preserved when the Information relating to individuals which can be obtained and disseminated is bound by sufficient Guarantees.''

This definition highlights what we aim to achieve with this conception: protecting informational privacy through assessing whether information is as available as it should be, and proposing ways of limiting such availability. As we will see in later chapters, whether information ''should be'' available is reliant on a further hypothesis of this thesis, which is that the various interests privacy protects (and which depend on informational privacy) have kept their societal value over time, and will continue to do so.

We will now develop an overview of how privacy, and especially informational privacy, was conceptualised in the legal system around which this thesis will be developed: EU law. This will guide our analysis as well as start to introduce the role of data protection and how it interacts with privacy in protecting the interest of informational privacy.

V. Sources of Informational Privacy in EU law: Privacy and Data Protection

Before studying the right to data protection and how it aims to protect informational privacy, the relationship between privacy and data protection requires some analysis. Indeed, we assert that the core overlap between those two rights is informational privacy.

The EU authorities have not been forthright about why the right to data protection was enshrined in the Charter of Fundamental Rights to the European Union's Article 8[125]. Nevertheless, the idea that privacy and data protection should remain steadfastly separate has been expressed by the academic community (such as Roger Brownsword stating that the link between the two needs to be "clean and clear"[126]), and the relation

[125] Charter of Fundamental Rights of the European Union, 26 October 2012, 2012/C 326/02, Article 8

[126] Brownsword, R. (2009) Consent in Data Protection Law: Privacy, Fair Processing, and Confidentiality' in *Reinventing data protection?* by S. Gutwirth, Y. Poullet, P. de Hert, & C. de Terwangne (Eds.). Dordrecht: Springer.

between privacy, data protection and informational privacy was held to be "not commonly understood"[127].

Some models base the relationship between the two on being aimed at protecting "human dignity" - with an early 1983 German case making that link[128], then backed by the Court of Justice of the European Union expanding the applicability of Article 8 to protect human dignity in cases related to the patentability of body parts[129] or a human embryo[130], or the prohibition of a laser game simulating killing[131]. Nevertheless, the right to human dignity remains relatively vague in the Big Data era[132], covering a number of subsequent rights and interests, including not just privacy and data protection but the right to life and to free expression[133], as shown by the fact that the ECtHR held that the State should have a large margin of discretion when applying that right[134].

Another vision is data protection simply as one element of privacy, as a new evolution in protecting a "right to be left alone", and only aimed at protecting privacy in the digital age. However, there are multiple privacy interests which have only a tenuous link to data protection such as a person's control over their own body[135]. Instead, data protection is an independent right, which serves a number of purposes. As observed by Orla Lynskey in "The foundations of EU data protection law": "While it is often accepted that data protection serves a multitude of purposes in addition to protecting informational privacy, there has been insufficient effort to identify and explain the rationale of these objectives."[136] In other words, though both the right to privacy and the right to data protection have a number of goals and interests that they seek to protect, one area where the overlap is most prominent is informational privacy. This is where our analysis will focus.

[127] Lynskey, O. (2015). The foundations of EU data protection law. *Oxford University Press.*
[128] Judgment of 15 December 1983, 1 BvR 209/83, BVerfG 65, 1.
[129] Case C-377/98 Netherlands v Parliament and Council [2001] ECR I-7079.
[130] Case C-34/10 Oliver Brüstle v Greenpeace eV [2011] ECR I-09821
[131] Case C-36/02 Omega Spielhallen- und Automatenaufstellungs-GmbH v Oberbürgermeisterin der Bundesstadt Bonn [2004] ECR I-9609.
[132] Caulfield, T., & Brownsword, R. (2006). Human dignity: a guide to policy making in the biotechnology era?. *Nature Reviews Genetics, 7(1), 72.*

[133] Lynskey (n.127)
[134] Evans v United Kingdom (2006) 43 EHRR 21
[135] Lynskey (n.127)
[136] Ibid

Meanwhile, case law from the ECHR also links together data protection and Article 8[137]. An example can be found S. Marper v. the United Kingdom[138], in which the judgment stated that "The mere storing of data relating to the private life of an individual amounts to an interference within the meaning of Article 8 of the European Convention on Human Rights"[139]. Meanwhile, there are other cases such as Ben Faiza v. France[140] where placing a real-time geolocation device on one's vehicle by the police constituted a breach of Article 8 of the ECHR, L.H. v. Latvia[141] where the collection of one's personal medical data by a State agency without consent was a violation of Article 8, and Malone v. the United Kingdom[142] in which interception of postal and telephone communications by the police was considered a violation of Article 8. All these show that there is a strong link between the right to private life and informational privacy.

In conclusion, "informational privacy" is one of the interests protected by EU law, using primarily a combination of privacy and data protection, as evidenced by the use of data protection and Article 8 of the ECHR to restrict the creation, dissemination and use of information. That interest has the aim of protecting individuals by restricting the information collected and processed by data controllers, and is enforced by a number of rights restricting different information-related practices[143], and implementing safeguards. A common criticism of the GDPR is that it uses tools which cannot follow the progression of technology and might hamper innovation[144]. We argue that this is because of a conception of informational privacy based on some assumptions about data and information that are being challenged in the Big Data era. Nevertheless, as we will show, the GDPR is showing signs of accepting a new framework, going beyond a binary definition of ''personal data'' and towards a contextual approach.

[137] De Hert, P., & Gutwirth, S. (2009). Data protection in the case law of Strasbourg and Luxemburg: Constitutionalisation in action. In *Reinventing data protection? (pp. 3-44).* Springer, Dordrecht.

[138] S. and Marper v. the United Kingdom [2008], Applications no. 30562/04 and 30566/04, ECHR 1581 § 67

[139] S. and Marper v. the United Kingdom [2008], Applications no. 30562/04 and 30566/04, ECHR 1581 § 67

[140] Ben Faiza v. France [2018], Application no. 31446/12, ECHR 153

[141] L.H. v. Latvia [2014], Application no. 52019/07, ECHR 515

[142] Malone v United Kingdom [1984], Application no. 8691/79, ECHR 10

[143] Rotenberg, M. and Jacob, D. (2013), Updating the law of information privacy: the new framework of the European Union, *Harvard Journal of Law & Public Policy, Vol. 36 Issue 2*

[144] Hildebrandt, M. and Tielemans, L. (2013) Data protection by design and technology neutral law. *Computer Law & Security Review* 29.5509-521.

Conclusion: Informational Privacy as a Self-contained Interest

"Privacy" and "Informational Privacy" are not synonymous. The former is a wide bundle of interests touching on various issues without unifying definition, but is protected under Article 8 of the ECHR. The latter is a more specific interest enshrined in European Law through, primarily, some uses of Article 8 and through Data Protection legislation.

As we have seen through studying different privacy conceptions, all of them, starting with Warren and Brandeis, have in common the idea of controlling the creation, dissemination, and use of information, whether it means addressing the appearance of widely distributed gossip columns in the 19th century or the NSA collecting data on individuals in the 21st century. Whether we are talking about surveillance in public, intrusion in private, or social media tracking somewhere in the middle, tools to gather information are present, and are limited by control mechanisms, constructed to limit these tools as they develop.

This interest is "informational privacy": controls on the creation, dissemination, and usage of information affecting individuals. As we will show, the GDPR has introduced new tools which affect these different controls, showing signs of moving towards the approach developed in this thesis.

Chapter 2. Informational Privacy and the Right to Data Protection

Introduction

This chapter is intended to show European data protection regulation through the prism of "informational privacy", and the declared goals and philosophy of the regulatory framework. An understanding of this framework is necessary in order to understand fully how it currently protects informational privacy, and the challenges to that protection brought about by the age of Big Data.

Indeed, "data protection", as a notion separate from "privacy", is a young right, only a few decades old[145]. Its appearance, as we will show, is due to the same changes that have driven the transformation of privacy. Hence, we will argue that data protection was created simply as being able to regulate behaviours in a way that privacy law could not, primarily because of the complex scope of privacy described in Chapter 1. Therefore, we will only analyse data protection within the bounds of the same concept of "informational privacy" within which we analysed privacy. Data protection covers a number of interests unrelated to privacy (including self-determination[146], no-discrimination[147] or competition[148]), though interests that go beyond informational privacy are outside the scope of this thesis.

In order to understand the European vision of data protection, it is necessary to understand the roots of that right and its goals. Through an analysis of the European data protection regime, we will show that though the GDPR attempts to avoids any question of "privacy" , it eventually recognizes a link between the two rights. We will also show that the European legislators have started using new tools, such as ones creating indirect social and economic pressure, to safeguard the right to data protection. This will highlight the fact that a multi-disciplinary approach, going beyond purely legal tools, is

[145] Bennett (n.5)
[146] Hornung, G., & Schnabel, C. (2009). Data protection in Germany I: The population census decision and the right to informational self-determination. *Computer Law & Security Review*, *25*(1), 84-88.
[147] Schreurs et al (n.16)
[148] Kerber, W. (2016). Digital markets, data, and privacy: competition law, consumer law and data protection. *Journal of Intellectual Property Law & Practice*, *11*(11), 856-866.

both an effective way to protect informational privacy, and an approach that the EU has already starting moving towards.

A. The Development of EU Data Protection Law: A Young Right

Though the keeping of records by governments of information about individuals is ages old[149], the concept of "data protection" appeared at the same time as the modern understanding of "data", in the second half of the XXth Century[150].

The main factor in the rise of Data Protection is naturally the rise of data-processing systems coming with the age of computing, in order to implement checks and balances to limit these new tools. However, this does not account for all of the development in the field. The rise in individual rights, especially against the State, after World War II has affected privacy, and later data protection.

Data Protection as a means to protect privacy takes its roots from Article 8 of the European Convention on Human Rights[151], which states that "Everyone has the right to respect for his private and family life, his home and his correspondence."

Data Protection arose in Europe with the influence of the Organization for Economic Cooperation and Development (OECD) and the Council of Europe, starting in the 1980s. The OECD established non-binding guidelines in 1980, holding principles which were later built on by subsequent legislation. They include that data should be accurate, that the purpose of collection should be specified, and that the data gatherer has a responsibility against the access, destruction, modification or dissemination of that data[152]. The Guidelines also provided for transparency, including the right to know what kind of data was being collected, and the right that the individual correct or erase that data. A familiar addition, which is the source of a maelstrom of conflict in the field of data

[149] Roch, M. P. (1996). Filling the Void of Data Protection in the United States : Following the European Example. *Santa Clara High Technology Law Journal*, *12*(1), 71–96. Retrieved from http://digitalcommons.law.scu.edu/chtlj/vol12/iss1/3
[150] Bennett (n.5)
[151] Article 8, ECHR
[152] Organization for Economic Co-operation and Development: Council Recommendation Concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data, reprinted in 20 I.L.M. 422 (1981) [hereinafter OECD Guidelines].

protection[153], is the limitation of cross-border flows of data to countries which have adequate existing protections[154].

These principles were quite vague, showing how the development of specific data protection provisions took some time to be fleshed out, though the basic principles the EU used have remained unchanged: top-down regulation limiting practices involving data, with some goal of user control and transparency.

Meanwhile, the Council of Europe, an entity created with the promotion of human rights in mind, began to study data protection legislation in 1968[155], and passed a resolution to protect "personal data" in 1974. Similar principles to the OECD can be observed, such as purpose limitation, rules on creation and dissemination of data, and transparency towards individuals[156]. Of particular interest is the fundamental place "personal information" holds in this conception, a structural decision which made sense then but does not now. These guidelines became the first binding data protection instrument in Europe as the 1981 Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data[157].

That Convention applies to both the public and private sectors - showing the European interest in focusing on the individual and those infringing on his dignity-based privacy interests, wherever they come from. Aside from the accuracy, purpose limitation, and legitimacy issues seen in the previous works, it altogether prohibits the gathering of information related to race, religion, health, and other matters considered particularly private[158]. This division of data based primarily on certain factors pertaining to the data itself in a vacuum is apparent from the first data protection instrument, which as we will show is the source of a growing problem.

Because the Convention allowed States to implement it as they pleased, harmony between European nations was not complete, which led to frictions and obstacles due to the cross-boundary transfer provisions not allowing such transfers if protection

---

[153] See *Schrems v. Data Protection Commissioner* (No.2), [2014] IEHC 351 (2014).
[154] Hoepman, J. H., Hubbers, E., Jacobs, B., Oostdijk, M., & Schreur, R. W. (2006, October). Crossing borders: Security and privacy issues of the european e-passport. In *International Workshop on Security* (pp. 152-167). Springer, Berlin, Heidelberg.
[155] Cole, P. (1985). New Challenges to the U.S. Multinational Corporation in the European Economic Community. *Data Protection Laws, 17 N.Y.U. J. Iirr'rL L. & POL. 893, 898 n.30*
[156]Roch (n.149)
[157] Council of Europe, Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, Jan. 28, 1981, Eur. TS. No. 108
[158] Ibid, Article 6

requirements between the two States were uneven[159]. As we will see, harmonization of data protection norms is possibly the most conflictual aspect of European data protection, especially in opposition with the US.

Not all data protection legislation at the European level came from supra-national entities. Sweden and West Germany were the first States to implement comprehensive data protection rules. Sweden started implementing data protection norms in 1973 with the Swedish Data Bank Statute, requiring governmental authorization to collect personal data, and a Data Inspection Board enforcing the statute and monitoring compliance[160].

In West Germany, privacy was enshrined in the Constitution after World War II which led to a first Data Protection Law in 1970, though it was limited to one State and only to the public sector[161]. In 1977, West Germany developed a precursor to the Council of Europe at the national level, the Bundesdatenschutzgesetz which had a wider scope and protected "natural persons", containing similar protections as the Convention, and including a supervisory body.

These national examples show us the trend in European countries: a unified data protection legislation, protecting individuals, supported and enforced by a supervisory agency. As we will see, all of European data protection regulation since has followed in the footsteps of this model.

These efforts coalesced in 1990, when the European Commission drafted three proposals to protect personal data rights. The first dealt with personal data collection, the second with electronic data networks, and the third with, specifically, information security[162].

The 1992 Amended Proposal on Data Protection acknowledges the conflict between the interests of data privacy and free flow of personal data[163], such as in order to protect freedom of the press. This limitation highlights the real focus of data protection: commercial and governmental use of data[164].

[159]Roch (n.149)
[160] Cole (n.155)
[161] Ibid
[162] Roch (n.149)
[163] European Commission, Amended Proposal for a Council Directive on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, 1992 OJ. (C 311) 4 [hereinafter Amended Proposal].
[164] Roch (n.149)

A recurring theme, which took a back seat with more recent developments, was a focus on “data quality”: the requirement that data must be kept current, updated, and accurate, with a burden on the data controller or processor. This has the side effect of encouraging limiting the data kept by data controllers, to avoid the burden of updating mountains of irrelevant data[165]. This links back to the original torts identified with misrepresentation or false light, which mainly grew out of an economic concern but evolved in the European context to be closer to a moral concern, of transparency and control over one’s identity and how it is perceived by others.

The development of a primarily rights-based European approach can be seen in the development of such control measures for individuals. Instead of a torts-based approach such as the one used in the US, European development of legislation is focused on establishing a high standard of rights, starting with the ones in the Amended Proposal.

The Network Draft designed by the European Commission in 1990 was intended to complement the Amended proposals in the area of telecommunications - adding extra regulatory constraints to that industry[166]. This continues the pattern of addressing problematic new developments with new constraints as they appear, with the usual latency that legal action takes. By the time the Data Protection Directive[167], based on these legal developments, came into effect in 1995, new types of technologies - in particular, the Internet, and later the explosion in digital storage leading to the phenomenon of Big Data - were already in their early stages. As we will show, one of the main limitations of the Directive was that it focuses too much on “personal data”, while developments in the field of Big Data have created new concerns which go beyond simply the “personal data” framework that has been in development since the OECD’s Guidelines. Nevertheless, we will show that the creation of ‘’Data Protection’’ as a separate, whole right is a legal innovation which has allowed new solutions to the challenges of the Information Age[168].

These developments mirror the evolution seen in the previous chapter: both privacy and data protection have had their major transformations directly related to new ways of processing data. As we will see, the latest development in EU data protection law, the

[165] Ibid
[166] Ibid
[167] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Off. J.L. 281 (Nov. 23, 1995) ("Data Protection Directive")
[168] Safari, B. A. (2016). Intangible Privacy Rights: How Europe's GDPR Will Set a New Global Standard for Personal Data Protection. Seton Hall L. Rev., 47, 809.

GDPR, has implemented new tools which show a move towards a new type of approach: one based on protecting informational privacy through compensating new tools giving more power to data controllers with guarantees limiting that power.

As a new legal right, data protection's scope has increasingly widened over time, both through the legal instrument mentioned above, but also through case law. The European Court of Human Rights has developed a number of such cases where the protection of personal data was an important element, especially regarding state surveillance, such as in cases of tapping phone conversations [169] or phone metadata[170].

B. Innovations in EU Data Protection Law: First Hints of a Balancing Approach

In the previous section we have shown how European data protection evolved to become the developed right it is today. In this section, we will show instances where the European data protection authorities have accepted the need to foray outside the bounds of the traditional means of protecting individuals, and how this is indicative of a shift towards the balance-based framework developed in this thesis. We will show that the new rights and principles created by the GDPR heavily rely on taking into account the context of the data processing, and the risk it presents to various interests of the data subject. This shows that even though attempts have been made to distance the right to data protection from other rights, especially privacy, the two are still strongly linked.

The General Data Protection Regulation came into effect on the 25th of May 2018. Replacing the outdated Data Protection Directive, it introduces a new and expanded set of rules for promoting individual control over data and for increased transparency, replacing the European framework on data protection previously enshrined mainly in the 1995 Data Protection Directive [171].

[169] See ECtHR, Klass v. Germany, Application no. 5029/71, Judgement of 6 September 1978, ECtHR, Amann v. Switzerkand, Application no. 27798/95, Judgement of 16 February 2000, or ECtHR, Halford v. United Kingdom, judgment of 25 June 2997, Reports, 1997-III

[170] See ECtHR, Malone v. United Kingdom, Application no. 8961/79, Judgment of 2 August 1984, ECtHR, P.G. and J.H. v. the United Kingdrom, Application no. 44787/95, Judgment of 4 May 2000, or ECtHR, Copland v. the United Kingdom, Application no. 62617/00, Judgment of 23 April 2007

[171] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Off. J.L. 281 (Nov. 23, 1995) ("Data Protection Directive")

In order to analyse the parts of the GDPR which show evidence of moving away from the ''personal data'' binary, and towards a ''balance-based'' framework, we will first study some new rights which signal the existence of this new way of conceptualizing informational privacy, and will then go over the ways in which the GDPR, in its attempts to expand its influence to a global scale, has opened itself up to using these forces to achieve its goals. In other words, we will show that where the Law is unable to protect informational privacy, other Guarantees come in - a core hypothesis of this work. These measures relying on new forces notably either first appeared with the GDPR, or were strongly reinforced by it. This highlights the two frameworks the GDPR is currently moving between: an instrument built on pillars that are becoming problematic, while attempting to implement new tools that can deal with emerging problems in innovative ways.

I. New Rights and Principles in the Age of Big Data: Contextual Tests and the Accountability Principle

1. Rights against Automated Individual Decision-making

A core new right that was created by the GDPR is Article 22 on automated individual decision-making[172]. According to Article 22:

"Member States shall provide for a decision based solely on automated processing, including profiling, which produces an adverse legal effect concerning the data subject or significantly affects him or her, to be prohibited unless authorised by Union or Member State law to which the controller is subject and which provides appropriate safeguards for the rights and freedoms of the data subject, at least the right to obtain human intervention on the part of the controller."[173]

The second paragraph in Article 22 then moves on to adding special requirements where the profiling is done based on special categories of personal data (namely, "suitable measures to safeguard the data subject's rights and freedoms"), with paragraph 3 prohibiting, specifically, "profiling that results in discrimination against natural persons on the basis of special categories of personal data".

[172] GDPR, Article 22
[173] GDPR, Article 11

The right to non-discrimination exists in EU law and was already well-established since well before even the Data Protection Directive, being referenced at Article 14 of the European Convention of Human Rights[174], Articles 18-25 of the Treaty on the Functioning of the European Union[175], and Article 21 of the Charter of the Fundamental Rights of the European Union[176]. Individual profiling is inherently challenging to informational privacy because it is a practice which is inherently discriminatory[177] in the fact that it consists of making assumptions about people based on certain characteristics. Because any characteristic about a person is, by definition, "information relating to" that individual, profiling is inherently an activity covered by data protection. Particularly, profiling is becoming more prevalent as more data about individuals is captured[178].

Of course, banning all profiling is not practical. Instead, the focus is on profiling which has an "adverse legal effect" on the data subject or "similarly significantly affects him or her". This part, particularly, shows the EU's wide interpretation of the interests data protection is meant to protect. The Article 29 Working Party attempted to clear up this definition in an opinion paper[179], mentioning child or housing benefits, or being subjected to increased surveillance measures, as potential legal effects[180]. This was then expanded by guidance from the Irish data protection authority on data protection impact assessments[181], which listed factors which may be considered to ''significantly affect'' data subjects, including ''being charged more for good or services than he or she would otherwise be charged'', ''being refused unemployment or access to good or services'', or ''loss of entitled to a particular social benefit conferred by law, such as child or housing benefit''[182].

[174] ECHR, Article 14
[175] Consolidated Version of the Treaty on European Union art. 22, 2010 O.J. C 83/01
[176] Charter of Fundamental Rights of the European Union, 26 October 2012, 2012/C 326/02, Article 21
[177] Goodman (n.28)
[178] Koops, B. J. (2014). The trouble with European data protection law. *International Data Privacy Law*, *4*(4), 250-261.
[179] Article 29 Working Party (2017), Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation, WP251, Adopted on 3 October 2017
[180] Wachter, S and Mittelstadt, B, and Floridi, L. (2016), Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation (December 28, 2016). *International Data Privacy Law, 2017.* Available at SSRN: https://ssrn.com/abstract=2903469 or http://dx.doi.org/10.2139/ssrn.2903469
[181] Comisiún Cosanta Sonraí (2018), Draft list of types of Data Processing Operations which require a Data Protection Impact Assessment
[182] Ibid

The Article 29 Working Party's Opinion opens up a wide variety of possible factors by stating that: "Processing that might have little impact on individuals generally may in fact have a significant effect on certain groups of society, such as minority groups or vulnerable adults. For example, someone in financial difficulties who is regularly shown adverts for on-line gambling may sign up for these offers and potentially incur further debt." If this definition can apply to a characteristic to which anyone is potentially vulnerable to (such as being in financial difficulties), it can apply to any kind of difficulty that affects one's behavior. The line is quite blurry: if one is not meant to target people in financial difficulties with gambling advertisements, what kind of financial difficulties qualify? This is an assessment and an approach that heavily relies on context, and heavily relies on assessing the ability the data controller has to make decisions which ''significantly affect'' the individual. Despite the fact that data protection has been identified as a separate right in its own right, isolated from the difficulties of defining "privacy", context still has to be taken into account, and part of the context is whether the data processing is affecting the right to privacy of the data subject.

2. The Right to be Forgotten

The right against automated decision-making is not the only new right which is heavily reliant on context. Another right, the "right to be forgotten"[183], has created challenging questions as to where data should and should not be kept.

The "data minimisation" principle stipulates that only the minimum amount of data need be kept to accomplish the purposes of processing[184]. This principle requires that only "relevant" data be processed, leading to instant conflicts about what "relevant" means[185].

The discussion over what is or is not "relevant" was inflamed by the Google Spain case[186], in which an individual argued successfully that search engine giant Google could be compelled to remove from its search results facts which "appear to be inadequate, irrelevant or no longer relevant or excessive in the light of the time that had elapsed."[187]

[183] GDPR, Article 17
[184] GDPR, Article 5
[185] Ausloos, J. (2012). The 'right to be forgotten'–worth remembering?. *Computer Law & Security Review*, *28*(2), 143-152.
[186] Case C-131/12, Google Spain, SL, Google Inc v Agencia Espanola de Proteccion de Datos (2014)
[187] Ibid

This statement constitutes direct evidence that the status of data changes with context. In this case, an individual's criminal record could de-indexed from Google after a certain amount of time - so that the person can live a life beyond their history, hence the name "right to be forgotten"[188].

This judgment has important implications, which include freedom of the press, public security, and deeper questions about individuality and change. The idea that time can allow personal data to be more or less ''relevant'' implies that factors beyond the data itself, the context surrounding the processing of this data, changes its relevance. Time is just one such factor[189], in fact any amount of factors about the data can change the rights associated with it. If age is relevant, what about other contextual elements? In the same way that we discussed previously with the "similarly affects him or her" element of Article 22, the ability to look at any relevant factors is also opening the door to a fully-contextual approach. Does information about the person's public involvement with political activists, an individual's shopping list or the car they once drove also become irrelevant as time goes by? Is it only for purposes that society frowns upon and thus might affect the individual in undue ways, and in which case how far does this go?

Data contextuality seems bound to be an element to be taken into account by attempts to regulate it, and so accepting those factors in legal decisions ends up being necessary, indicating that the GDPR is showing signs of accepting this contextual approach.

Another aspect where the new provisions of the GDPR touch on new tools to use in the goal of protection for individuals can be found, not in the new rights, but in a new principle: the principle of accountability.

3. Accountability: Reinforcing data protection compliance social norms

The principle of accountability requires a level of awareness and transparency throughout the whole organisation of the data controller, and the ability to demonstrate

[188] Ambrose, M. L., & Ausloos, J. (2013). The right to be forgotten across the pond. *Journal of Information Policy*, *3*, 1-23.
[189] Korenhof, P., Ausloos, J., Szekely, I., Ambrose, M., Sartor, G., & Leenes, R. (2015). Timing the right to be forgotten: A study into "time" as a factor in deciding about retention or erasure of data. In *Reforming European data protection law* (pp. 171-201). Springer, Dordrecht.

that the principles of the Regulation are being followed[190]. This new principle, which did not exist in the Data Protection Directive, has entirely changed the way data protection regulation is approached by organisations: making compliance about being accountable, direct, open[191]. Transparency, through clear information and notices, is a cornerstone of the GDPR[192]. This idea of increasing trust is directly mentioned in the GDPR's Recital 7, which mentions "the importance of creating the trust that will allow the digital economy to develop across the internal market"[193].

The idea of "accountability" was primarily driven by a number of factors - such as the need to develop a framework ensuring an unfragmented approach to data protection[194]. But additionally, the goal of this new principle is to change a general public perception that the rights of individuals were not being protected[195]. The GDPR attempts to do so by ensuring that trust-building between the different parties is strengthened. In other words, the new Regulation's "cornerstone is the concept of trust: trust in data controllers to treat personal information responsibly, and trust that the rules will be effectively enforced"[196]. This goes beyond a legal responsibility, and into a greater, socio-ethical responsibility to perform tasks in a way which is in line with greater societal goals[197].

There was already some element of accountability in the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data in 1980, which stated that a "data controller should be accountable for complying with measures which give effect to the principles"[198], though the Data Protection Directive did not explicitly recognize accountability. It was in 2010 that the Article 29 Working Party advised that some additional tools to ensure the application of privacy laws needed to be added to the

[190] GDPR, Article 5(2)
[191] GDPR, Recital 39
[192] Cradock, E., Millard, D., & Stalla-Bourdillon, S. (2015, May). Investigating Similarity Between Privacy Policies of Social Networking Sites as a Precursor for Standardization. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 283-289). ACM.
[193] GDPR, Recital 7
[194] Lindqvist, J. (2017). New challenges to personal data processing agreements: is the GDPR fit to deal with contract, accountability and liability in a world of the Internet of Things?. *International Journal of Law and Information Technology. Volume 26, Issue 1, 1 March 2018, Pages 45–63*
[195] GDPR, Recital 9: "The present differences in the level of protection of the right to the protection of personal data may prevent the free flow of personal data throughout the Union and therefore constitute an obstacle to the pursuit of economic activities at the level of the Union, distort competition and impede authorities in the discharge of their responsibilities under Union law."
[196] Buttarelli, G. (2016). The EU GDPR as a clarion call for a new global digital gold standard, *European Data Protection Supervisor*, accessed at 14/4/2018 at https://edps.europa.eu/press-publications/press-news/blog/eu-gdpr-clarion-call-new-global-digital-gold-standard_en
[197] Butin (n.29)
[198] Organisation for Economic Co-operation and Development. (2002). OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. *OECD Publishing.*

regulatory framework[199], particularly the recognition of accountability, and stressing that "the increase of both the risks and the value of personal data per se support the need to strengthen the role and responsibility of data controllers"[200] (showing the close links between "accountability" and "responsibility"). This Opinion also goes back to the idea of "trust", stating that this principle will be necessary to protect both the public and private sector from suffering "significant negative effects"[201]. Interestingly, it mentions "devastating consequences in both in economic and particularly in reputational terms" and the need to ensure "the trust of citizens and consumers"[202].

We assert that this developing accountability is part of a greater shift[203]. Businesses are not just becoming more accountable for data protection, but for a host of other social causes as well[204]. This is known as the doctrine of "corporate social responsibility"[205], which encompasses a variety of interests surrounding the pressures put on businesses to show awareness and efforts to promote certain social values. This includes interests such as "listening to customers; services for disabled customers; healthy living; and data protection"[206], as well as environmental awareness and gender equality. These values are being requested more and more by newer generations, both of customers[207] and of employees[208]. This can be seen from concepts like the rise of free-range eggs[209] to taxi companies offering to plant trees to offset their carbon footprint[210]. This ''social pressure has seen a rise in recent years and is being used by the GDPR as a tool to ensure transparency and accountability in organisations. This means that businesses now have two forces pushing them to be accountable: the GDPR's legal obligations, and the social

---

[199] Cerasaro, E. F. (2017), Accountability principle under the GDPR: is data protection law moving from theory to practice?, *LUISS Law Review*, 2/2017

[200] Article 29 Working Party (2010), Opinion 3/2010 on the principle of accountability, WP 273, adopted on 13 July 2010

[201] Ibid

[202] Ibid

[203] De Hert, P., & Papakonstantinou, V. (2012). The proposed data protection Regulation replacing Directive 95/46/EC: A sound system for the protection of individuals. *Computer Law & Security Review*, *28*(2), 130-142.

[204] Pearson, S. (2017). Strong Accountability and Its Contribution to Trustworthy Data Handling in the Information Society. In *IFIP International Conference on Trust Management* (pp. 199-218). Springer, Cham.

[205] Jones, P., Comfort, D., & Hillier, D. (2005). Corporate social responsibility and the UK's top ten retailers. *International Journal of retail & Distribution management*, *33*(12), 882-892.

[206] Ibid

[207] Pelteret, M., & Ophoff, J. (2016). A review of information privacy and its importance to consumers and organizations. *Informing Science: The International Journal of an Emerging Transdiscipline*, *19*, 277-301.

[208] McGlone, T., Spain, J. W., & McGlone, V. (2011). Corporate social responsibility and the millennials. *Journal of Education for Business*, *86*(4), 195-200.

[209] Smith, N. C. (2003). Corporate social responsibility: whether or how?. *California management review*, *45*(4), 52-76.

[210] An example can be found in UK taxi firm CityCabs at https://www.citycabs.co.uk/news/general/city-cabs-carbon-neutral/ (last visited on 4/2/2018)

pressure coming from individuals (both customers and employees). In a domino effect, businesses tend to prefer partners which also follow these norms (a simple example is restaurants with "responsible" policies buying products from similarly "responsible" suppliers)[211]. This means that there is now a triple pressure on businesses: the legal pressure from the authorities, the social pressure from customers and employees, and economic pressure from business partners, all as a consequence of the wording of the GDPR and its focus on accountability[212].

Overall, as we can see, though the European Regulation is based at its core on a set of rules-based principles attempting to set out a self-contained framework, in the end practical considerations and contextual elements have made their way into the GDPR. We propose that this is evidence of a contextual, multi-faceted approach finding its way into European data protection.

These indications are not only found in the rights and principles of the GDPR. We will show that when European Law finds itself needing to interact with other legal systems, the EU uses every tool at its disposal to achieve compliance – which leads to further indications of a contextual approach to protect informational privacy.

II. A Global Instrument

The core assertion of this thesis is that informational privacy can only be adequately protected when the ability of data controllers to process data is sufficiently limited by various controls, but also that this balance-based approach is already showing signs of being adopted by the EU. The rights and principles we have analysed in the previous section show that through a number of its new rights and principles, the GDPR recognizes the need to use a number of different controls. These controls take into account not just legal tools, but also market forces, social norms, and new technologies, both to protect data subject rights directly and to incentivize data controllers to do so. As we will show, this strategy of the EU to move towards creating incentives for compliance with the tools it introduces in the GDPR is consistent with the approach developed in this thesis.

[211] Maloni, M. J., & Brown, M. E. (2006). Corporate social responsibility in the supply chain: an application in the food industry. *Journal of business ethics*, *68*(1), 35-52.
[212] Koops, B. J., & Leenes, R. (2014). Privacy regulation cannot be hardcoded. A critical comment on the 'privacy by design'provision in data-protection law. *International Review of Law, Computers & Technology*, *28*(2), 159-171.

1. The “Adequacy Test”: Multiple paths to compliance.

One of the most important goals of the GDPR is to be able to protect the data of European residents anywhere. Because of that, sending personal data of European residents outside of the EU is subject to strict requirements, and can only be sent under a set of conditions laid out in Chapter V of the Regulation[213]. These include numerous mechanisms, such as binding corporate rules to allow transfers between branches of the same entity[214], standard data protection clauses approved by the European Commission[215], or approved certifications[216]. However, countries which are considered to have “adequate” protection (as decided by the European Commission) can simply receive and transfer data from EU countries with no additional requirements. The Commission takes into account various factors to see whether the other State is adequate. This includes respect of the rule of law and human rights[217], having relevant data protection rules, as well as effective individual rights towards data[218].

However, many of these countries have legal frameworks which are based on different foundations. For example, Canada has a number of legal acts governing the privacy of information, the main one being the federal “Personal Information Protection and Electronic Documents Act”[219] (also known as “PIPEDA”). Though this act was originally passed to obtain an adequacy decision in the EU[220], it is very different from data protection as understood by the EU. It focuses on the concept of “data privacy”, which, as the name suggests, only covers the “privacy” dimension of data protection. This means that the other dimensions of data protection, such as non-discrimination, individual self-empowerment, or competition law, are not included in PIPEDA[221]. Despite this, Canada has been held to be an “adequate” country for data protection purposes.

“Adequacy” was always a nebulous concept, and was only first fully detailed in the 2015 Schrems decision[222] which invalidated the Safe Harbour Agreement. In that case, the Court of Justice of the European Union established additional criteria for adequacy,

[213] GDPR, Chapter V, Articles 44-50
[214] Ibid
[215] Ibid
[216] Ibid
[217] GDPR, Article 45
[218] Ibid
[219] *Personal Information Protection and Electronic Documents Act,* S.C. 2000, c. 5
[220] Austin, L. (2006). Reviewing pipeda: Control, privacy and the limits of fair information practices.
[221] Ibid
[222] Case C-362/14, Schrems, 6 October 2015 (ECLI:EU:C:2015:650)

which included that the country ensure a level of protection of fundamental rights "essentially equivalent" to the EU's, that this protection apply to limitations on the powers of public authorities, and that individuals have legal remedies for access and correction of data. The Directive, and the GDPR after it, have not been particularly clear about what is meant by "essentially equivalent", though the Article 29 Working Party has issued guidance on the topic[223], based on three core criteria: "content principles" (such as the core rights and principles), "additional principles" for special cases such as sensitive dad or direct marketing, and "procedural/enforcement/remedial mechanisms" which must achieve "a good level of compliance", provide "support and help to individual data subjects", and "appropriate redress to the injured parties"[224]. Importantly, this assessment shows that there is more than one way to be adequate. This includes saying that "Basic data protection concepts and/or principles should exist. These do not have to mirror the GDPR terminology but should reflect and be consistent with the concepts enshrined in the European data protection law."; and giving examples of parts of the data protection framework that are, essentially, "non-negotiable", such as the ability to object to direct marketing and the need to follow the principles of data quality, data security, or data retention[225]. As G. Greenfield puts it, "Assessment of adequacy is therefore a complex matter, a question of balancing positives and negatives, and without black-and-white criteria for inclusion or exclusion."[226]

The primary cause behind the need for the matter to be this complex is that that Europe is a precursor in this notion of "data protection" as a unique, separate right. Even countries which accept taking on data protection legislation will not have the same historical and cultural development as Europe has had, and as such will have differing legislation. If not, differences will appear, whether in application, enforcement, or case law[227]. Overall, this shows that the EU accepts that there are multiple paths to protecting individuals, and as long as some core provisions are followed, some leeway can exist.

---

[223] Article 29 Working Party (2014), Adequacy Referential (updated), WP254, Adopted on 28 November 2017

[224] Greenleaf, G. (2017), *Questioning 'Adequacy' (Pt I) – Japan* (December 7, 2017).*150 Privacy Laws & Business International Report, 1, 6-11; UNSW Law Research Paper No. 1*. Available at SSRN: https://ssrn.com/abstract=3096370

[225] Article 29 Working Party (2014), Adequacy Referential (updated), WP254, Adopted on 28 November 2017

[226] Greenleaf (n.224)

[227] Curtin, D. (2018). Does the GDPR Change the World or is the World Changing Beyond the Regulation?. In *Das öffentliche Recht vor den Herausforderungen der Informations-und Kommunikationstechnologien jenseits des Datenschutzes| Information and Communication Technologies Challenging Public Law, Beyond Data Protection| Le droit public au défi des technologies de l'information et de la communication, au-delà de la protection des donnés* (pp. 35-48). Nomos Verlagsgesellschaft mbH & Co. KG.

One place where some have argued that there is, if anything, too much leeway, is the US/EU Privacy Shield agreement, the successor to the Safe Harbor agreement which was struck down by the Schrems decision mentioned above[228]. In November 2017, the Article 29 Working Party released their first annual Joint Review on the topic of the Privacy Shield[229]. In this review, the Working Party recognizes that "some of the main points of concern for the WP29 in [the area of access by public authorities to data transferred under the Privacy Shield], have yet to be fully resolved."[230] Data transfers between the EU and US are necessary for the economy of both blocks[231]. Meanwhile, the US will not stop snooping on data coming in from the EU. This causes a stalemate where the European authorities are not satisfied with the transfers, but have to let them continue for the time being. In other words, market forces are too strong for legal forces to be effective, creating an imbalance.

This discussion over international transfers of data tells us three things. One, that the Article 29 Working Party recognizes that there are multiple paths to protecting individuals, as long as they are indeed sufficiently protected - showing hints of an adaptable approach. Second, that practical concerns such as the demands of the market can hinder data protection efforts. Third, that while the EU is alone in its understanding of a separate "right to data protection", it has managed to push forward this concept through various means, one of those being the market. In the following section, we take a closer look at this tool in the EU's toolbox of data protection, as it highlights perfectly the ability of using law to influence other forces in order to protect informational privacy.

2. The Market as a tool for EU Rule-making: The Brussels Effect

Like Canada, many countries have changed their legal provisions in order to align with European data protection law. This is part of a wider approach in EU law, in the same way that the rule of law "'is undoubtedly a value that the EU relentlessly seeks to export "beyond the borders of the Union by means of persuasion, incentives and negotiation,

[228] Case C-362/14, Schrems, 6 October 2015 (ECLI:EU:C:2015:650)
[229] Article 29 Working Party (2017), EU - U.S. Privacy Shield – First annual Joint Review
[230] Ibid
[231] Alo, E. R. (2013). EU Privacy Protection: A Step Towards Global Privacy. *Mich. St. Int'l L. Rev.*, *22*, 1095.

but other more "punishing" means have also been used…''[232] Thus, EU Law is dedicated to upholding its principles, and where the global nature of the Internet come into conflict with it, its law should prevail[233], as stated by the European Court of Justice itself in the Schrems case[234]. This constant drive to establish EU standards of individual protection throughout the world is also part of the GDPR, and the phenomenon of countries changing their laws to be in accordance with the EU's is not a coincidence, but a dedicated effort to globalize these norms[235]. Even as recently as last year Japan announced it will be working to reach a level of data protection high enough to be in line with the GDPR and receive an adequacy decision[236].

Why is it that so many countries are working on achieving the European standard? And why is it that at the same time, many companies worldwide, from Microsoft to Amazon, are rushing to comply with the GDPR on a global level despite being US-based companies?[237]

Getting back to the roots of the European project is necessary to understand this dynamic. It has always been the drive of the EU to use "strength in numbers" to develop enough influence to shift global trends to its advantage[238]. The simple reality is that businesses go where they have an economic interest to go, and so the EU's goal has always been to put companies (and States) in a situation where trading with the EU while following EU standards of rights is their best option[239]. When the Data Protection Directive passed, many third party countries became "inadequate" for purposes of data transfers, data transfers which were an important component to trade between these third countries and the EU[240]. This meant that these countries rushed to adopt relatively similar legislation to reach adequacy and resume the transfers without hindrance. The

---

[232] Pech L. (2012), The Rule of Law as a Guiding Principle of the European Union's External Action. *CLEER Working Paper 2012/3. The Hague: Asser Institute*

[233] Kuner, C. (2017). The Internet and the Global Reach of EU Law, *LSE Law, Society and Economy Working Papers 4/2018, London School of Economics and Political Science*

[234] See Case C-362/14, Schrems, 6 October 2015 (ECLI:EU:C:2015:650), at paras. 84-87, which criticize the Safe Harbour Agreement for giving US law primacy over EU rights.

[235] Greenleaf, G. (2012), The influence of European data privacy standards outside Europe: implications for globalization of Convention 108', *2 International Data Privacy Law 68 .*

[236] Greenleaf (n.224)

[237] A particular example of that phenomenon is Cloud providers. For more information, see the CISPE Code of Conduct at cispe.cloud.

[238] Treaty Establishing the European Coal and Steel Community, ECSC Treaty, EUROPA, http://europa.eu/legislation_summaries/institutional_ affairs/treaties/treaties_ecsc_en.htm

[239] Bennett, C. J. (2018). The European General Data Protection Regulation: An instrument for the globalization of privacy standards?. *Information Polity*, *23*(2), 239-246.

[240] Moerel, L. "Back to Basics: When Does EU Data Protection Law Apply*?" International Data Privacy Law 1, no. 2 (January 24, 2011)*: 92–110

reason behind so prompt a reaction is because data protection compliance has a heavy cost for companies - the closer the two legislations are (even without an actual adequacy decision), the less the cost involved in joining the European market[241]. This was especially true of small companies, which did not (and still do not, if GDPR compliance for SMEs is to be believed[242]) have the tools to achieve compliance. This meant that any company wanting to trade with the EU would have in its best interest the adoption of a European standard.

The idea of pushing the EU as a regulatory powerhouse is part of a deliberate, stated strategy. An article, published in Politico on the 31st of January 2018[243], elaborated on the very open way in which the EU is stating this goal. From Věra Jourová, the European commissioner for justice, stating that "We want to set the global standard"[244], to John Bowman, former lead negotiator for the U.K. government on Europe's new data protection rules, asserting that "Europe wanted to be seen selling a global standard", "that's crystallized through its adequacy decisions." Incidentally, the same article shows how successful this strategy has become, with Pansy Tlakula, chairperson of South Africa's Information Regulator (South Africa's newly created data protection regulator), stating that "We regard Europe's directives as best practice"[245]. In short, the GDPR is becoming a global standard for data protection due to a strategy adopted by the EU to use legal tools to influence Market incentives.

In conclusion, the clash of conceptions that come with cross-border data transfers has pushed EU authorities to allow some flexibility (in the form of the adequacy test and the use of various tools to increase the regulatory influence of the GDPR) where normal legal tools would fail. In other words, where one safeguard is ineffective, European authorities have used others to ensure data protection. As we will show in Chapter 4, this mindset comes in conflict with some core pillars of the European data protection regulatory system.

The new trajectory for EU data protection law can also be found in the GDPR's new ''risk-based'' approach. As we will show, this new approach is aimed specifically at protecting

---

[241] Ibid

[242] Movius, L. & Krup, N., U.S. and EU Privacy Policy: Comparison of Regulatory Approaches, *3INT'L J. COMM. 167, 173 (2009).*

[243] Scott, M. and Cerulus, L. (2018), Europe's new data protection rules export privacy standards worldwide, POLITICO, last accessed on 3/2/2018 at https://www.politico.eu/article/europe-data-protection-privacy-standards-gdpr-general-protection-data-regulation/

[244] Ibid

[245] Ibid

informational privacy by identifying situations where a data controller has a potentially dangerous power over information, and limiting it through a number of safeguards.

III. Informational Privacy in the GDPR's Risk-Based Approach: the Separation of "Risk" and "Harm"

1. ''Risk'' in the GDPR

While the Data Protection Directive only referred to ''risk'' in relation to the security principle, the notion of ''risk'' appears a number of times in the GDPR. Data controllers have to take into account the risks to data subjects when they implement measures which protect the rights and freedoms of the data subject[246], which means that every decision made by data controllers as to how to protect the data subject has to incorporate an assessment of risk. This alone places risk at the core of the GDPR and its commitment to data protection by design and by default[247].

Various derogations to obligations of the GDPR are negated if a risk to data subjects is present, such as the derogation to the need to appoint a representative in the EU for non-EU data controllers[248] and the derogation to document data processing activities for organisations with fewer than 250 employees[249].

Risk also extends to whether certain types of processing require extra measures to protect individuals. Certain data breaches have to be reported to the ICO if they are likely to result in a risk for the rights and freedoms of natural persons[250], and have to be communicated directly to data subjects if they represent a high risk to data subjects[251]. Similarly, if the data controller is about to engage in a type of processing which is likely to result in a high risk to the rights and freedoms of data subjects, that controller is legally required to perform a data protection impact assessment[252]. Additionally, if that

[246] GDPR, Article 24
[247] GDPR, Article 25
[248] GDPR, Article 27
[249] GDPR, Article 30
[250] GDPR, Article 33
[251] GDPR, Article 34
[252] GDPR, Article 35

assessment results in a high risk in the absence of enough measure to mitigate it, the data controller has to consult the supervisory authority[253]. It is also part of the role of the data protection officer to have ''due regard to the risk associated with processing operations''[254].

The GDPR has some examples of what risks could materialize, which include ''discrimination, identity theft or fraud, financial loss, damage to the reputation, loss of confidentiality of personal data protected by professional secrecy, unauthorised reversal of pseudonymisation, or any other significant economic or social disadvantage''[255]. This touches on certain privacy interests, such as proprietary privacy (damage to reputation), or decisional privacy (identity theft), but only gives some examples without defining the scope of the concept, leaving room for interpretation.

What constitutes ''risk'' was later clarified by the Article 29 Working Party Opinion on Data Protection Impact Assessments[256]. In this paper, the Article 29 Working Party laid out a set of 10 criteria which may present a ''high risk'' to data subjects. As we will show, the risks established in this paper are all directly linked to informational privacy, and very rarely linked to a particular interest, whether it is a privacy interest of an interest linked to another right. Studying these ''high risk'' factors is vital to understanding the GDPR's risk-based approach and how it links to the approach developed in this thesis.

The first criteria is ''evaluation or scoring, including profiling and predicting''[257], especially concerning someone's location, movements, interests, and other types of information. This criteria does not specify why certain types of information should be more protected, or which interests their lack of protection would infringe upon. As such, the only interest that this criteria protects explicitly is informational privacy, while specifically avoiding addressing which other interests may be impacted.

The second criteria includes ''automated-decision making with legal or similar significant effect''[258], which as we have seen in this section is a largely undefined notion with the potential to include a variety of possible interests. While some examples were given of particular harms (such as adverts for online gambling being targeted at individuals in

[253] GDPR, Article 36
[254] GDPR, Article 39
[255] GDPR, Recital 75
[256] Article 29 Working Party (2017), Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679, WP 248, Adopted on 4 October 2017
[257] Ibid
[258] Ibid

financial difficulties), no clear scope of interests has been defined for what is covered by ''significant effect''. Meanwhile, informational privacy, the ability to create and use information about individuals, is unequivocally present.

The third criteria covers ''systematic monitoring''[259], particularly monitoring publicly available spaces. This brings in a connection with the ''public'' space, which once again proposes a link between this criteria and privacy interests.

The fourth criteria includes data that is ''sensitive'', which includes any information that ''may more generally be considered as increasing the possible risk to the rights and freedoms of individuals''[260]. This admission of generality and non-definition of interests shows how the EU legislators are purposefully refusing to define which interests to protect. Nevertheless, once again this activity relates to the ability of data controllers to obtain and use information, and as such is related to informational privacy.

The fifth and seventh criteria (the sixth, concerning aggregation of datasets, is the subject of a separate analysis in the following section of this chapter) concern the data itself: whether it is data that is ''large scale'' and whether the data concerns ''vulnerable data subjects''[261]. These both are based on the ''power'' of the controller, whether that power comes from having more data (for large-scale processing) or from the data subject being in a position of less power than the data controller (for a vulnerable data subject). This ''vulnerable'' status is contextual - an employee can be ''vulnerable'' towards their employer, or children towards their teacher. This ability of the data controller to obtain and process data of individuals that do not possess the power to stop them is another sign of an approach based on contextual guarantees. If the right controls were in place, vulnerable individuals would not be vulnerable. The fact that this is considered to be a ''high risk'' practice shows that according to the Article 29 Working Party, an imbalance of control over the data between the data controller and data subject is a danger to data protection. This shows again the link between informational privacy (control over information) and data protection, as well as the recognition of an approach which is based on a balance between what the data controller can do with the data, and the guarantees against that ability.

The eighth criteria concerns innovative use of technology or applying new technological solutions, such as face recognition or fingerprinting. These might represent a high risk

[259] Ibid
[260] Ibid
[261] Ibid

because ‘’the personal and social consequences of the deployment of a new technology may be unknown’’[262], with a particular example, the Internet of Things, considered to have ‘’a significant impact on individuals’ daily lives and privacy’’. The mention of privacy shows, once again, that data protection and privacy are still linked - through informational privacy.

The ninth criteria is based on transfers of data beyond the EU (which touches on measures the EU takes to protect their residents outside the EU, as we explored in the previous section of this chapter), while the tenth is based on processing which ‘’prevents data subjects from exercising a right or using a service or a contract’’[263] such as a bank screening.

In conclusion, the way that ‘’risk’’ is defined in the GDPR shows a move towards the protection of informational privacy specifically. Even though data protection covers interests beyond informational privacy, the GDPR focuses on it when considering what ‘’risks’’ can result from data processing. This shows that a core focus of data protection is protecting informational privacy, and that focusing on protecting informational privacy will allow other privacy interests to be protected as well. In the following section, we will see how the GDPR sets out that in order to limit risk, some safeguards can be implemented which compensate for risk-prone processing. As we will show, this approach to protecting data (and thus informational privacy) is aligned with the framework developed in this thesis, showing that EU data protection law is already moving towards that framework.

2. Risk-mitigating Safeguards and Informational Privacy

The sixth ‘’high risk’’ factor defined by the Article 29 Working Party covers the aggregation of datasets originating from different operations. This sixth factor, in particular, becomes high risk when the data is used for purposes ‘’that would exceed the reasonable expectations of the data subject’’[264]. This ‘’reasonable expectation’’ for purposes was clarified in another opinion of the Article 29 Working Party regarding purpose limitation[265] which showed that one of the factors is the context in which the data has been collected, which includes the nature of the relationship between the controller

[262] Ibid
[263] Ibid
[264] Ibid
[265] Article 29 Data Protection Working Party (2013), Opinion 03/2013 on purpose limitation, WP203, Adopted on 2 April 2013

and the data subject as well as the balance of power between the data subject and the data controller.

This context element also includes the impact of the further processing of the data subjects, which takes into account ''both positive and negative consequences'', including emotional impacts, irritation, fear and distress ''that may result from a data subject losing control over personal information, or realising that it has been compromised.''[266] This is a direct acknowledgment that not only does the EU legislator stay away from defining the scope of interests that should be protected, but instead makes a direct reference to informational privacy, the control over information, as the core element defining whether behavior is permissible or not.

The last element of the purpose limitation's ''reasonable expectation'' test is perhaps the most significant. It that not only is the core interest that data protection aims to protect informational privacy, but that the EU recognizes that protecting informational privacy is best done with a balance between the ability of data controllers to process data and the safeguards limiting that ability. This last element is based on ''the safeguards applied by the controller to ensure fair processing and to prevent any undue impact on the data subjects''[267]. The Article 29 Working Party asserts that additional measures could serve as ''compensation'' for a change of purpose, including additional steps such as better transparency, anonymisation measures, or the ability to object to processing. Further guidance on what the safeguards should achieve focus on two aspects: data security (which includes availability, confidentiality, and integrity of personal data) as well as data protection (which includes transparency, isolation and ''intervenability'') with a focus on the isolation element. This element is about limiting what information exists, how available it is to be used by the data controller, and whether any privacy enhancing technologies are applied.

In other words, to ensure whether datasets can be combined together in a way that the data subject can ''reasonably expect'' without creating a risk to the data subject, a number of guarantees can be applied to ensure that there are sufficient safeguards in place in order to protect informational privacy. This approach can also be found in data protection impact assessments, where having sufficient safeguards can mitigate risk and allow processing operations to be carried out despite representing a risk[268]. This

[266] Ibid
[267] Ibid
[268] Van Dijk, N., Gellert, R., & Rommetveit, K. (2016). A risk to a right? Beyond data protection risk assessments. *Computer Law & Security Review, 32(2), 286-306.*

approach, allowing or disallowing processing based on whether sufficient guarantees are in place, is new to the GDPR and shows that the framework which will be developed in this thesis is already being implemented by the EU.

IV. The Balancing Approach in the GDPR: the "Legitimate Interests" Test

The idea of regulating the processing of data based on a balance between the Information and Guarantees limiting that processing is not only found in the GDPR's risk-based approach. As we will see, the principles-based approach of data protection, reinforced in the GDPR, is made up of a balance between risks and safeguards limiting thost risks.

This can be identified by studying the legal bases for processing personal data. A legal basis in the GDPR already present in the Data Protection Directive is whether the controller has "legitimate interests" for processing data[269]. Legitimate interest as a legal basis requires a balancing test of the interest of the data controller and the risk to the data subject[270]. Particularly, whether a risk is present to the data subject partly depends on how "identifiable" the data subject is. For example, December 2013 comments by Neelie Kroes, Vice-President of the European Commission responsible for the Digital Agenda, stated[271]:

"Sometimes, full anonymisation means losing important information, so you can no longer makethe links between data. That could make the difference between progress or paralysis. But using pseudonyms can let you to analyse large amounts of data: to spot, for example, that people with genetic pattern X also respond well to therapy Y. So it is understandable why the European Parliament has proposed a more flexible data protection regime for this type of data.Companies would be able to process the data on grounds of legitimate interest, rather than consent. That could make all the positive difference to big data: without endangering privacy."[272]

---

[269] GDPR, Article 6

[270] Furey, E., & Blue, J. (2018). She Knows Too Much–Voice Command Devices and Privacy. In *2018 29th Irish Signals and Systems Conference (ISSC)* (pp. 1-6). IEEE.

[271] Hintze, M. (2016). Viewing the GDPR through a De-Identification Lens: A Tool for Clarification and Compliance. *preprint*, 1-22.

[272] Kroes, N., "Data isn't a four-letter word," IAPP Europe Data Protection Congress/Brussels, 11 Dec. 2013.Available athttp://europa.eu/rapid/press-release_SPEECH-13-1059_en.htm.

This shows that the legitimate interests legal basis was intended, at least in part, to allow for greater flexibility and breadth of processing where sufficient safeguards would ensure there is no risk to data subjects. This flexibility however means legal uncertainty, as limits of this flexible legal basis have not been strictly set, such as in the case of Asnef and Fecemd v. Administración del Estado[273] where the ECJ supports a wide interpretation of legitimate interests which allows the credit industry to process negative financial data of consumers based on legitimate interests[274]. The Article 29 Working Party similarly acknowledged that Google was able to base most of its business needs on legitimate interests including providing, maintaining, protecting and improving services[275].

More guidance over the "legitimate interests" legal basis was developed by the Article 29 Working Party in a dedicated opinion[276]. In it, the Article 29 Working Party "recognises the significance and usefulness of the Article 7(f) criterion, which in the right circumstances and subject to adequate safeguards may help prevent over-reliance on other legal grounds". In other words, whether processing can be allowed on "legitimate interests" grounds depends on a balance between how much information can be processed, and the safeguards on that processing. It is the processing itself which requires safeguards, not specific harmful events that may materialize which need to be prevented. This, we argue, is the direction EU data protection law is inevitably moving towards.

A test of whether there is an appropriate balance between the information and safeguards does includes factors such as "the impact of the data subject"[277], but as is the case where we discussed "risk" in the previous section, what constitutes "impact" remains largely undefined.

A strong statement related to the need for a flexible approach is detailed by the Article 29 Working Party in that Opinion: "[...] the need for some flexibility also comes from the very nature of the right to the protection of personal data and the right to privacy. Indeed,

---

[273] C-468 & 469/10, ASNEF and FECEMD v. Administración del Estado, [2011] ECR I-12181.
[274] Ferretti, F. (2014). Data protection and the legitimate interest of data controllers: much ado about nothing or the winter of rights?. *Common Market Law Review*, *51*(3), 843-868.
[275] Article 29 Working Party, "Letter from the Article 29 Working Party addressed to Google along with the recommendations" (Brussels, 16 Oct. 2012), at ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2012/20121016_letter_to_google_en.pdf
[276] Article 29 Data Protection Working Party (2013), Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, WP217, Adopted on 9 April 2014
[277] Ibid

these two rights, along with most (but not all) other fundamental rights, are considered relative, or qualified, human rights. These types of rights must always be interpreted in context. Subject to appropriate safeguards, they can be balanced against the rights of others. In some situations-and also subject to appropriate safeguards-they can also be restricted on public interest grounds."[278]

This balance between the information and the safeguards can also be found when studying Article 9 of the GDPR, which lists the exemptions which allow a data controller to process special categories of personal data[279]. Some extra safeguards are implemented (for example, consent is replaced by "explicit" consent, and "public interest" is replaced by "substantial public interest") which show, as we discussed in the previous section of this chapter, that higher-risk processing is offset by additional guarantees. Additionally, legitimate interests do not appear in that Article, which means that special categories of personal data cannot be processed using "legitimate interests" (unless another Article 9 exception allows it, such as if the data controller is a foundation or not-for-profit). This shows that according to the GDPR, some processing can be made legitimate by sufficient guarantees, but some cannot if it is too high risk.

The Article 29 Working Party, in an attempt to clarify where the concept of ''legitimate interests'' applies, developed its various elements, especially what "Interests or rights of the data subject" means. As we will see however, that notion remains largely undefined even after the Working Party's attempts at clarifying it.

An important mention when discussing these interests or rights is that it is necessary to pay particular attention to the interests and rights of data subjects because: "At a time of increasing imbalance in 'informational power', when governments and business organisations alike amass hitherto unprecedented amounts of data about individuals, and are increasingly in the position to compile detailed profiles that will predict their behaviour (reinforcing informational imbalance and reducing their autonomy), it is ever more important to ensure that the interests of the individuals to preserve their privacy and autonomy be protected."[280] This shows that there exists a balance of power between data controllers and data subjects, and that processing can be made legitimate where the balance of power is equal (as in these cases, "legitimate interests" can be used as a legal basis). This also shows that not only the processing itself, but also all elements of

[278] Ibid
[279] GDPR, Article 9
[280] Article 29 Data Protection Working Party (2013), Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, WP217, Adopted on 9 April 2014

the power balance between the controller and data subject should be considered. This also shows the wide scope of what "interests and rights of the data subject" can encompass.

The elements of the test developed by the Article 29 Working Party when balancing the interests of the controller with the rights of the data subject have a very broad scope. The impact on the data subject is meant to assess "both positive and negative consequences"[281] of the processing, which may include discrimination, defamation, damage to reputation or to the autonomy of the data subject. Adverse outcomes can include irritation, fear, the chilling effect on "freedom of research or free speech that may result from continuous monitoring/tracking", and should be considered "much broader than the impacts that may result from a data breach"[282]. Importantly, this wide notion of harm is very similar to the one found in privacy cases such as the UK case Gulati & Ors v MGN Limited[283] in which "harm" to the right to privacy in the context of a phone-hacking case included elements such as "loss of privacy or "autonomy" resulting from the hacking or blagging that went on; there is compensation for injury to feelings (including distress); and there is compensation for "damage or affront to dignity or standing"[284].

The judge in the case mentioned that "While the law is used to awarding damages for injured feelings, there is no reason in principle, in my view, why it should not also make an award to reflect infringements of the right itself, if the situation warrants it. The fact that the loss is not scientifically calculable is no more a bar to recovering damages for "loss of personal autonomy" or damage to standing than it is to a damages for distress. If one has lost "the right to control the dissemination of information about one's private life" then I fail to see why that, of itself, should not attract a degree of compensation, in an appropriate case"[285]. The judgment, especially its acknowledgment of "loss of autonomy" as a specific harm, shows both the link between data protection and privacy through the similarity between their related harms, but also the difficulty and subjectivity in assessing these harms - in this case, the defendant argued that only "distress or injury to feelings" should be a harm worth compensation[286].

[281] Ibid
[282] Ibid
[283] Gulati & Ors v MGN Limited[5] (confirmed by the Court of Appeal in Representative Claimants v MGN Limited [2015] EWCA Civ 1291
[284] Ibid
[285] Ibid
[286] Ibid

The Working Party goes beyond even that scope of privacy or data protection harms: "relevant 'impact' is a much broader concept than harm or damage to one or more specific data subjects. 'Impact' as used in this Opinion covers any possible (potential or actual) consequences of the data processing."[287] This includes positive or negative consequences of not just one processing, but an accumulation of separate occurrences which together may have an impact on the data subject. It can extend from annoying marketing phone calls to loss of life, it can change based on the nature of the data itself, but also how it is being processed[288]. It can also take into account the status of the data subject and data controller, particularly whether the data controller is in a position of power to the data subject. The balancing test is based in principle on the "average individual"[289] but may incorporate a case-by-case approach in some specific situations, particularly vulnerable persons (mirroring the "high risk" processing areas covered by the Working Party when detailing situations where a DPIA may be necessary).

This reinforces our findings in the previous section of this thesis: "impact" is largely undefined, its scope so wide as to encompass a great number of rights or interests with no clear definition of where it applies. In order to handle this scope, the Article 29 Working Party is showing hints of moving towards a definition of "risk" not based on the actual impact, but on the types of processing involved, such as whether sensitive data or data about vulnerable persons is processed. This approach, basing impact not on actual harm or consequences but on the balance of power between holders of information and subjects of that information, is what we argue in this thesis is the natural and necessary way forward. This move is not yet fully realised, but strong signs are clear in the way the GDPR approaches the "legitimate interests" legal basis.

The fact that this approach is not fully realised, and that the Article 29 Working Party still attempts to base the decision on "impact" and "likelihood" despite how broad these concepts are means that "legitimate interests" as a legal basis remain undefined and too tempting for data controllers looking for an easy way to process data without obtaining consent. While a flexible approach is necessary in the Big Data age, "legitimate interests" attempt to cover all possible interests and rights that may be impacted as a result of processing, all possible harms and consequences that may materialise, leading to too wide a scope. Instead, we argue that the focus should continue to move towards

[287] Article 29 Data Protection Working Party (2013), Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC
[288] Ibid
[289] Ibid

informational privacy and a balance between information and limits on the use of that information.

Conclusion

The GDPR makes some moves towards a contextual approach with notions such as “significantly affects data subjects”, with some leeway when it comes to the “adequacy” requirement, and particularly with the move towards a risk-based approach which is based on a balance between risk-prone data processing activities and the safeguards against these activities. The need to accept a more contextual and multi-factor approach comes from the fact that some problems cannot be solved with legal tools alone, and need to incorporate other safeguards to protect individuals.

On the other hand, the GDPR also relies on the existence of these different forces in order to push forwards its objectives. Whether it is by using “accountability” to include market forces and social trust into the legal framework or by using the political and economic power of the EU to ensure other countries attempt to align their legislation with the EU’s, the European data protection authorities have shown that protecting informational privacy through data protection can be done using a number of tools from multiple disciplines.

As such, the GDPR makes protecting informational privacy the focus of data protection, instead of defining a strict scope of what information should be protected and what harmful consequences could occur. This focus on informational privacy to address the challenges of the modern age, and use of safeguards to balance out the risks involved in data processing, all show that European data protection law is moving towards a balance-based approach between the ability of data controllers to process data, and guarantees against that processing breaching informational privacy.

In the next chapter, we will analyze the transformations that have led to the challenges to informational privacy: the phenomenon of Big Data and the new way data and information have transformed.

Chapter 3. Information in the Age of Big Data

Introduction: Big Data and "Information"

The fact that this approach is not fully realised, and that the Article 29 Working Party still attempts to base the decision on "impact" and "likelihood" despite how broad these concepts are means that "legitimate interests" as a legal basis remain undefined and too tempting for data controllers looking for an easy way to process data without obtaining consent. While a flexible approach is necessary in the Big Data age, "legitimate interests" attempt to cover all possible interests and rights that may be impacted as a result of processing, all possible harms and consequences that may materialise, leading to too wide a scope. Instead, we argue that the focus should continue to move towards informational privacy and a balance between information and limits on the use of that information.

Big Data is often characterised as the "Three V's": Volume, Velocity and Variety[290]. Volume is the most familiar understanding of Big Data, and refers to that ever-bigger amount of data being created. Simply storing and accessing that data is a new challenge.

Meanwhile, Velocity is fast becoming the most difficult of the challenges to meet. To process, move and use that amount of information is just as - or even more - useful than having that data in the first place. Processing and understanding it takes complex technology, and it is the development of such technology which currently is the main obstacle to the expansion of Big Data[291].

[290] Tene, O., & Polonetsky, J. (2011). Privacy in the age of big data: a time for big decisions. *Stan. L. Rev. Online*, *64*, 63.
[291] Ibid

Finally, Variety is what expands Big Data to a wide array of fields: data can come from many sources, be they in a different language, or different types: visual (photos and films), audio (sound files), spreadsheets, paper files, etc. Aggregating many types of data together creates an exponential amount of possibilities in what can be deciphered from the data, and overcoming the technical challenges in doing so is a main avenue of expanding the size of "Big Data".

Above all, Big Data is a matter of scale. Individually, every part of what constitutes "Big Data" existed before it. The recording of "Data" has always existed, in the form of bookkeeping or censuses, photographs and recordings. Over the last half-century data began to be accumulated and analyzed in new ways due to the rise of computers in every part of our lives, as well as the appearance of new ways to collect data - cheap electronics from sensors to smartphones and cameras[292]. This led to an exponential growth of the data and its analysis over time: it is estimated that 5 exabytes of data (10^{18} bytes) were created by humans in all of history up to 2003[293], while it had expanded a thousand fold by 2012 to 2.72 zettabytes (10^{21} bytes). It is estimated to double every two years, which creates opportunities never before possible[294]. With the rise of Big Data came the expansion of the Information Age.

"Information", as understood in this thesis, will be capitalized when referred to as opposed to the common meaning of the word "information". We will first study the basics of what Information is, and what makes it different from "data". We will then assemble a conception of Information from the various developments made earlier in this thesis, from the conceptual levels of definitions of informational privacy all the way to the modern understandings of the "suitable safeguards". This will provide a thorough description of Information, which will then be expanded by the implications of such a notion and why it allows for a better insight informational privacy issues in the Digital Age.

American organizational theorist Russell Ackoff delimited a "scale" from "data" to "wisdom"[295]. Under that understanding, "data" is raw, exists objectively, and can have any and every form, whether it can be used or not. "Information", meanwhile, is data

[292] Smith, M., Szongott, C., Henne, B., & Von Voigt, G. (2012, June). Big data privacy issues in public social media. In *Digital Ecosystems Technologies (DEST), 2012 6th IEEE International Conference on* (pp. 1-6). IEEE.
[293] Intel IT Center, Planning Guide: Getting Started with Hadoop, *Steps IT Managers Can Take to Move Forward with Big Data Analytics,* June 2012
[294] Sagiroglu, S., and Sinanc, D. (2013) Big data: A review. *Collaboration Technologies and Systems (CTS), 2013 International Conference on*. IEEE.
[295] Ackoff, R. L. (1999). *Ackoff's best: His classic writings on management*. John Wiley & Sons.

which has been given “meaning by way of relational connection”[296]. This meaning can be useful, and allows answering questions such as “who, what, when, where, and how many.”[297] Meanwhile, “knowledge” is the collection of information into something useful, such as “2 + 2 = 4” and requires the ability to store and process information in a creative or complex way. Then, “understanding” is a process to create new knowledge from previously-held knowledge. Finally, “wisdom” is a further step where new understanding is created from previous understanding and judges whether something is right or wrong, good or bad[298].

These categories go in rather philosophical directions as they reach higher levels of abstraction, and we do not need to ask questions of what “wisdom” means in order to conceptualize informational privacy. However, it is important to understand that these words matter: characterizing the output of data processing as “wisdom” is quite different from judging it to be simply “information”. It would be tempting to judge “wisdom” as being more valuable than “data”. However, we would argue that creating layer after layer of understanding based on a set of data is not a ladder leading to wisdom, necessarily. Instead, every step from the raw data to the decisions made from that data requires the input of human beings, whether it be directly - by looking at a list of student grades then targeting the worst performers for additional support - or indirectly - by designing an algorithm that decides who gets the better rates for insurance. “Information” is the combination of data and human agency, and the deeper one goes in the study and processing of that data, the more human agency is added in the mix, to the point where the end result often is so distant from the original data as to be something altogether different.

In this thesis, ‘’Information’’ will relate to any data that has seen an element of human agency towards it, which means that it includes ‘’understanding’’ and ‘’wisdom’’.

It is important to note that Information does not have a physical support. Instead, it is the result of the data being observed. Nevertheless, it is possible to assess what kind of Information can be deduced from what data the individual data controller is holding.

To show why “Information”, and not “data”, is the notion to focus on, an example can be useful. If a British individual - with no expert knowledge or experience, speaking only English - were to come into possession of a report holding valuable personal data written

[296] Ibid
[297] Ibid
[298] Ibid

in German, the data would be sensitive and valuable, but the Information that the individual can deduce from it would be non-existent. Without more data, or an understanding of German, the data is completely useless to that one individual. This is to show the fact that whether “personal data” leads to “personal Information” is highly dependant on factors which may have more to do with the observer than with the data. With instant translation services now available online, the same data could now lead to Information now which would have been more difficult to obtain a decade ago.

With this basic view of Information established, we will now see how Big Data was primarily a transformation of the creation and transmission of Information in a way never seen before. We will now analyse the new technologies of Big Data from another perspective, this time looking at what they mean not for the data created, but for the Information.

Data collection is increasing to never-before-seen levels, through technologies such as the Internet of Things, smartphones, cookies, or the increasing capabilities of software. This increase of the “Volume” and “Variety” aspects of Big Data have been accompanied by an increase in the third “V” - “Velocity”, but this increase has not kept up with the explosion of the other factors. In “data/Information” terms, this means that despite a huge amount of data being collected, the capabilities of data processing do not allow them to be fully processed into Information. As such, not only are “data” and “Information” separate concepts, but what will lead to the creation of one might not necessarily affect the other in a linear fashion. We will now see how the goal of all data processing is to create new and more powerful Information, and how viewing Information as a separate construct allows for a better understanding of the challenges of Big Data.

A. Data Processing and Information Creation: A Combination of Data and Human Agency

Data does not remain the same over the course of its processing. It becomes changed, curated, aggregated, translated, moved and otherwise transformed at every stage of processing. This has two major effects: it changes the properties of the data, and it changes the Information that can be obtained from that data. As we will see, these two

effects are at the source of essentially everything that can be done with data, including its informational privacy impacts, and so understanding how this process works is important. We will start with an overview of the process and what has an impact on the data for each stage, then move on to detail how that impacts each of the main effects. For each stage, we will analyse the relevant factors which affect the Information which can be obtained at that stage.

There is a myriad of such steps and they can occur multiple times in one processing activity, but for practical reasons we use this typology: collection, storage, aggregation, processing, mining, and use. As we will show, each stage can modify the data and the Information it creates.

I. Data Collection and Information: Choosing the Data

At the first step of the process is “data collection”. In a way, it is the most important, and what brought on the age of Big Data. Data has always existed in the “wild”, but not until recently was it possible to collect it efficiently[299]. It used to be that finding the perpetrator of a crime required the imperfect memories of witnesses. Since then, we have learned how to collect new types of data: we can search for, and analyze, DNA samples[300]. We can study the CCTV footage of the area. We can look through social media to learn about the victim.

Additionally, data is more easy to access than ever because of the rise of computers in our daily environment and their ability to record automatically vast amounts of information. Written records tend to be slow to prepare: copying one book used to be the matter of weeks or months. Now gigabytes of data can be generated, saved and searched in a matter of seconds.

2. What data exists

[299] Tene (n.290)
[300] Beyleveld, D., & Brownsword, R. (1998). Human dignity, human rights, and human genetics. *The Modern Law Review, 61(5), 661-680.*

First off, what will dictate the Information created is the kind of data from which it is created. If no data can be collected, no Information can be created. There are two aspects to this: on the one hand, data can be collected that did not use to be collected, and on the other, data can now be collected in greater and more reliable numbers, which gives it a value it did not use to have.

The first aspect can be seen in the capabilities of new technologies. Obtaining an individual's search engine queries, browser history, IP address or social media data is all entirely new data that never existed before. In the Big Data age, much of what we do has moved to a digital format – including a large portion of socialising – which has led to this outpouring of new data which can be collected. This is a continuing process, but it is also important to remember that it has a limit. There is a tremendous amount of data that as of yet cannot be collected, and may never be, even though if it could it would have tremendous economic value. This includes emotions, dreams, thoughts, urges, and more. This data cannot be collected as we simply do not possess the technology to do so. Though this seems like the realm of science fiction, so was the idea that it would be possible to spy on individuals, seeing and hearing them, through their microphones and cameras, without their knowledge or consent, yet that has happened, and is happening in a widespread manner[301]. Data collection is still limited to what can be collected within the limits of our current technology, but this can change, and this will affect the Information held by various parties.

The second aspect is more indirect, but also more problematic. A vast amount could be gathered previously, but was not practical to gather until now. This leads to a complicated situation: on the one hand, the data itself was already being gathered, legally, by various actors. On the other, it was gathered so sporadically and expensively that any widespread, privacy-invasive uses of that data were impossible. This meant that the data kept the same legal status while slowly becoming more and more transparent - until a breaking point was reached. An example is in United States v. Knotts[302], in which government agents tracked a drum of chemicals to follow the movements of the defendants. The Court held that since the drum's location was followed through public spaces, where agents could have physically tracked it, the data-gathering was held as equivalent. This essentially justifies the surveillance of all public spaces, since technically it would be possible for a police agent to be present at every street corner and following every individual in public.

[301] Kelion, L. (2017), Wikileaks: CIA has tools to snoop via TVs, *BBC News,* 7 March 2017
[302] United States v. Knotts 460 U.S. 276, 278 (1983).

The reason behind this change is because in order to obtain Information one needs to be able to gather data in sufficiently large quantities and with sufficient reliability. Technologies now available for public surveillance don't just allow the data collection to be more accurate and more widespread. They also allow for such data collection to take place over longer periods of time, thereby giving a more complete picture – so much so that there have been calls to decide whether surveillance goes beyond a reasonable expectation of privacy based on that duration factor[303]. As such, the increase in data collection can manifest itself in a number of ways, with the overall consequence that more of the same data is being collected, to the point where the Information which can be obtained from it can be used in ways that may impact informational privacy. As such, the Information that gets created will change based not just on new data that can be collected, but also on existing data that can be collected in more efficient ways.

2. What data can be collected

What Information can be created also depends on what data is actually available. Though this observation is obvious so far, there is an important factor to take into account. What data is "available" will change based on who is doing the data collection. Depending on the observer, the means of data collection will be very different, to the point where different Information will be created by different parties. Additionally, as we have seen, not just the types of data but also the ease of access to it are relevant factors. In short, depending on what data can be accessed by different entities, as well as how advanced their data gathering tools are all, change what Information they can create.

An example can be seen in the observations we have made earlier regarding identifiability. Under Article 4 paragraph 1 of the GDPR, "personal data" is based on whether the person is "identifiable"[304], based on "the means likely reasonably to be used either by the controller or by any other person to identify the said person". The "means likely to be used" will change depending on which controller is attempting to obtain Information.

Gathering data can be expensive, and data controllers will only do so if it is useful for their purposes and worth the expense. As such, whether data will be gathered – the

[303] Bellovin, S. M., Hutchins, R. M., Jebara, T., & Zimmeck, S. (2013). When enough is enough: Location tracking, mosaic theory, and machine learning. *NYUJL & Liberty*, *8*, 556.
[304] GDPR, Article 4

"means likely to be used" – is based on a number of factors. First, whether the data gathering tools exist. Second, whether gathering that data is possible using the resources available to the entity. Third, whether these resources are worth the price the entity is willing to pay for that data. The last factor, in particular, is very important, as it means that the "means likely to be used" test will depend not just on what the current technological environment permits, and not even just what various entities could potentially do, but additionally whether they are likely to go through the expense of gathering those tools. For example, a large company only interested in customers' purchases for inventory purposes will be unlikely to use data collection tools to gather vast amounts of other data on their customers, while companies which base their business model on creating detailed profiles of their customers to sell to advertisers (such as Facebook) will go through a lot more effort to gather all possible data. In such a situation, though both entities have a vested interest in data-gathering, and both have the resources and tools to do so, one's "means likely to be used" are at a disproportionately high standard.

Because of those factors, we can already see that the concept of "personal data" or "public/private" spaces is beginning to be ineffective when considering the fact that Information, and the means one can use to procure it, changes from individual to individual. Privacy in public still exists against those who do not routinely use trackers to follow people, while one's personal social media data is still well-hidden from the strangers one passes on the street. However, that informational privacy protection is not there against police authorities or Facebook. Since the State, under the national security exception in Article 23 of the GDPR, "may restrict by way of a legislative measure the scope of the obligations and rights"[305] provided by the Regulation, and since the State has a massive amount of tools available to carry out such data collection, is there privacy in public if the government has access to that data? As we will show, just because one entity is able to obtain vast amounts of Information on individuals does not necessarily have an impact on informational privacy. After all, one's partner or family also holds huge amounts of personal information about an individual - yet that is not considered a major privacy concern.

3. How is the collection performed

[305] GDPR, Article 23

As we have seen, data collection will depend on what data the observer has the ability to collect. Additionally, however, the means behind that data collection will change what will be collected and how. We have studied how data collection is now ubiquitous through various tools including the Internet of Things. However, it is important to remember that "data" is not a perfect reflection of our world, but a translation of real world phenomenon into another form. For example, a smartphone's GPS location data is not literally a log of the phone's location - it is a log of results from computations done using the phone and the satellites it is linked to. GPS is not perfect, and depending on the position of the user, the data may be inaccurate to some degree, with forests or weather patterns affecting this accuracy[306]. As such, the data is not a perfect representation of the phone's movement, which in turn is not a perfect representation of the movements of the individual who carries it. This is important because it means the Information created from the data will be inaccurate as well, and every step following that, which uses that data, will carry on these inaccuracies.

In the case of GPS, of course, such an issue might seem minor - if one is only interested in a general idea of their location, a few meters of inaccuracy might not matter. However, every piece of technology has its limitations, and no matter how data is collected, there will be a chance for inaccuracies. We have a tendency to assume that the data is correct, because computers have no reason to hold bias and are perceived as unable to make a mistake[307]. If one follows this tendency, the Information one will create will be inaccurate. Meanwhile, the same data, captured with less inaccurate means, would lead to different Information - an example would be tracking an individual's purchase habits using data from stores at which they purchase at against using an online questionnaire to ask the person what they buy - individual memory is not as accurate as the actual record of what is purchased.

4. Who is doing the collecting, and why

This leads us to the last of our factors: we have previously established that depending on the observer's technological skills, different data can be collected. However, additionally, what data will be collected will also depend on the purpose of that collection. Because different data leads to different Information, while a different purpose for

[306] Wing, M. G., Eklund, A., & Kellogg, L. D. (2005). Consumer-grade global positioning system (GPS) accuracy and reliability. *Journal of forestry*, *103*(4), 169-173.
[307] Soffer, P. (2010). Mirror, mirror on the wall, can i count on you at all? Exploring data inaccuracy in business processes. *Enterprise, business-process and information systems modeling*, 14-25.

collection leads to different data being collected, naturally a different purpose will lead to different Information at the end of the process.

This is particularly important to mention in the context of human bias. If the data collector is looking for particular Information, they will look only for data which creates that Information. For example, if a company is interested in knowing what their customers buy, they will only gather data which they believe will lead to that Information. This is important because it affects the "means reasonably likely to be used" by the company - it is not likely for such a company to gather data it has no use for or reason to gather. Looking for particular Information is not a problem in itself, but problems arise when one is looking for certain conclusions. This phenomenon is essentially a confirmation bias, where the individual or company will only integrate the data which, in a vacuum and ignoring data which might contradict it, leads to Information which confirms those biases. Examples can be an individual who dislikes their neighbor and remembers every possible transgression while forgetting any acts of kindness or mitigating circumstances, or a company convinced in a product's future success with a target market only taking into account surveys which reinforce that conclusion[308].

This bias can also be present in the data-collecting tools. Tools and programs are designed by individuals, and aimed at specific uses, sometimes in specific contexts. A rough example would be using a GPS tracking of a cell phone to ascertain its owner's location. There are a number of factors which may make the Information inaccurate, and as such the tool used is imperfect in its own right.

5. New Knowledge: ''N=All'' and the Internet of Things

In the age of Big Data, creation of data is constantly expanding into new directions, with processes such as the Internet of Things[309]. The Internet of Things has been defined as " a dynamic global network infrastructure with self configuring capabilities where physical and virtual 'things' have identities, physical attributes, and virtual personalities. They use intelligent interfaces, and are seamlessly integrated into the information network".[310] Any item can be part of the Internet of Things: clothes, headphones, fridges, lamps, cars or

[308] Krumholz, H. M. (2014). Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. *Health Affairs*, *33*(7), 1163-1170.
[309] Sarkar, S., Chatterjee, S., & Misra, S. (2018). Assessment of the Suitability of Fog Computing in the Context of Internet of Things. *IEEE Transactions on Cloud Computing*, *6*(1), 46-59.
[310] Hoepman, J. (2012). In things we trust? Towards trustability in the internet of things. Constructing Ambient Intelligence. *Springer Berlin Heidelberg*. 287-295.

whole buildings all can tap into this new architecture[311]. The Internet of Things is finding its way into more and more aspects of our everyday lives, interacting with each other to combine data, creating more powerful tools - exploiting the "Variety" aspect of Big Data[312].

Internet of Things devices can be used for any number of purposes[313], which can end up gathering very powerful data on individuals[314]. An example that is becoming more and more common is the use of monitoring devices to gather information about the individual's behaviour in order to adjust prices, such as in "Pay-as-You-Drive" insurance[315]. These types of insurance have been including black boxes in cars, which monitor an array of data types such as where the car is, where it is parked or how fast it was going, and can lead to changes in pricing to disincentivise certain behaviours[316].

This highlights one of the two major sides of the impact of Big Data on privacy: the ability of monitoring individuals in the public sphere no matter where they are. This does not only come from pressure on the side of industry with phenomena such as "Weblining", but also from products which draw their functionality and use from their ability to gather data automatically. The most obvious is smartphones: equipped with a wide array of devices able to locate and track the individual in real time and record sound and images, smartphones and their many apps using these functionalities obtain access to an unprecedented amount of personal information. These devices go beyond simply the smartphone. From health trackers such as FitBit and FuelBand to smart objects such as smartwatches or even smart clothes[317].

Though it is still a very young field, technologies such as "augmented reality", "geolocation" or "QR codes"[318], as well as the development of permanently-online devices from phones to cars, have resulted in a new level of data collection. Instead of

---

311 Ibid

312 Treacy, B. and Bapat, A. (2013). The 'Internet of Things' — already in a home near you?, *Privacy and Data Protection, 14 2 (11)*

313 Ford, D. T., & Qamar, S. (2017). Seeking opportunities in the Internet of Things (IoT):: A Study of IT values co-creation in the IoT ecosystem while considering the potential impacts of the EU General Data Protection Regulations (GDPR).

314 Treacy (n.312)

315 Wilson, S. (2014), The Collision between Big Data and Privacy Law. Au*stralian Journal of Telecommunications and the Digital Economy, Volume 2 Number 3,* October 2014. Available at SSRN: http://ssrn.com/abstract=2548079

316 Ibid

317 Peppet, S., Regulating the Internet of Things: First Steps Toward Managing Discrimination, *Privacy, Security & Consent (March 1, 2014). Texas Law Review, Forthcoming.* Available at SSRN: http://ssrn.com/abstract=2409074

318 Ibid

data being entered by individuals into databases through the intermediary of a computer terminal, data now is created by the devices themselves, without human input. A company at the forefront of these technologies, Tesla, has recently boasted about their data-gathering capabilities having extended to 1.3 billion miles of autopilot data[319], something not possible only a few years ago.

There is a double trend in IoT devices. On the one hand, they capture more and more data and are more and more common, which create a ubiquitous web of data collection. On the other, however, these sensors become more and more personal, intimate, and integrated in our lives. Examples are inventions such as the PillCam or the SmartPill, ingestible technologies which can monitor bodily functions or sensors such as the BioStamp, a device which can be worn like a Band-Aid and measures various health-related metrics such as brain activity or hydration levels[320]. Recently, an individual was charged with arson using evidence from their own pacemaker[321]. As these devices become more and more common, the question becomes not whether data will be generated and collected, but how to control this flow of data.

A particular phenomenon which has changed the way we understand data is the concept of "n=all". On any normal research project, a sample "n" is used from which to gather data from. Countless challenges have stemmed from choosing how to select "n" and get sufficient high-quality data[322]. These challenges are disappearing as we grow closer to "n=all": a stage at which the sample size will be every single individual. Tesla uses their cars currently on the road, driven by consumers, to gather data. In the same way, Google obtains data from every Google search, Facebook from every one of its users, and so on[323]. A noteworthy example came from the Google Flu Trends[324] project in 2009: using searches for popular terms linked to flu, such as "flu symptoms" or "pharmacies near me", the Google team used a database of 50 million searches, letting their algorithm identify the patterns. This early example of "Big Data" was extremely successful, able to

[319] Lambert, F. Tesla has now 1.3 billion miles of Autopilot data going into its new self-driving program, *Electrek*, 13 November 2016, found at: https://electrek.co/2016/11/13/tesla-autopilot-billion-miles-data-self-driving-program/
[320] Ibid
[321] Varghese, S., Cops use pacemaker data to file arson charges, *iTWire*, 2 February 2017, available at: https://www.itwire.com/data/76677-cops-use-pacemaker-data-to-file-arson-charges.html
[322] Mayer-Schönberger, V., and Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. *Houghton Mifflin Harcourt*.
[323] Harford, T. (2014) Big data: A big mistake?. *Significance* 11.5: 14-19.
[324] Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science, 343(6176), 1203-1205.*

predict when a flu outbreak was going to happen and where with better accuracy than the government's health services, and much faster[325].

This ubiquitous web of data collection can be found in one particular type of innovation, which has been gaining traction in EU communities: "Smart Cities". Increasingly, IoT devices are being proposed as the solution to the intricate problems of cities, such as traffic congestion, lack of services, crime, or pollution[326], leading to the notion of "smart cities". There is no clear definition of what a "smart city" is, but some of their principal features have been catalogued by Lilian Edwards in "Privacy, security and data protection in smart cities: A critical EU law perspective."[327] Such cities include in their infrastructure networks of IoT devices including "roads, cars, fridges, electricity meters, domestic appliances and human medical implants" in order to gather the data (the "Variety" of Big Data's "3 V's", networks of digital communications to allow real-time analysis and use of that data (the "Velocity") and the high-capacity infrastructure that can accommodate such a high amount of data (the "Volume")[328].

All of these technologies not only are able to collect and analyse a huge amount of data, but are also generally compliant with the existing privacy and data protection framework for two reasons: first, the data collection is done in public, and as we will see later on in this thesis privacy in public is a notion that is quickly becoming outdated. Second, the data collection is based on a public good, social and economic welfare of the city's denizens, which is a powerful interest, justifying its superseding of privacy.

This leads to the question of surveillance by the State creating these smart cities, as well as that data being transmitted to third parties, the private entities and contractors necessary for any smart city project. Without careful consideration, this smart city might end up having all of its data in the hands of technological monopolies[329].

A major challenge to informational privacy in this mixture of public and private bodies working together to create a "smart city" is that currently, the technology behind it is fragmented across all of those entities. This leads not only to possible leaks at some point in the many transfers of data between entities, but also security vulnerabilities.

[325] Ibid
[326] Edwards, L. (2016), Privacy, security and data protection in smart cities: A critical EU law perspective. *Eur. Data Prot. L. Rev.* 2 (2016): 28.
[327] Ibid
[328] Ibid
[329] Ibid

These vulnerabilities are well-known[330], such as a report by the FTC in 2015 on the Internet of Things noting security risks as its greatest worry both in terms of vulnerability of these actual devices and their potential to spread vulnerabilities through networks to other systems[331].

The Internet of Things is made up of a great number of small devices, each of which collects data but does not necessarily have a host of security measures to prevent it from being used for the wrong purposes; and these devices are usually "only secured as an afterthought, or worse, not secured at all, transmitting data in the clear."[332] Adding to that the fact that the devices' users have little understanding of their functioning (or even of their existence in certain cases) and are unable to make them more secure, IoT devices are a magnet for cybercrime.

Overall, new processes and technologies are working together to change how knowledge is created. This transformation is unprecedented and profound, and its consequences for the boundaries separating the public and private spheres and personal and non-personal data challenge these legal notions. We will now analyse the current European data protection framework, and how these technological developments are challenging its core principles.

The ability to gather data from everyone, all the time, is an entirely new way of gaining knowledge. It is no surprise that a new way of gaining knowledge requires similarly new means of regulation.

6. New Knowledge: Browsers and Cookies

Outside the Internet of Things, the primary contact individuals have with the realm of Big Data is through their devices, especially smartphones and computers. Every single action made on a computer may be recorded, whether it be something that clears as a survey or an order form asking for personal information for a delivery, clicks on a webpage or the movement of a mouse. Through the various interactions individuals have

[330] Ibid

[331] FTC Staff report Internet of Things: Privacy and Security in a Connected World, January 2015 at https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013- workshop-entitled-internet-things-privacy/150127iotrpt.pdf (herafter FTC, 215).

[332] Akamai, akamai's [state of the internet] report (2014), p. 1, at https://www.antel.com.uy/wps/wcm/connect/2e38bd0047ad6c9682d3e7af6890d810/q3-2014-state-of-the-internetreport+(2).pdf?MOD=AJPERES .

with their devices, the controllers of those websites can easily record and track individuals[333]. Because any connection requires identification, some exchange of data is necessary. In order to identify the individual over multiple visits, a website will install a cookie on the user's computer, a small text file which will contain information about the website and the user. Cookies can be stored for a lengthy amount of time, and most users are not aware of how to find, much less review or remove, these cookies[334]. Cookies can also be shared through multiple websites and across various services[335], which creates an invisible web in which individuals are reinforcing the data profiling operations of various third parties without being aware of it.

Originally, cookies, much like the rest of the Web, were varied and tailored to and by each website. As time went by, unified protocols and standards - Hypertext Transfer Protocol (HTTP), Uniform Resource Locator (URL) - ensured that everything about the connections between computers became easier[336]. As websites became more complex, organisations started creating cookies which could be identified through their multiple servers[337]. This laid the groundwork for the modern cookie: a tool which can be used to communicate from the user's computer with various servers to transmit information.

Through the use of cookies, it is possible to track individuals through multiple websites: if an individual has a cookie from website X, and then visits website Y, it is possible for website Y to identify the individual from that cookie[338]. Many online advertisers have such cookies, tracking and identifying individuals throughout multiple websites, and using the cookies to record data about user activity[339]. Variety being a major theme in the trends of Big Data, the ability to link together data from multiple sources is vitally important, and has exponential effectiveness as the number of sources grows.

---

[333] Hirsch, D. (2011), The Law and Policy of Online Privacy: Regulation, Self-Regulation, or Co-Regulation? Seattle University Law Review, Vol. 34, No. 2, 2011. Available at SSRN: http://ssrn.com/abstract=1758078

[334] Nenoist, E. (2008) Collecting Data for the Profiling of Web Users, in *Profiling the European citizen. Springer Netherlands*

[335] Brimsted, K. (2010), Behavioural Advertising: time to tame the cookie?, *Privacy and Data Protection, 11 2 (, 7)*

[336] Kristol, D. (2001). HTTP Cookies: Standards, privacy, and politics. *ACM Transactions on Internet Technology (TOIT)* 1.2 : 151-198.

[337] Ibid

[338] Lin, D., & Loui, M. C. (1998). Taking the byte out of cookies: privacy, consent, and the Web *(Vol. 28, No. 2, pp. 39-51). ACM.*

[339] Brimsted (n.335)

This practice has become very common across the industry[340]. Because more and more consumers have been deleting cookies on a regular basis in response[341], finding cookies which last longer and are more reliable has been the source of much research. Because cookies are a main way in which websites track the number of visitors to a website, in particular in order to attract advertisers towards a website, being able to gauge traffic accurately has an important impact[342].

Several practices have sprung up in order to create ''stickier'' cookies. An example is an online advertising company named "United Virtualities", which developed a "Persistent Identification Element", tagged to the user's browser but not deletable[343]. These "Flash Cookies" first came to prominence in 2005: they cannot be removed, have no expiry date, are stored in a different location from regular cookies, and are not controlled by the browser (and as such cannot be removed through clearing browser history). Even "Private Browsing" included in most browsers does not deter Flash Cookies, and a majority of major websites use them[344].

Cookies are being used in increasingly-creative ways to collect data. For example, some ultrasonic pitches have been embedded in various audio media - into TV commercials, browsers or apps - which are then detected by nearby devices, allowing cookies to pair a single user to multiple devices and keeping track of what the individual sees, how long they watch the ad, and how the person acts in reaction to the ad (such as buying a product, for example)[345]. There is a vast amount of value in creating a cohesive net capturing all the data linked to one individual, which prompts such creations[346], but also represents a challenge to one's informational privacy.

This cross-linking of data to ensure one individual is identified is not only limited to cookies, and is clearer than anywhere in the example of Google. In 2012, Google announced that all its accounts would be merged into one, whether they be Google+, Gmail, Youtube, or any other Google services[347]. One of the purposes of this ambitious

---

[340] Soltani, A,, et al. (2010). Flash Cookies and Privacy. *AAAI spring symposium: intelligent information privacy management*. Vol. 2010.
[341] Ibid
[342] Pierson, J., & Heyman, R. (2011). Social media and cookies: challenges for online privacy. *info*, *13*(6), 30-42.
[343] Ibid
[344] Ibid
[345] Soltani (n.340)
[346] Ibid
[347] The Guardian Staff and agencies (2012), Google user data to be merged across all sites under contentious plan, The Guardian, available at: https://www.theguardian.com/technology/2012/jan/25/google-merge-user-data-privacy

move was, once and for all, to have only one individual, one identity, and one account owner to apply one profile to[348], giving Google a very strong competitive position, to the point that Google was fined €4.34 billion for abusing its competitive position in favor of its search engine[349]. This move has been followed by multiple platforms, including Microsoft, uniting such wide services as Microsoft Office, Bing, Skype, MSN, or Microsoft's Cloud service OneDrive (the name "One Drive" being another indicator of this trend to unite, aggregate, centralise a full suite of services).

This push towards creating one entity, one account, and merging digital and real-world identities has several implications, but the most important one is that it becomes possible for one entity not only to possess tremendous amounts of data about a person's habits - tracking them from their e-mail to their office work all the way through their Cloud storage and online chat - but also become gatekeepers of that knowledge.

As we can see, data is being collected and centralised in wholly new ways, ways which allow us to gather a myriad of small, previously-unrecorded data points. As we will show later on, this change is part of what makes the notion of "personal data" more and more obsolete.

Overall, as we have seen, the "data collection" step, which should in principle be as unbiased and objective as possible since the data that is collected is raw, unchanged, actual data, in fact has a large number of factors which will determine what kind of Information can be created from that data. Depending on what data can be collected in the first place, what means the observer has to collect it, the limitations of the technology and the possible biases which drive the collection, the Information which will come out of the data collection will be entirely different.

This modularity of Information undermines the idea of "personal data" as understood by European data protection regulation. Under the "Information" paradigm, the same data can lead to privacy-invasive Information or not, depending on these factors. Nevertheless, one might argue that at the collection stage, the data collected can still be classified as "personal" and "non-personal". It is in the following stages that things become even more difficult. We will now investigate the storage and aggregation steps and show how the resulting "sculpture" becomes more and more uncertain.

[348] Ibid
[349] Commission Press Release, IP/18/4581 (Jul 18, 2018)

II. Storage and Aggregation: the "Volume" and "Variety" elements of Big Data

In order to explain the importance of the storage and aggregation steps, it is necessary to reiterate the fundamentals of Big Data: Volume, Velocity, Variety. The storage step is what dictates the "Volume" factor, as if data cannot be stored, it cannot be used, while the aggregation stage is where data from a Variety of sources come together in one database. Even though at that stage nothing is done to the data in and of itself, the simple fact that more data exists in one place is important, mainly due to issues of access.

1. What data is being stored

In the early days of computing, data was stored without much structure. Tailor-made (and time-consuming) programming models were necessary in order to understand the data[350]. The core limit here was the third "V": Velocity. The data was there, but not in a format that could be used effectively. This led to the rise of databases, and in turn to the appearance of data warehouses, buildings entirely dedicated to storing all that data in one place where it can be examined as a whole. At the same time, data became processed into forms which could make easier the linking of various pieces of data together: "structured data" started to appear[351].

Most data remains unstructured, generally, which means it does not follow a specified format[352]. Until recently, not much could be done with raw, unstructured data beyond manual analysis. This includes raw data, weather data, CCTV, seismic imagery, and so on: data which is created automatically by machines and captured in a form meant for using, not analysing, that data[353]. Unstructured data also comes from human action, whether it be social media, text messages, written forms, or e-mails.

Structured data systems are generally preferred by data analysis, as they are easier to mine for data, which has driven a push to make structuring data easier[354]. Nevertheless, structured data is not "superior" data; instead, it is simply data which has been processed

[350] Katal, A,, Wazid, M., and Goudar, R. (2013) Big data: issues, challenges, tools and good practices.*Contemporary Computing (IC3), 2013 Sixth International Conference on*. IEEE.
[351] Ibid
[352] Ibid
[353] Ibid
[354] Ibid

and modified in order to fit with a specific ability to analyse it. As such, this involves a variety of methods, which may impact the data and the outcome of the data analysis. This means that the structured data may be vastly different from its unstructured origins, and no method of structuring data is perfect or without bias or room for human error[355].

As technology advances, data becomes more and more structured, because of the value residing in such data. Nevertheless, some challenges exist which hamper this development. Studying these challenges, and the consequences were they to be overcome, is vital.

The main challenge facing Big Data today is not Volume or Variety, but Velocity. The amount of data being collected, and its variety, cannot be handled efficiently by the current systems, hence the push for more structured data, but some analyses require both a huge amount of data, and data that is very structured[356]. Satisfying both requirements presents technical difficulties, especially when it comes to various types of data each with its own specific features[357].

This is not the only challenge, however. The storage required for that huge amount of data is expensive, in particular due to social media and its various sensory data types, such as geolocation data or audio and video data. Because Big Data requires this large amount of data to be linked together, uploading, accessing and studying it takes a prohibitive amount of time[358]. Transportation of data between where it is stored and where it is studied can be done in multiple ways, but limitations in network speeds have led some companies physically transporting hard drives in trucks, such as Amazon's AWS Snowmobile[359].

"Big Data" groups together a vast array of tools and processes, all of which change what data is created and how. As we will show in later chapters, this means that traditional understandings of how to handle data may be outdated.

Data gathering has increased exponentially with the advent of the Internet of Things and the ubiquitous nature of smartphones, but data storage requires actual physical

[355] Tene, O., & Polonetsky (n.8)
[356] Ibid
[357] Ibid
[358] Ibid
[359] Miller, P., *Amazon wants to ship your data to the cloud using a literal truck,* The Verge, 30 November 2016, Available at: http://www.theverge.com/circuitbreaker/2016/11/30/13797212/amazon-aws-snowmobile-snowball-cloud-storage-truck

infrastructure, servers to hold the data, as well as ways to move these huge amounts of data from server to server. Because of that, eventually data storage becomes too expensive, and the idea of storing everything that can be collected is ill-advised. For example, Amazon has recently showcased a service called the AWS Snowmobile: instead of transferring one's data to Amazon's Cloud service directly and suffering the bandwidth limitations making this transfer impractically slow (when talking about petabytes of data), the AWS Snowmobile is a truck that will drive up, plug itself into the data controller's network, and store up to 100 petabytes of data[360]. It will then drive to Amazon's cloud storage data warehouse, and upload the information there for the data controller to access through Amazon's cloud service.

The fact that this is required is important to showcase the limits of Big Data technology - data collection has expanded to the point that it is faster to put it on a truck and drive it somewhere else than actually transfer it. The physical infrastructure has to take over where the digital one has failed. Meanwhile, some ventures are being made into whole fields such as using DNA for data storage[361], while flash storage, a faster but more expensive type of data storage than the traditional spinning hard drives, are considered by 40% of business owners to be too expensive to be practical[362].

This are just some examples of the roadblocks which limit the breadth of Big Data. It doesn't matter how much data is collected if it cannot be stored. These roadblocks are being overcome quickly, year after year, but there is no certainty that this will go on indefinitely.

The cost and logistical difficulty of data storage has some implications for every data controller, which means it has implications for the Information that becomes available. Storing and keeping huge amounts of data is more of a problem for some entities than others, while some entities do not actually need all of the data that they could collect to be stored. For example, a smartphone app could have permission to collect all manners of data, such as camera or GPS localisations. However, when that is not useful

[360] Hern, A. (2017), Amazon's Snowmobile will let you upload stuff by the truckload – literally, The Guardian, 5 December 2016, accessed 20/3/2017, https://www.theguardian.com/technology/2016/dec/05/amazon-snowmobile-upload-truckload
[361] Patel, P. (2017), Tech Turns to Biology as Data Storage Needs Explode, *Scientific American,* 31 May 2016, accessed 20/3/2017, https://www.scientificamerican.com/article/tech-turns-to-biology-as-data-storage-needs-explode/
[362] Bourne, J. (2017), Two in five execs grumble flash technology is too expensive, research finds, CloudTech, 29 July 2016, accessed 20/3/2017, https://www.cloudcomputing-news.net/news/2016/jul/29/execs-grumble-flash-technology-too-expensive-research-finds/

whatsoever to the data controller, there is no reason to collect it, as its storage will be an unneeded expense.

2. What data is aggregated

Aggregation, incarnating the "Variety" aspects of Big Data, is where the exponential value of data is achieved. There, through the aggregation of various types of data, a new database is created which has the potential to create new, powerful Information. Nevertheless, at the storage and aggregation stage, that Information has yet to be actually created - this is the role of the mining stage. Nevertheless, an important feature of aggregation is where the data comes from. The true power of data is in its ubiquity - individual and group profiling both work best when they have as clear as possible a picture of what they are looking for. This is a true challenge for data controllers, as while it is true that everything we do online and much of what we do in the physical world is monitored, there is no one observer, but a myriad of them.

Overall, the greatest scandals and upsets surrounding data protection did not revolve around just data collection, but about how that data aggregated. A recent example is the "Snoopers Charter", which allowed the UK government, on the one hand, to gather more data through requiring Internet service providers to store for a certain time many types of data, while allowing a large number of public authorities to access that data[363]. These organizations range from the police forces of every part of the UK, to the Department of Health, Home Office, or the Gangmasters and Labour Abuse Authority[364], and all can access the same pooled data.

The Snoopers Charter has led to severe criticism[365], but it is important to remember that little to none none of this criticism is directed at new collections of data. Instead, it is an increase in both storage and aggregation: storage because data is kept for longer and more diligently by these data controllers, and aggregation because all of these individual agencies can now aggregate that data with their own data to create new, powerful Information. As such, the Snoopers Charter does not actually involve a significant amount new collections of data or deployment of new data-collecting technologies.

[363] Cellan-Jones, C. (2016) Snoopers law creates security nightmare, BBC News, 29 November 2016
[364] Ibid
[365] Ibid

Meanwhile, the power of aggregation is not missed by data controllers. Every major software company is working on consolidating platforms that aggregate all of their data for profiling purposes. An example is Facebook, which has made very costly purchases in Instagram in 2012 ($1bn)[366] and WhatsApp in 2014 ($19bn)[367]. Through these purchases, not only has Facebook opened itself up to larger data collection, but then revealed that it would share data between the different services for ad targeting - in other words, data profiling of users throughout these platforms[368]. Additionally, "WhatsApp will also be sharing the data with the "Facebook family of companies"[369], which includes the aforementioned Instagram, as well as Oculus VR, which signals their clear intentions towards the aggregation of that data. The amount of money Facebook has invested to acquire these companies shows how important not just data profiling, but accurate and complete data profiling, is to their objectives and business model. In these cases, once again, there is no new data gathering taking place. The data Facebook, Instagram, and WhatsApp collect remains the same. However, once that collected data is aggregated, it leads to being able to create much more powerful, and much more valuable, Information.

A similar practice that we have studied previously is the use of cookies, not just to collect data, but to identify individuals across devices to link these devices together. In that case, once again, the ability to link devices together allows for the aggregation of data for data controllers to use to create new Information.

Finally, a relevant example, showing how ubiquitous is this push towards aggregating as much data as possible into one identity for better profiling, can be made from essentially all of the major software companies currently on the market. Microsoft has over the last few years united its accounters over various programs and services into one - Xbox Live, Skype, Microsoft Office, Windows, OneDrive, Bing and more, in order to (among other purposes) aggregate data over these various services[370]. Google has done the same, creating one account for Google Drive, Gmail, Google Search, Youtube and others.

---

[366] BBC News (2012), Facebook buys Instagram photo sharing network for $1bn, BBC Technology, 10 April 2012, accessed on 22/3/2017, http://www.bbc.co.uk/news/technology-17658264

[367] BBC News (2014), Facebook to buy messaging app WhatsApp for $19bn, BBC Business, 20 February 2014, accessed on 22/3/2017, http://www.bbc.co.uk/news/business-26266689

[368] Lomas, N. (2016), WhatsApp to share user data with Facebook for ad targeting — here's how to opt out, Techcrunch, 25 August 2016, accessed on 22/3/2017, https://techcrunch.com/2016/08/25/whatsapp-to-share-user-data-with-facebook-for-ad-targeting-heres-how-to-opt-out/

[369] Ibid

[370] Dailey, K. (2012), Could Google's data hoarding be good for you?, BBC News Magazine, accessed on 17/6/2018 at https://www.bbc.co.uk/news/magazine-16749076

Facebook is promoting its “logging in through Facebook” functionality and the aforementioned WhatsApp and Instagram. Apple, which has been known for a long time to have a strategy based around “owning the consumer”[371] and having a separate, complete environment providing all of the services an individual needs through one platform, has the Apple ID through which the consumer logs in to iTunes, the iCloud, FaceTime, Apple’s App Store, etc.

In conclusion, we can see that even after data is gathered, its ability to create Information will depend on what data is being stored and aggregated. Two data controllers with powerful data gathering processes might keep different data depending on their storage capabilities (leading to different Information which can be created), while two data controllers with the same data collection capabilities would find themselves with entirely new Information if they chose to aggregate that data in a shared platform.

So far, however, the data itself remain largely unchanged, despite being aggregated among other forms of data. As such, concepts of “personal data” can still hold, and attaching legal obligations to such data based on “localising” it in previous steps (whether consent was obtained during collection, whether it constitutes “personal data”, whether the data collected follows data minimisation purposes, and so on) is still possible. However, as we will show, the next step, that we call ‘’processing’’, and its transformation of data use the previous steps to create something that is altogether new. Put back into the terms of the “Sculptor’s Work”, if data collection is gathering the raw materials to use in the sculpture, storage and aggregation is the sculptor assembling all of those materials in one place, ready to begin the work.

III. Processing: Transformation Through Human Intervention

Once data is aggregated, it is processed into a form that can subsequently be mined for valuable Information. Because of how “big” Big Data is, human beings are completely unable to process that huge amount of data, and so leave the task to a wide world of software, professional data processors, or other data management services.

[371] Montgomerie, J., & Roscoe, S. (2013, December). Owning the consumer—Getting to the core of the Apple business model. In *Accounting Forum* (Vol. 37, No. 4, pp. 290-299). Elsevier.

1. How is the aggregated data processed

Processing Big Data is not a simple process. It requires complex software, trained personnel, and very powerful hardware. As such, once again, the limitations of cost restrict what the data controller may do, which brings to mind the "means likely to be used" test used for anonymised data. In this case, depending on the value the processor is expecting to gain from the Information obtained by the processed data, the tools used by the data processor will change, and with them the chances of informational-privacy-challenging Information to emerge.

There is a plethora of tools and methods used to process data and make it ready for mining. The objective is to obtain data that is both accurate, and valuable. Behind this seemingly-simple goal lies a highly-complex architecture, with constant innovation changing the status quo. The first main hurdle is to obtain the relevant data, and that hurdle is the one which is being resolved with the greatest speed, as we have seen (even if non-homogeneously). The next hurdle is to process that data into not just a readable and understandable form, but one that provides the highest possible value.

In the previous age, the age of "small data", the goal was to have the most precise data possible: few points of data meant that accurate data was vital, as any error in one of the data points would have a strong impact[372]. However, in the Big Data age, allowing for imprecision (what Viktor Mayer-Schönberger and Kenneth Cukier describe as "messiness" in their paper "Big data: A revolution that will transform how we live, work, and think."[373]) allows for the processing of much more data. If less work has to go into ensuring the relevance and accuracy of each data point, more data can be processed, and this idea was embraced quite thoroughly by the Big Data community, hoping that the greater amount of data processed would allow for more-accurate Information to be produced, offsetting the loss caused by the messiness.

Additionally, messiness is introduced based on the Variety of data points used. For every point of data introduced in the processing attempt that is of a different type, there is a chance that the tools used to process it are inefficient or inaccurate, which only becomes more likely for data which is very hard to process. For example, voice data is difficult to process if one is trying to use voice recognition, because voice recognition technology

[372] Mayer-Schönberger (n.322)
[373] Ibid

is, as of yet, imperfect. However, voice data is easy to process if one is only interested in the volume of the voice, as the tools to capture that data are more accurate. As such, a company recording phone conversations with customer service trying to process calls automatically using voice recognition might open themselves to a higher degree of inaccurate Information than one only trying to process the same calls to, for example, identify wheter individuals are talking loudly to their customer service (such as to identify patterns of potential staff abuse).

There are, along the same lines, certain types of data which are easier or harder to process. We have mentioned before that there exists “structured”, “semi-structured”, and “unstructured” data. Though the goal of processing is to produce structured data, the data used for the processing exercise can itself be any of these types. For example, two databases of structured data can be processed to create a new database combining them, while a myriad of unstructured data groups can be processed into one new structured database. In this example, the first case is generally much easier to process than the second one.

Between unstructured data types, as well, there are differences. Text-based unstructured data, such as e-mails, books or Word documents, can be processed relatively well due to the amount of support that exists for text search. Meanwhile, images, audio or video are much more difficult to process as the means to process them completely are not always as advanced[374].

The advantage of Big Data in terms of “messy” data is the law of large numbers - accuracies can be rectified by an overwhelming amount of positive data. This rule has been enormously successful. An example is the spell checker in Microsoft’s Word program in 2000[375]. Two researchers trying to improve it ended up failing to find a good solution through sophisticated algorithms, ended up feeding the program huge amounts of data, by a factor of a thousand, with amazing results such as accuracy increasing from 75% to 95%[376]. Because the data used there was very messy, it showed that it is possible for very unstructured data to lead to very powerful, useful and accurate Information. As such there is no clear rule for how to create accurate Information - it can come from unstructured data or structured data, from large amounts of data or small ones. Overall, however, there is a tendency for structured data, lending itself to well-developed tools, in large quantities, to construct good Information.

[374] Ibid
[375] Ibid
[376] Ibid

What does that mean in terms of informational privacy? The main observation is that tools that the data controller has at his disposal, or chooses to use, will significantly change the Information obtained. Some programs are better for unstructured data, while some are designed to handle structured data. Some programs are good at financial data while some are best used to identify language patterns. Data processing is a complex environment, and as such the same aggregated data can lead to different Information based on how the processing is done.

2. Who is doing the processing

We have mentioned how the processing stage, due to its complexity and the need for expertise for it to function, can be delegated to third parties. These third parties - data processors - are specifically targeted by the GDPR because of their unique situation - holders of large amounts of data without having the responsibility of being the data controller. As such, despite technically being third parties, they are under the GDPR responsible for many of the same obligations as data controllers. An important factor is that despite being contracted by the data controller to study the data, the decision-making process of the data processor is different from the data controller's[377]. Biases in data analysis in either of the parties are compounded when they are both involved in the processing.

Meanwhile, this raises the issue of aggregation once more - the same data processor might have a myriad of data points from various controllers. If they were to combine that data, though they might have more complete Information, it would be a breach of privacy. This is usually remedied by contractual clauses, however, showing that restrictions over what data is shared depends on not just regulatory restrictions, but contractual ones as well.

3. What is the purpose of the processing

[377] Article 29 Working Party (2010), Opinion 1/2010 on the concepts of "controller" and "processor" WP 169, 16 February 2010, http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp169_en.pdf

Finally, it is vital to note that because processing is usually aimed at discovering specific Information, the transformation of the data is not a simple ‘’increase’’ of the Information which can be obtained from it. In the case of sensitive data, for example, this processing can involve anonymization which attempts to keep some value in the data while protecting individual privacy. In this way, depending on the aims of the processing and the priorities of the data controllers and processors, the result of the processing can take different forms. Importantly, this means that a lot of the data gets discarded over the course of processing. Let’s take the example of a company trying to track the popularity of a certain product. Aggregating the sales records of its stores, that data is then processed to output only one statistic - how much of that product was sold over a certain period. In this example, the processing stage strips out almost all of the data. It does creates very accurate Information, the Information that the data controller was interested in, yet the data produced by this processing has very little chance of infringing the privacy of customers. Indeed, as the processed data is just a metric of products and quantities, all of the data which relates to them is trimmed out of the analysis.

The idea of various possible purposes changing the data analysis is reflected in the GDPR as well, since one of the major notions is the idea of ‘’purpose limitation’’ - under Article 5 of the GDPR, each purpose of processing has its own legal basis[378], determines what the data retention period should be, as well as how much data can be processed to accomplish this purpose. Ideally, each purpose of processing will be a flow, from collection to deletion, all following one distinct purpose. However, practices such as data aggregation and group profiling can take these flows and muddle them, mixing data and creating new Information from it[379]. This new Information, new ‘’data’’, may well be ‘’personal data’’ but is not covered under the purpose for which the data was originally collected. In those cases, and considering that this Information can appear even without the data processor having intended to create it, trying to limit data processing to one ‘’purpose’’ may be impossible[380]. These kinds of issues are one of the reasons that the oft-criticized ‘’legitimate interest’’ ground for processing data has been included in the GDPR, as a begrudging admission that some flexibility will be necessary[381].

[378] GDPR, Article 5

[379] Rauhofer, J. (2015). Of Men and Mice: Should the EU Data Protection Authorities' Reaction to Google's New Privacy Policy Raise Concern for the Future of the Purpose Limitation Principle. *Eur. Data Prot. L. Rev.*, *1*, 5.

[380] Goodman (n.28)

[381] Moerel, L., & Prins, J. E. J. (2015). Further Processing of Data Based on the Legitimate Interest Ground: The End of Purpose Limitation, IAPP News, accessed at https://iapp.org/news/a/on-the-death-of-purpose-limitation/ on 11/10/2018

The fact that data, under the GDPR, is held to be personal or not at the collection stage overlooks the importance of this processing stage: processing personal data might lead to non-personal data, while processing non-personal data, through the finding of unseen correlations, may create personal data. But more importantly, every instance of processing changes what the data is, and this change needs to be taken into account, regardless of whether it crosses the boundary between personal and non-personal data.

Overall, processing and how it is performed changes what Information can be obtained from it, depending on the means used by the data controller or processors. There are many types of data and ways it can be processed, based on the existing technologies, which ones are available to the controller or processor, which ones they actually choose to use, how they use it, and for what purpose. The fact that these decisions can be split between various parties - data controllers and processors - only adds to the variance in how the data is transformed and what Information can be created from it.

Going back to the Sculptor's Work analogy, there are infinite forms a sculpture can take from the same materials, and every sculpture fulfills a different purpose. As such, the materials the sculpture is made of can be of less importance than the way it is shaped.

IV. Data Mining: Further Opportunities for Bias

Once data is collected, aggregated, stored, and processed into a structured or semi-structured database, it becomes mined for valuable information. That data mining is where value is extracted from the data - the "mountain" of data being "mined''. In that stage, the processed data is studied to create finally the Information which the data controller is looking for. As such, the data mining and processing stages are intrinsically linked and are often performed by the same entity. Nevertheless, it is important to separate them because the data mining can be performed by anyone who has access to the processed data. The basic consequence of this is that some data can essentially only be mined appropriately by the processor, if the formatting is done in a certain way. An example is pseudonymised data: data which can only be identifiable if one has access to additional Information. The same database processed into pseudonymised data can lead either to no or complete Information in the data mining stage depending on whether the observer has access to that additional Information.

Going back to the Sculptor's Work analogy for the data/Information conundrum, the data mining stage is the one most prone to bias. At this stage, the sculpture is finished, and the sculptor looks at his work, hoping to find in his complete work what he built it for. Because "beauty is in the eye of the beholder", whoever will look at that structure may have a different vision of it. Their own expertise, their experience, their biases, expectations, and contextual information will all come into play when it comes to evaluating that structure and using it for their purposes. Because of the amount of subjectivity that comes into data mining and how varied the Information that it creates is, the data mining stage has many facets.

1. How is the data mined

Mining the data can be done in a vast number of ways, which has a strong impact on what the outcome of that mining is. Because each method of data mining changes the information created by it, and suffers from different limitations and biases, the term "mining" - which puts in mind the idea of an external effort on a mountain of data - is inaccurate. A preferable term is "data sculpting": select pieces of the mountain of data are used to construct something specific the creator had in mind, which changes depending on the sculptor and the observer, created for a purpose and function. The same mountain can lead to an infinity of different sculptures depending on these factors, and the same data can lead to different Information.

The most simple type of data mining does not require particular expertise and is done every day - taking two data points and creating a correlation from it: "A happens in 50% of cases when B happens". These correlations are "linear correlations", with a clear pattern of cause to effect[382]. Data mining uses many tools, primarily found in the AI community - which develops processes such as machine learning and pattern recognition - and the mathematics community. There are two main approaches to data mining: top-down and bottom-up[383].

Top-down analysis starts by looking for confirmation (or lack thereof) from established patterns and models of behavior. This allows for the identification of patterns which follow the norm, or depart from it[384]. As such, it primarily aims at identifying and understanding behaviors. Because it tests a hypothesis, when the data becomes too complex and too

[382] Hildebrandt M., Backhouse J. (2005), Descriptive analysis and inventory of profiling practices. In *FIDIS Project Deliverable 7.2.* Available at: http://www.fidis.net
[383] Ibid
[384] Ibid

many variables are involved, accuracy drops. This is an issue because it is often the case that important variables which impact the outcome of the study are either not considered, or not kept through the collection, storage, and aggregation phases and as such not included in the data mining process.

Bottoms-up analysis manipulates the data itself to uncover patterns, attempting to generate hypotheses that can explain behavior or predict future behavior[385]. This type of analysis has no real interest in causation: it observes behaviors and the link between them without looking at the reason why these links are the way they are, leaving the interpretation to the observer. As interpretation can be flawed or biased, bottoms-up analysis is also imperfect.

For Big Data analysis, use of computer algorithms is almost always required, as manual analysis is prohibitively time-consuming and expensive as the Volume and Variety of data increases. Nevertheless, dealing with complex data holds complexities at every level. Choosing the framework in which the analysis is performed - whether it be which program is used, which data is examined, or what to do if unexpected events occur and how best to direct the algorithm - are all major factors in deciding what happens to the data[386].

The use of more and more data in data processing can reveal new, unexpected correlations, depending on the tools used for the data mining; and as the phenomenon of Big Data expands, these unexpected correlations become more common[387]. Because the conclusions might not necessarily match what was expected of the mining, there can come a time where a data controller, who has agreed not to use the “personal data” in ways beyond the stated purpose (according to the purpose limitation principle[388] found in the GDPR[389]), finds him/herself with new Information, that goes beyond that purpose, without even meaning to do so. Meanwhile, that Information was obtained while using the data for its stated purposes – in that case, is there a breach of purpose limitation or not? The difficulty to outline properly what can be obtained from data processing causes privacy statements to be vague in order to give data controllers room to manoeuvre,

[385] Ibid
[386] Tene, O., & Polonetsky, J. (2013). Judged by the tin man: Individual rights in the age of big data. *J. on Telecomm. & High Tech. L.*, *11*, 351.
[387] Colonna, L. (2013), Mo’ Data, Mo’ Problems? Personal Data Mining and the Challenge to the Data Minimization Principle, *Conference proceedings of Big Data and Privacy: Making Ends Meet hosted by Stanford Law School and The Center for Internet and Society*
[388] Mayer-Schonberger, V., & Padova, Y. (2015). Regime Change: Enabling Big Data through Europe's New Data Protection Regulation. *Colum. Sci. & Tech. L. Rev.*, *17*, 315.
[389] GDPR, Recital 28

weakening their overall effectiveness and transparency, while making it harder for individuals to understand the processes involved.[390] According to the Article 29 Working Party, the purpose limitation is quite strict in that the GDPR Article 5(1)(b) requirements of the purpose being "collected for specified, explicit and legitimate purposes"[391] require the purpose of the collection to be "clearly and specifically identified" and "detailed enough to determine what kind of processing is taking place"[392].

As such, there is an irremediable limitation there, where data controllers are forced to stretch as much as legally possible their purpose specification, just in case the data they discover happens to go beyond the scope that was originally planned. This is, once again, because "data" and "Information" are different constructs, and the final construct is so different that binding it by the rules of the raw data it was built from makes less and less sense as these constructs become more elaborate. In other words, going back to the Sculptor's Work analogy, the GDPR asks the sculpture's commissioner to decide exactly what they will see in the completed work even though at that point all they have is the raw materials in a pile.

Thus, depending on what data is used, unexpected correlations can appear. However, these correlations are not necessarily correct, either. Because the tools used to mine the data are varied[393], and because the Information which can come from it can be complex, unexpected, or incomplete, data mining can lead to inaccuracies[394]. Data analysis is a whole set of skills which requires expert knowledge, and as such the same data might lead to different Information depending on how these skills are applied.

This is an important reason behind the separation between "data" and "Information". "Data" is rarely inaccurate: it is how that data is framed, interpreted, understood, contextualised - the "Information" - which can be inaccurate. Data mining produces Information, not necessarily facts.

---

[390] Mantelero, A. (2014). The future of consumer data protection in the EU Re-thinking the "notice and consent" paradigm in the new era of predictive analytics. *Computer Law & Security Review*, *30*(6), 643-660.
[391] GDPR, Article 5
[392] Article 29 Data Protection Working Party (2013), Opinion 03/2013 on purpose limitation, WP203, Adopted on 2 April 2013
[393] Altman, M., & McDonald, M. P. (2001). Choosing reliable statistical software. *PS: Political Science & Politics*, *34*(3), 681-687.
[394] Mohan, P., Thakurta, A., Shi, E., Song, D., & Culler, D. (2012). GUPT: privacy preserving data analysis made easy. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data* (pp. 349-360). ACM.

2. Who is doing the mining

On top of the purpose limitation considerations come the aforementioned fact that different observers might have different understandings of the same data. This simple observation, which we will develop further in the coming section, means that the Information can be misunderstood or manipulated depending on what Information the observer wants to obtain. An example is confirmation bias[395], and is found in every level of data processing, from individuals self-monitoring their activity and limited by their limited sample size of 1[396] to large corporations looking to prove the worth of certain products despite signs to the contrary, to political parties only taking into account statistics that match their worldview[397].

Because of this, data mining is not as simple as making a conclusion based on the processed data. Instead, interpreting the data in order to obtain relevant and accurate Information is a whole world unto itself. This is another reason for the importance of distinguishing “data” and “Information”. “Data” in and of itself says nothing - it is only through the “Information” we create from it that decisions are made, that inaccuracies, prejudices and biases show their effects.

Through this section, we have gone over the major stages of the processing of data. As we can see, this process is not a clear line from “messy data” to “clean facts”. It is not mining a mountain to uncover a hidden gem. It is in fact a lot more akin to sculpting the data in a specific shape: the work is just as much based on the materials you start with as it is on the person doing the sculpting, and how and why they do so. There is a whole host of decisions that are made at every stage that will influence the final output, every single one changing, in subtle or not-so-subtle ways, the Information that can be obtained from it. We will now look deeper into specific issues and patterns that find themselves present in every stage of the process, in order to discuss architectural facts which mean the “personal data” binary is unable to reflect that reality and even attempting to anchor data points to an individual is unsustainable.

Overall, as we have seen, the process of data analysis is complex, and involves many steps, each of which relies on various different pieces of technology which themselves

[395] Swan (n.7)
[396] Ibid
[397] Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 755-769.

have inherent limitations, but together constitute an extremely powerful tool which creates knowledge in entirely new ways. This new knowledge brings to light underlying data protection challenges. We will now study a particular type of data processing exercise which creates the ability to obtain information about individuals, in a way that was never possible before.

B. Profiling and Group Profiling: A New Way of Creating Knowledge and Predicting Behavior

I. Defining Profiling

We will now study a few of the distinctions in profiling, to highlight the opposition between the types of profiling that came before, and the ways profiling has changed and is now challenging traditional understandings of data. The GDPR defines profiling as ''any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person''[398] A more thorough definition was outlined in "Defining Profiling"[399] along these lines: "Profiling is a technique to automatically process personal and non-personal data, aimed at developing predictive knowledge from the data in the form of constructing profiles that can subsequently be applied as a basis for decision-making. A profile is a set of correlated data that represents a (human or non-human, individual or group) subject. Constructing profiles is the process of discovering unexpected patterns between data in large data sets that can be used to create profiles. Applying profiles is the process of identifying and representing a specific subject or to identify a subject as a member of a specific group or category and taking some form of decision based on this identification and representation."[400] This definition gives us further elements which we will develop to assess the impact of profiling on informational privacy.

Organic profiling is profiling as performed by non-human entities, which is developed as a survival trait. An animal understanding that someone is predator or prey depending on

[398] Ferraris, V. and Bosco, D. and Cafiero, G. and D'Angelo, E. and Suloyeva, Y., Defining Profiling (December 11, 2013). Available at SSRN: http://ssrn.com/abstract=2366564 orhttp://dx.doi.org/10.2139/ssrn.2366564
[399] Ibid
[400] Ibid

the input of information from its senses is organic profiling. Human profiling, by opposition, is different because it is reflective: we can understand the conclusion we draw, and study and perfect them[401]. Finally, machine learning is done based not on mechanics decided by survival evolution, but by the input of a man-made software architecture[402]. As we can already see, a vital difference between human and machine profiling lies in how decisions are made, and in their accountability - a human may understand and rectify their biases, where a computer or animal cannot.

Another distinction regards this factor: the reasoning involved in the decision. Animals and machines perform automated profiling - the aggregation and processing of data where no decisions are made based on outside reflection[403]. Meanwhile, autonomic profiling is a process where there is human intervention and reflection, but at a low level - the machine makes all of the decisions, with a minimized human role. Finally, non-automatic profiling is where machines are not involved in the decision-making process[404]. As we are focusing on Big Data and its implications, we are focusing this study on autonomic profiling.

Finally, an important distinction when it comes to profiling is the difference between group profiling and personalised profiling. As mentioned before, profiling creates a correlation between pieces of data. Once the data is mined and the correlations established, a set of assumptions is created. These assumptions however can be made about two types of entities: an individual, or a group. For example, if one finds data correlating incomes with shopping patterns, one can either create assumptions such as “Individuals with incomes of X or higher buy more of product Y”, or “Particular individual A with income of X buys a lot of product Y”. Both profiles - the group one and the individual one - can be built on the same data, but have different implications.

To understand the implications of group profiling, a distinction needs to be made between distributive and non-distributive profiles. A distributive profiles identifies a group where all members share all of the attributes correlated in the profile. These profiles have certain implications because all members of the group will be assumed to follow certain characteristics, while this is actually rare[405]. For example, one could assume that all individuals in an extremely poor neighborhood would be poor and apply a group profile

[401] Ibid
[402] Van der Hof, S., & Prins, C. (2008). Personalisation and its influence on identities, behaviour and social values. In *Profiling the European Citizen* (pp. 111-127). Springer, Dordrecht.
[403] Ibid
[404] Ibid
[405] Ibid

to them, using it for further constructions later on such as targeted advertising. In fact, most profiles are non-distributive, which means they are probabilistic: every member of the group is *likely*, not *assured*, to share certain characteristics[406].

This phenomenon is at the source of many instances of data inaccuracy: on the one hand, most profiles are non-distributive, which means blanket assumptions made on the individuals will be wrong at least some of the time. On the other hand, data controllers will want to use them, even when there is a chance their profiles will not apply some of the time. When the probabilities are wrong, or even the assumptions are wrong, the consequences on informational privacy can be significant.

II. Profiling in the Law

Profiling was not addressed directly in the Data Protection Directive, mainly because the Directive was created in a context where profiling and its implications in conjunction with Big Data were not a major concern. However, the GDPR now has in its Article 22 provisions specifically addressing it[407]. Under this Article, data subjects have a right not necessarily to avoid the exercise of profiling, but instead to avoid being "subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."[408] Recital 58 provides as examples the "automatic refusal of an on-line credit application or e-recruiting practices without any human intervention."[409]

In the GDPR, profiling is defined in the context of data protection as "any form of automated processing of personal data, intended to analyze or predict the personality or certain personal aspects relating to a natural person, in particular the analysis and prediction of the person's health, economic situation, performance at work, personal preferences or interests, reliability or behaviour, location or movements."

This Article has some important implications. A concern about the automation of decision-making falls under the field of discrimination, as unthinking machines - machines using the automated profiling explained above - use no reasoning to come to their conclusions, which means if they are not programmed to avoid discrimination, there

[406] Ibid
[407] GDPR, Article 22
[408] GDPR, Article 22(1)
[409] GDPR, Recital 58

can be some breaches of the right to non-discrimination (protected amongst others by Article 14 of the European Convention on Human Rights[410]). Specifically, GDPR Recital 71 states that data controllers need to "implement appropriate technical and organizational measures" that "prevents, inter alia, discriminatory effects" on the basis of processing sensitive data ("sensitive data" being data considered especially personal such as race or religion)[411].

This focus on "sensitive data" in both Article 22 and Recital 71 are relevant. They can either be interpreted as applying these extra protections only to sensitive data. This would involve carefully piecing together the data points which make up the data processing and finding whether some of them are sensitive[412]. However, a second interpretation is that any data processing operation containing some sensitive data needs to have this level of protection. In either case, it involves pinpointing the sensitive data being examined, which can be difficult - the correlations created by data become more and more complex and difficult to identify the bigger the dataset becomes, and the more likely it is that both some sensitive data appears, and that sensitive data becomes harder and harder to assess. Overall, paradoxically, the bigger the data, the harder it is to find specific information within it.

The ability to create profiles about individuals is a new tool in creating Information. All of this new Information has been considered to be increasing and improving human knowledge, but that is not necessarily the case[413]. In fact, what makes data different from Information is the human element, which leads to the potential for human error[414].

C. The Uncertainty of Human Agency

Now that we have gone over each part of data processing, we will study several global factors which are present across various stages, and which ensure that what comes out of the processing of data is not simply "better" data, but instead is a whole new construct. Because of how different that construct can be, any ties to the individual can end up

[410] ECHR, Article 14
[411] GDPR, Recital 71
[412] Goodman (n.28)
[413] Pauleen, D. J., & Wang, W. Y. (2017). Does big data mean big knowledge? KM perspectives on big data and analytics. *Journal of Knowledge Management*, *21*(1), 1-6.
[414] Carmichael, L., Stalla-Bourdillon, S., & Staab, S. (2016). Data mining and automated discrimination: a mixed legal/technical perspective. *IEEE Intelligent Systems*, (6), 51-55.

disappearing, making some data which seems “non-personal” still able to allow the creation of very personal Information.

I. Data and Information Inaccuracy

The first of those factors is data inaccuracy. Because the various steps in the process are fraught with choices, complexities and biases, the data obtained and processed at various stages can end up inaccurately representing reality.

Data inaccuracy is a major concern for controllers for a simple reason - data accuracy is a legal requirement under Article 5(1) of the GDPR, which states that “Personal data shall be [...] accurate and, where necessary, kept up to date”[415]. Data inaccuracy also has economic downsides, as inaccuracy leads to less benefit and less utility to the Information it creates. However, when it comes to informational privacy, several issues arise, the main one being the potential tarnishing of one’s reputation by wrongful but negative facts being treated as accurate. Another issue is discrimination based on inaccurate data, such as an individual being wrongfully considered to have bad credit because of a mistake somewhere in the data gathering or mining process.

As we have seen, how you create your Information is as important as how you create your data, and despite it looking like data is a mountain that has to be mined, there is a lot more human involvement at every stage of the process - including the creation of these algorithms. As such, evaluating all Information creation under the same rules, be it by an algorithm or an individual, and taking into account the fact that there is no such thing as objective Information, would allow one to succeed where the GDPR struggles in dealing with accurate data leading to inaccurate or discriminatory Information.

An example can be seen by going back to the Google Flu Trends example we have outlined earlier in this thesis. Though Google’s attempt at discovering outbreaks of the flu using search engine data was quite successful, the way the data was analysed deserves a second look.

Google’s engineers obtained their results by going through millions of searches looking for 1152 data points[416], removing the data points that they deemed irrelevant. However,

[415] GDPR, Article 5
[416] Ibid

that method failed to detect a non seasonal influenza pandemic in 2009, because it was mainly interested in seasonal patterns and could not predict such a pandemic which would happen outside of winter. Meanwhile, despite improvements in the service, Google's service has overestimated flu trends consistently, with certain patterns: for example, the direction and magnitude of the errors varies with the season, showing that these inaccuracies could have been avoided by other research methods[417]. Unpredictable behavior such as sudden media frenzies, unexpected weather patterns, or the appearance of whole new terms not taken into account by the model are all able to render Google Flu Trends' results inaccurate[418].

Meanwhile, Google has been attempting to extend the reach of its predictions by modelling software through the Google Correlate tool, which can be modified by service providers according to their business model[419]. This leads to so-called "blue team" errors, where the service providers' changes and decisions on what data to take into account or not for their data mining efforts are mistaken. This means that data that has been collected, stored, aggregated, and processed correctly, using Google's experienced, powerful tool, can still, nevertheless, lead to inaccurate Information based on who uses it.

To add to the complexity, there is the possibility of "red team" attacks. This case, a true example of the wider digital world in action, is what happens when the subjects of the research themselves attempt to manipulate the data for their own goals, whether for economic or political benefit. An example is attempting to make one's interests trend on Facebook or Twitter, to spread rumours, smear political candidates, or change perceptions of what is happening based on what is shown through various services[420]. The brigading of online polls is a common example of this.

II. Technical Limitations of Big Data

In addition to the uncertainty linked to the human-error element of data processing comes the difficulty in the process itself. There is a wide variety of tools, with various degrees of reliability and complexity, with various possible uses and requiring a whole array of skills to use effectively. A simple look at the vast literature related to data mining techniques

[417] Ibid
[418] Ibid
[419] Ibid
[420] ibid

is enough to show this, as well as the fact that data science in and of itself is an entire scientific discipline[421]. Because of this, and because of the inherent difficulties in assessing what kind of Information can be obtained from various data processing exercises, there is a lot of variance in every step of the process.

This is a limitation we have mentioned before, when discussing anonymisation, which is a type of data processing in and of itself. Because of the various possible processes involved, we held that anonymisation was not an ultimate solution, did not lead to certain results, and as such cannot be trusted to protect privacy on its own, a view which is gaining support in the privacy community[422]. The exact same is true of all aspects of data mining, and this complexity, especially at steps which happen after the collection stage (at which the current regulatory system decides if data is personal or not), means that how the data is "sculpted" changes not just based on the biases of the observer, but also on the resources they use and how they use them.

Finally, a common theme in every step of the data processing is the fact that the forces attempting to improve and expand the capabilities of data gathering activities are getting better and better at it in an exponential way - the rapid progress of Big Data. That does not mean that data is becoming more accurate - as we have seen more data analysis does not necessarily imply more data accuracy - but it does mean that the power of these processes, and the Information created from them, are becoming greater and greater. This creates changing power dynamics. The different speeds at which different developments occur each change the balance of Information creation: giving more power to those who can benefit the most from these innovations, while not necessarily compensating it with safeguards for those whose data is being processed. As such, it is not just the case that Information creation is uncertain because of inaccuracies based around human error and bias, and because of the complexities of data science. It is also uncertain in a third dimension, time.

[421] For some examples, see "Berry, M. J., & Linoff, G. (1997). Data mining techniques: for marketing, sales, and customer support. John Wiley & Sons, Inc." for an early example from the late 90s, "Berkhin, P. (2006). A survey of clustering data mining techniques. *Grouping multidimensional data*, *25*, 71." for a specific look at how complex various parts of the process are, or "Agrawal, R., & Srikant, R. (2000, May). Privacy-preserving data mining. In *ACM Sigmod Record* (Vol. 29, No. 2, pp. 439-450). ACM." for a look at how data mining practices attempt to preserve privacy.

[422] See "Ohm, P. (2010), Broken Promises of Privacy: Responding to the Surprising Failure of Anonymisation*. UCLA Law Review, Vol. 57"*, or "Barocas, S., & Nissenbaum, H. (2009). On notice: The trouble with Notice and Consent."

Conclusion

In this chapter so far, we have shown how the raw data present in the world - the ‘’raw materials’’ the sculptor will use - leads to a unique data set which contains a number of elements affecting its contents and characteristics - the “sculpture”. We have shown that the process which leads from one to the next transforms, transfers, prunes and aggregates the data in a way which can be so profound as to render it wholly unrecognizable. Compounded with the complexities of interpreting data, this also means the Information obtained by this process, given the same data, can be completely different. This has many effects, chief amongst which being the fact that “personal” data can lead to non-personal Information, and vice versa. However this is a mere side-effect of the huge difference between the data before and after the application of the various steps. In fact, this difference means that it is often pointless to judge the data and the Information which can be created from it based on its state in previous stages.

It is important to remember that this phenomenon is not always so pronounced as to be very noticeable. Most of the time, the database will be a list of consumer addresses, or the aggregation of user feedback forms showing clear patterns. The fact that the transformation of data throughout the different stages is usually not so drastic is a main reason why the current regulatory framework has continued relying on the personal data binary. It is the age of Big Data, and the increasingly-complex tools used to process it, which has changed the status quo, and even then most data operations are still akin to the relatively-simple ones described above.

Chapter 4. Informational Privacy in EU Law: Challenges in Data Protection and Privacy Law

Introduction

Earlier in this thesis, we have laid out an analysis of what privacy and data protection are. We have then set into place the various stakeholders and the processes which have created the environment in which we are evolving. We have shown that new technologies which are being developed create new ways to approach data, and that the processes of data collection, aggregation, mining, and use are becoming more complex and wide-reaching.

We will now analyse how this new reality is clashing with the classical EU approach to data protection and privacy. We will do so by identifying how some essential parts of the EU framework protecting those rights are facing challenges, and then identify the roots of those various challenges. Through this, we will show that there are architectural limitations to this classical framework, and that the signs of a new approach show that the EU is shifting away from this classical framework.

This new approach to data protection (and informational privacy), as we have shown in previous Chapters, consists of assessing the ‘’risk’’ involved in processing and mitigating risk-prone processing with additional safeguards.

A. Challenges to EU Data Protection Law

The first aspect of EU Data Protection Law we will examine is the main way that data which holds some of the value of “personal data” can be examined without the dangers to informational privacy associated with the processing of that personal data, and as such a vital tool in ensuring the protection of individuals: anonymization. As we will show, anonymisation is limited, which highlights the limitations of the concept of ‘’personal data’’. We will then analyze another such core part of EU data protection law, consent, and similarly show its limitations in regards to protecting informational privacy in the age of Big Data.

I. The Limits of Anonymisation

In the European context, anonymisation is the set of technological practices aimed at ensuring that personal data is separated from its "personal" component, which is achieved in the EU conception by ensuring individuals cannot be "identified"[423]. The Article 29 Working Party published an Opinion in 2014 on anonymisation[424]. In this Opinion, the Working Party argued that the concept of "anonymisation" in the European view is closely linked to that of "identification". This shows that European authorities approach anonymisation with the same view as the one behind the distinction in what constitutes "personal data": the "identified" requirement. Because personal data can be mined for great monetary value, but at the risk of endangering individual privacy, anonymization - ideally - would allow one to obtain that value while protecting the right to data protection (ie without identification).

1. The "Means likely to be used" test.

The GDPR refers to anonymisation in its Recital 26, and excludes anonymized data from its framework:

"To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly.

To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.

The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or

[423] GDPR, Recital 26
[424] Article 29 Working Party (2014), Opinion 05/2014 on Anonymisation Techniques, WP216, Adopted on 10 April 2014

to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable.[425]".

The extra test which Recital 26 of the GDPR specifies for identifiability, "the means reasonably likely to be used", shows that the regulators are aware of an ever-present technical risk: anonymisation is not one barrier which stops all possible identification attempts, but rather a set of various technical practices which may protect data from identification to a certain (or uncertain) extent[426]. If an entity deploys "unreasonable" efforts in ensuring identification, the failure to stop it is not a violation under the GDPR.

Finding where anonymisation becomes sufficiently protective can be challenging. To the Article 29 Working Party, even de-identified data is still personal if it can be attributed to one person, whether or not any other characteristics apply - the only way to make that type of data anonymous would be to aggregate it to the point where no identification is possible[427]. Essentially, to the Article 29 Working Party, there is no such thing as a non-personal dataset created from only one data subject[428]. This means that as long as the data set is made from one individual, no matter how anonymized or innocuous, it constitutes personal data.

Even when data is anonymised, there is still a possibility of de-anonymization or of re-identifying the individual from other, less obvious pieces of data. This makes it very difficult to turn personal data into ''non-personal'' data falling outside the scope of the GDPR.

Anonymisation is only effective at limiting identification when the effort needed to de-anonymize data is prohibitively time- or resource-intensive compared to the effort needed to anonymize that data in the first place. However, in the last couple of decades, research into anonymisation has shown that true anonymisation is impossible, and the means of re-identifying are becoming easier to obtain[429]. This puts anonymisation in a difficult

[425] GDPR, Recital 26
[426] Bayardo, R. J., & Agrawal, R. (2005, April). Data privacy through optimal k-anonymization. *In Data Engineering, 2005. ICDE 2005. IEEE.*

[427] Article 29 Working Party, Opinion 5/2014 on Anonymisation Techniques, WP216
[428] Article 29 Working Party, Opinion 5/2014 on Anonymisation Techniques, WP216
[429] Ohm, P. (2010), Broken Promises of Privacy: Responding to the Surprising Failure of Anonymisation*. UCLA Law Review, Vol. 57, p. 1701*; U of Colorado Law Legal Studies Research Paper No. 9-12. Available at SSRN: http://ssrn.com/abstract=1450006

situation, as the practice is widely used to protect personal information, especially in the medical and research contexts[430], but cannot be completely trusted.

The distinction based on “means likely reasonably to be used” provides some flexibility since anonymisation only needs to protect up to a certain threshold, but the shift in balance of power between anonymization and de-anonymisation means that under this distinction, anonymisation technology does not mean de-identifiability, and as such does not change the legal nature of the data, and is simply a security measure.

“Anonymisation” is not one process which when applied “increases anonymity”. Instead, it is a wide array of different technologies, each of which has its own advantages, flaws, and methods. For example, the process of “k-anonymity” is aimed at ensuring that the information of each person in the data set cannot be distinguished from other individuals in that data set, but achieving that (or coming as close as possible, as no anonymisation is perfect) can be done in multiple ways.

Because of this, re-identification can also be done in various ways. The main difficulty is that with enough other pieces of data being aggregated, as long as the data has any relation to the individual (which is always true, since the data is generated by the individual’s behavior), there will be pieces of data which may allow re-identification. Because of the unpredictability of that process, efforts to ensure true anonymity are never perfect.

An example of how the “means likely to be used” can be unpredictable is the AOL search data situation[431]. In 2006, AOL had released twenty million search queries for researchers to mine for data. At that point, in the view of AOL, that data was considered fully and irremediably anonymised - meaning they did not have the “means likely to be used” to re-identify it[432]. However, as they quickly realized, this did not mean anonymisation: reporters from the New York Times showed that they could easily re-identify parts of the information. They were able to identify an individual based on their searches:

“[S]earch by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for "landscapers in Lilburn, Ga," several people with the last name Arnold and "homes sold in shadow lake subdivision gwinnett county georgia."

[430] Ibid
[431] Schwartz (n.17)
[432] Ibid

It did not take much investigating to follow that data trail to Thelma Arnold, a 62-year old widow who lives in Lilburn, Ga., frequently researches her friends' medical ailments and loves her three dogs. "Those are my searches," she said, after a reporter read part of the list to her.'[433]

Meanwhile, another example involved the popular streaming website Netflix in a study by Arvind Narayanan and Vitaly Shmatikov[434]. At the time that Netflix was still a DVD rental business. Their research demonstrated that some people in a supposedly anonymised data set could be identified through their ratings on the website IMDb. The researchers demonstrated that "Given a user's public IMDb ratings, which the user posted voluntarily to selectively reveal some of his […] movie likes and dislikes, we discover all the ratings that he entered privately into the Netflix system, presumably expecting that they will remain private."[435]

According to a study by computer science professor Latanya Sweeney, all it takes is a ZIP code, birth date, and gender to identify 87% of Americans[436]. When considering how unexpected correlations can came out of unexpected places, the fact that there is no algorithm or program that can take into account every piece of data and what can be deduced from it, the idea that any data is truly "anonymous" becomes doubtful.

So as we have seen, anonymisation is a set of techniques aimed at ensuring that individuals cannot be identified, while nevertheless using the data to extract value, but it is not perfect and is getting weaker as re-identification technology improves.

2. The status of pseudonymised data

The limitations of anonymization are problematic because they dispel the conception that personal data and data value can co-exist without conflict. The goal of anonymisation was to create sets of data which could be left to be used and mined free of protective

[433] Barbaro, M. & Zeller, T. (2006), A Face Is Exposed for AOL Searcher No. 4417749, *N.Y. TIMES,* Aug. 9, 2006, at Al.
[434] Narayanan, A. & Shmatikov, V. (2008), Robust De-Anonymization of Large Sparse Datasets *(2008 IEEE Symp. on Sec. and Privacy 111, Feb. 5, 2008),* available at http://www. cs.utexas.edu/-shmat/shmatoak08netflix.pdf.
[435] Ibid
[436] Sweeney, L. (2000), Simple Demographics Often Identify People Uniquely, *Carnegie Mellon Univ., Sch. of Computer Sci., Data Privacy Lab., Working Paper No. 3, 2000.*

restrictions and limitations[437]. This idea of "release-and-forget anonymization", data which can be anonymized then released into the wild with no further oversight, cannot have been fully effective as data can often be re-identified, and there is no way to know with certainty whether re-identification is possible[438].

If anonymisation fails, then data sets used in certain sectors, including healthcare, cannot be used to create useful information, as the sensitivity of the data is too important to warrant the risk of re-identification. As long as data created by individuals is involved, identification will be a risk - the only real way to be perfectly sure the data will be protected is by not collecting it at all. This means that either an alternative to anonymisation needs to be found - another process which can protect data while allowing the data to be used for creating value - or a balance needs to be struck between the right to data protection and the responsibilities imposed on data controllers.

An alternative has been proposed, taking into account the fact that the main factor in re-identification is the combination of "anonymized" data with other data[439]. This alternative, "pseudonymised data", has been primarily proposed in fields where studying data is necessary, but where the data is also very sensitive - including, in particular, healthcare[440]. Pseudonymised data is data which, on its own, is anonymised (by removing personal identifiers) but which may be re-identified if it is linked to other pieces of data, kept separately. It is appropriate that this concept be prominent in the medical field, as it is akin to isolating a patient in a germ-free room: the data is safe and protected as long as it is not "infected" by "foreign agents" - other pieces of data[441]. It is defined by the GDPR's Recital 26 in these terms: "Personal data which has undergone pseudonymisation, which could be attributed to a natural person by the use of additional information, should be considered as information on an identifiable natural person."[442]

The ICO's "Code of Practice on Anonymisation"[443] makes a distinction between data aggregation exercises which result in non-individualized data (and as such anonymised under the DPD's definition) and processes which remove certain identifiers from person-

[437] Ohm (n.429)

[438] Stalla-Bourdillon, S., and Knight, A. (2016), Anonymous data v. Personal data–A false debate: An EU perspective on anonymisation, pseudonymisation and personal data. *Wis. Int'l LJ. APA*

[439] Ohm (n.429)

[440] Ibid

[441] Tsakalakis, N., Stalla-Bourdillon, S., & O'hara, K. (2016). What's in a name: the conflicting views of pseudonymisation under eIDAS and the General Data Protection Regulation.

[442] Recital 26, GDPR

[443] Information Commissioner's Office, Code of Practice on Anonymisation: Managing Data Protection Risk, (2012).

specific data but leave individual-level information (carrying higher risks)[444]. The later includes pseudonymised data, which is defined as "distinguishing individuals in a dataset by using a unique identifier which does not reveal their 'real world' identity."[445] What the ICO envisions as a means to turn this pseudonymised data into anonymised data is not specified.[446]

The General Data Protection Regulation gives some leeway to pseudonymised data, in order to incentivize the use of the practice. Nevertheless, as established, it is not enough to exempt the data processing operation from the GDPR: pseudonymised data is still often personal data[447].

This shows that the European regulators are aware of the evolution of the Data Protection landscape, and that the binary approach of the Data Protection Directive is gaining some shades of grey, emphasising the "means likely to be used" as a prominent criteria in what constitutes enough protection or not.

The limitation of pseudonymised data is that it only protects data as long as it stays isolated from "contamination" by other data. As such, pseudonymisation provides limited usefulness, especially in the wider commercial online context, where data is frequently sold and combined with other pieces of data without the inherent protective apparatus found in fields like healthcare.

In conclusion, anonymisation has not been successful in creating "privacy-free" data. Data can never be truly free of the possibility of endangering informational privacy, no matter what techniques are used, because unexpected Information can always result from it. Pseudonymisation actually only proves that further with its protection which, while promising, is threatened by contact with external data. As long as complete non-identification is the requirement, anonymisation will fail to achieve it.

3. The Risk-Based Approach to Identification: Anonymisation and Pseudonymisation

[444] Stalla-Bourdillon (n.438)
[445] Information Commissioner's Office, Code of Practice on Anonymisation: Managing Data Protection Risk, (2012).
[446] Stalla-Bourdillon (n.438)
[447] Recital 23, GDPR

We have previously identified that the GDPR is moving to a risk-based approach, protecting informational privacy through a contextual, safeguards-based approach. As we will now see, this risk-based approach extends to the ''identification'' criteria, with the chance of identification increasing the risk. This shows that even in the ''identifiability'' criteria, which is on the face of it a binary test (either data is personal or it is not) the GDPR recognizes that certain guarantees, such as tools to make identification more difficult, do have an impact on the protection the data needs to be given.

The Article 29 Working Party's opinion on anonymisation seems to make suggestions towards a risk-based approach to anonymization, such as with a statement that "legal regulations…must therefore be formulated in a technologically neutral manner and ideally take into account the changes in the developing potentials of information technology.". However, it requires an almost-null probability of re-identification for the criteria of anonymization to be fulfilled. In the Big Data era, this is nearly impossible. Additionally, only if the initial raw dataset is destroyed can the data be considered anonymized. As such, despite some indications that the risk-based approach would be used in the context of anonymization, the fact of the matter is that it seems rather limited at present.

Nevertheless, some provisions of the GDPR seem to go beyond the binary approach of "personal data" or "anonymised data"[448]. While acknowledging the binary approach, the GDPR allows for a wider spectrum of de-identification[449]. Besides the pseudonymisation mentioned above, further types of data were identified in the GDPR by scholar Mike Hintze[450], which include "identified data", which directly identifies a specific person (such as a name or e-mail address), and "identifiable data" which related to a specific person whose identity is not clear but there is a reliable way to create a link with identifying data (pseudonymous data is considered by Hintze as a subset of identifiable data).

Besides those two relatively general types of data, Article 11 of the GDPR[451] adds data which resembles identifiable data, but where there is no way for the data controller, using his available means, to actually identify the data - this includes data sets which used to be considered anonymous (because the data controller did not believe they could be identified) but ended up actually having some re-identification, such as cases with AOL search data or Netflix Prize data[452]. Finally, "anonymous/aggregate data" is data that

[448] Hintze (n.271)
[449] Ibid
[450] Ibid
[451] Article 11 GDPR
[452] Hintze (n.271)

cannot identify the individual - anonymisation must be irreversible and eliminate any potential future re-identification attempts. Of these, only anonymous data is exempt from the GDPR.

The idea that the efforts of the data controller with regard to the anonymisation of data has an impact over its legal framework leads to an interesting understanding of the GDPR's forays into a more flexible, context-specific understanding of data protection. This is not only found in Article 11; in Article 6(4) of the GDPR[453], mention is made of de-identification allowing the processing of data for other purposes, bypassing the purpose specification principle. Under that Article, "Where the processing for a purpose other than that for which the personal data have been collected is not based on the data subject's consent . . . the controller shall, in order to ascertain whether processing for another purpose is compatible with the purpose for which the personal data are initially collected, take into account, inter alia . . . (e) the existence of appropriate safeguards, which may include encryption or pseudonymisation."[454].

This is important because it means that there is an acknowledgment of encryption and pseudonymisation efforts being part of a spectrum of "safeguards" which improve the chances of a processing beyond the original purpose being considered compliant with the GDPR, beyond the binary approach. As we will see later on, this is a symptom of the larger trend surrounding the need for an overhaul of data protection which embraces fully a conception taking into account all relevant factors instead of a binary approach.

Besides anonymisation, a pillar of EU data protection law is using consent: getting the consent of individuals allows for the processing of their personal data legitimately. However, we will show that this pillar, too, is failing, and will only keep failing more over time.

II. The Role of Consent: A Control and Autonomy Tool Facing Challenges in the Big Data Era

Now that we have shown that anonymisation is failing in the world of Big Data due to the blurring of the line between "personal" and "non-personal" data, we will show that this

[453] Article 6 GDPR
[454] Article 6 GDPR

blurring, and this ever-increasing complexity, is endangering another pillar of data protection legislation: consent. We will show that in this the age of information, the intricacies of data processing have rendered the idea of consent as a means to ensure autonomy ineffective. Showing consent is failing will prove two things: first, that the idea of putting the main legislative barrier to data processing at the collection stage (where consent is obtained) is ineffective. Second, it will show that the widening expertise gap between individuals and data controllers in terms of data processing is eroding the very idea of autonomy as a pillar of the data protection framework. Thirdly, it will show that the ‘’control’’-based approach to protecting informational privacy faces difficult challenges in the age of Big Data. Control requires the ability to make significant, informed choices, which may not be possible as the technological literacy divide between data processing and the subjects of that processing is ever-widening.

1. Consent in the EU Data Protection Regulation

i. The role of consent in data protection and elsewhere

Consent has always played a central role in conceptions of informational privacy. According to J. Feinburg, “The root idea … of privacy is that of a privileged territory or domain in which an individual person has the exclusive authority of determining whether another may enter, and if so, when and for how long, and under what conditions. Within this area, the individual person is… boss, sovereign, owner.”[455] A report by privacy advocate Simon Davies compiling the opinions of over 180 privacy specialists from 19 countries listed consent as the third most influential privacy theme for 2013 – after data aggregation and regulatory changes[456].

This reliance on consent is based on the image of consent as the expression of an individual’s autonomy[457]. However, despite its influence in the European context, it is debatable whether “consent” should be a fundamental principle of data protection in its own right or whether – because it is so difficult to define and apply in practice – it should only play a supportive role to the package of other principles[458]. As consent is frequently set aside, and is difficult to obtain, it is open to question whether it should be the basis for an informational self-determination foundation of privacy.

---

[455] Feinberg, J. (1984). Offense to others (Vol. 2). *Oxford University Press on Demand.*

[456] Davies, S (2013), Predictions for Privacy A report on the issues and trends that will dominate the privacy landscape in 2013, *LSE Enterprise, the London School of Economics*

[457] Le Métayer, D. (2010). Privacy by design: a matter of choice. In *Data protection in a profiled world* (pp. 323-334). Springer, Dordrecht.

[458] Brownsword, R. (2004). The cult of consent: fixation and fallacy. *King's Law Journal*, *15*(2), 223-251.

As an action expressing the free choice of an individual, consent is linked with personal autonomy[459]. Rawls argues that a person is action autonomous when "the principle of his actions are chosen by him as the most adequate possible expression of his nature as a free and equal rational being."[460] Faden and Beauchamp take a similar interpretation of personal autonomy, with a more practical formulation: "the personal rule of the self by adequate understanding, while remaining free from controlling interferences by others and from personal limitations that prevent choice."[461]

Generally, consent is a form of autonomous action which is aimed at the authorization of another party's actions. The right for an individual to choose is reflected in their ability to consent. Because of how important consent is in expressing an individual's autonomy and ability to choose, those who consent need to do so through an adequately clear and affirmative action.[462] Because informational privacy means control over personal information, taking away that control is a breach of informational privacy, making an absence of consent, usually, a breach of informational privacy[463].

A survey published in June 2011 suggests that "nearly all Europeans would like to give their specific approval before collection and processing of their personal information"[464]. A survey by the European Commission in 2008 showed that 64% of EU citizens are concerned about the handling of the personal data gathered by data controllers and processors, and also showed that the vast majority of internet users feel insecure about the state of their personal data on the internet[465]. As such, not only is consent a way for individuals to express their autonomy, but it is also the mechanism individuals trust to protect their interests through the exercise of such autonomy.

---

[459] Beyleveld, D., & Brownsword, R. (2007). *Consent in the Law*. Hart Publishing, Oxford

[460] Schermer, B. W., Custers, B., & van der Hof, S. (2014). The crisis of consent: How stronger legal protection may lead to weaker consent in data protection. Ethics and Information Technology, 16(2), 171-182. Chicago

[461] Buccafurni, D. (2008) , Reconsidering the Facilitation of Autonomous Decisionmaking in Genetic Counseling, *ProQuest*, p.139

[462] Ibid

[463] Austin, L.(2005), Is Consent the Foundation of Fair Information Practices? Canada's Experience Under Pipeda*. 56 University of Toronto Law Journal 181,* p.5

[464]Special Eurobarometer 359 (2011), Attitudes towards data protection and electronic identity in the European Union. *Survey conducted by TNS Opinion & Social at the request of Directorate-General Justice, Information Society & Media and Joint Research Centre Survey*

[465] European Commission (2011), Flash Eurobarometer 225, Attitudes towards data protection and electronic identity in the European Union. *Survey conducted by the Gallup Organization Hungary upon the request of the Directorate General Justice, Freedom and Security*

Because consent has the ability to make an illegal act legal, the one who has control over the giving or not of consent has control over this ability, which gives them power. A constant issue in subject/controller relationships is the balance of power: for example, if one wants to join in on social media, one has to agree to the social media platform gathering and processing their data. In most fields, an individual consumer can only choose from what the market offers, and to ensure data protection, there needs to be a communal push for that protection. Where this is not the case, an individual will have little recourse where their preferences differ from the majority of people[466].

The Article 29 Working Party issued an opinion on consent in the employment law context. In this opinion[467], the Article 29 WP explains that consent under Article 7(1)(a) of the Data Protection Directive is not a commonly used ground for the processing of personal data in the employment context[468]. Instead, the other grounds listed in that Article are commonly used.

Because it is possible to use the grounds of Article 7(1)(b) and 7(1)(c) for most processing purposes (contractual and legal obligations), the Article 29 Working Party maintains that consent should only be used as a basis when the other options are not available. The reason behind the use of these grounds is that they are based not on autonomy of the data subject, but on necessity. Consent needs to be freely given under the definition of consent in Article 2(h), but this can only happen if this consent can be given without outside influence. This is problematic in employment law as the bargaining powers of the employer and employee are usually extremely uneven, with pressure to comply with the employer's wishes, especially if the giving of consent is a condition of employment. As such, using more definite grounds for processing, especially an obligation laid down in the law or in a contract, is more secure.

This is particularly interesting for our purposes: in the employment context, the systematic inapplicability of one of the conditions - "freely given" - challenges consent as a basis for processing. It is along this pattern that we will identify the greater limitations of consent. Now that we have gone over the basics of the notion of consent, we will show how the concept has evolved, particularly in the evolution from the Data Protection Directive to the GDPR.

[466] Austin (n.463)
[467] Article 29 Working Party (2014), Opinion 08/2001 on the processing of personal data in the employment context, WP48, Adopted on 13 September 2001
[468] Ibid

ii. Consent in the GDPR

In the GDPR, personal data may only be processed if the data subject has unambiguously given his consent, or if the processing is necessary for the performance of a contract to which the data subject is party, if the processing is necessary for compliance with legal obligations to which the controller is subject, if the processing is necessary for the performance of a task carried out in the public interest, or if the processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party[469]. This means that there are multiple grounds for the processing of personal data which can bypass the need for consent. Some even argue that most instances of processing will be able to be justified outside of consent[470]. This is important because, despite the importance of consent in the autonomy of individuals, it is only one amongst other ways of obtaining and processing data under the Data Protection Directive.

Consent in the GDPR is defined as "any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed"[471]. There are two distinct qualifications for consent in the Directive: for the processing of "special categories of personal data" - which include "personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life"[472] – for which the data controller needs to obtain "explicit consent".

In the GDPR, "unambiguous consent" is where there can be no doubt of the data subject's consent. Doubt is removed when consent is based on an express, positive action carried out by the individual[473].

Article 2(h) of the Data Protection Directive defines consent as "any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed". This highlights the three main requirements for consent in EU Law: "freely given", "specific", and "informed".

---

[469] GDPR, Article 4

[470] Bygrave, L., Data Protection Law. Approaching Its Rationale, Logic and Limits, *Kluwer Law International,* p.66

[471] GDPR, Article 4

[472] Ibid

[473] Ibid

"Freely given" means that consent can only be valid if the data subject is able to exercise a real choice, with no risk of deception, intimidation, coercion or significant negative consequences, if consent is not given. If the consequences of consenting undermine the freedom of choice of an individual, their autonomy would not be respected. An example is an employment relationship, where the data subject can be in a situation of dependence with the data controller. While subordination is often the main reason preventing consent from being free, other elements can influence the decision of the data subject. The processing being performed by a public authority, for example, can influence the data subject[474].

"Specific" means that the data processing operation needs to identify clearly its purpose and extent. Consent must be intelligible, and refer clearly and precisely to the scope and consequences of the data processing[475]. This means that the context in which consent applies is limited. This leads to there being a direct link between the "informed" requirement and the "specific" requirement.

The last element of the definition of consent - but not the last requirement for valid consent - is that it needs to be informed. According to the Article 29 Working Party's opinion on the Data Protection Directive, this means that "consent by the data subject (must be) based upon an appreciation and understanding of the facts and implications of an action. The individual concerned must be given, in a clear and understandable manner, accurate and full information of all relevant issues, in particular those specified in Articles 10 and 11 of the Directive, such as the nature of the data processed, the purposes of the processing, the recipients of possible transfers, and the rights of the data subject. This includes also an awareness of the consequences of not consenting to the processing in question".[476]

A new feature of the GDPR is that it introduces a brand new Article relating to consent, its Article 7 on "Conditions for consent"[477]. This provision adds a burden of proof on controllers obtaining consent , the requirement for consent to be clearly separated from the rest of the agreement , and the right of the data subject to withdraw their consent . Finally, according to Recital 43 of the GDPR, consent is not a legal basis for processing, where there is a significant imbalance between the data subject and the data

[474] Article 29 Working Party (2014), Opinion 08/2001 on the processing of personal data in the employment context, WP48, Adopted on 13 September 2001
[475] Ibid
[476] Article 29 Working Party (2007), Working Document on the processing of personal data relating to health in electronic health records (EHR), WP131, Adopted on 15 February 2007
[477] GDPR, Article 7

controller[478]. These conditions show that the General Data Protection Regulation's intention for consent is for the "free" and "informed" aspects of consent to be ensured through additional protections on top of the Directive's already-heavy level of protection. This shows that the GDPR is aware that "imbalances" are a possibility. What we will argue however, is that these "imbalances" are becoming so widespread as to make every situation one where consent becomes a non-valid basis for data processing. When this becomes the case, paternalistic intervention becomes necessary.

The challenge comes from a dissonance between what legal provisions are considered to lead to the best protections to individuals in principle, and the actual technological and practical realities. The multiplicity of sources for information gathering, processed by layers upon layers of service providers, means the imposition of explicit consent requirements at certain control points in the flow of information could frustrate individual user experience without increasing their privacy. A common worry of scholars is that privacy becomes a "box-ticking exercise", where data controllers and legal authorities would only care for the legal requirements being filled, regardless of actual outcome[479].

iii. Consent in the EU's vision of ''Informational Privacy as Control'' approach

We have seen in previous chapters how the EU is moving towards a new approach, one based not on the ''personal data'' binary but on ensuring that appropriate safeguards are in place to limit the risk involved in the data controller's activities. Those safeguards are means to prevent negative impacts ''that may result from a data subject losing control over personal information, or realising that it has been compromised.''[480] This idea of ''control'' as a core way to protect informational privacy takes inspiration from the German vision of privacy as a tool of self-determination: in a 1983 decision, the German Federal Constitutional Court ruled that the right to informational self-determination included ''basic right warrants (...) the capacity of the individual to determine in principle the disclosure and use of his/her personal data''[481].

This idea of ''control'' is a core focus of the GDPR, with rights such as the right to data portability being created ''[t]o further strengthen the [data subject's] control over his or

[478] GDPR, Recital 43
[479] Zanfir, G. (2014). Forgetting about consent. Why the focus should be on "suitable safeguards" in data protection law. In Reloading Data Protection (pp. 237-257). *Springer Netherlands.*, p.5
[480] Article 29 Data Protection Working Party (2013), Opinion 03/2013 on purpose limitation, WP203, Adopted on 2 April 2013
[481] BVerfGE 65, 1 – Volksz¨ahlung Urteil des Ersten Senats vom 15. Dezember 1983 auf diem¨undliche Verhandlung vom 18. und 19. Oktober 1983

her own data''[482]. Consent is the core mechanism which the GDPR relies on to ensure control over personal data, as it is a legal basis which is entirely in the hands of the data subject. When a higher level of consent is required, such as in the case of ''explicit consent'' required in case special categories of personal data are being processed[483], it is because that processing ''poses significant data protection risks and a high level of individual control over personal data is therefore deemed appropriate.''[484]. This shows that not only is consent there to provide control to the data subject, but that the amount of control given to the data subject needs to be proportional to the risk involved in the data gathering. Research into consent reinforces this view as control being an important element of informational privacy: Stewart and Segars found that ''consumers are less likely to view a given information practice as privacy invasive if they are able to maintain, even to a small degree, some measure of control over their personal information''[485]. Examples of ''small degree on control'' include asking for permission to use data for new purposes or providing access to what data is processed.

One main aim of the GDPR is to ensure that the individual can "control" personal data. Control is a core aim of the GDPR (as can be seen by the strong rights given to the data subject aimed at giving control to the data subject[486]), and is used as a guarantee to protect data subjects. The strongest evidence of the role of control as a guarantee can be seen by the fact that one of the core risks to data subjects named in the GDPR is the loss of control of personal data: Recitals 75 and 85 mention the inability to exercise control over personal data as a type of risk similar to "discrimination, identity theft or fraud, financial loss, damage to the reputation"[487] amongst others.

This idea of "privacy as control" has been part of the privacy debate for a long time[488], especially in the US[489]. The first to mention this idea in the EU were cases in Germany

[482] GDPR, Recital 68
[483] GDPR, Article 22
[484] Article 29 Working Party (2017), Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, WP251, Adopted on 3 October 2017
[485] Stewart KA and Segars AH (2002) An empirical examination of the concern for information privacy instrument. *Information Systems Research* 13(1), 36-49.
[486] See Recital 68 of the GDPR on the right to data portability: "To further strengthen the control over his or her own data, where the processing of personal data is carried out by automated means, the data subject should also be allowed to receive personal data concerning him or her which he or she has provided to a controller in a structured, commonly used, machine-readable and interoperable format, and to transmit it to another controller."
[487] GDPR, Recitals 75 and 85
[488] Camenisch, J., Sommer, D., Fischer-Hübner, S., Hansen, M., Krasemann, H., Lacoste, G., ... & Tseng, J. (2005, November). Privacy and identity management for everyone. In *Proceedings of the 2005 workshop on Digital identity management* (pp. 20-27). ACM.
[489] Allen, A. L. (1999). Privacy-as-data control: Conceptual, practical, and moral limits of the paradigm. *Conn. L. Rev.*, *32*, 861.

related to “informational self-determination”, which later inspired the Data Protection Directive[490].

As we have seen throughout the discussion on consent, “control” is also a way to justify data processing and ensure that sufficient guarantees are in place: if the data subject has complete control over whether the processing takes place or not, there is a strong guarantee that nothing will take place that will breach their informational privacy, since they should be fully aware (through the privacy notice) of the Information being gathered and the Guarantees in place, and thus informed judges on whether they thought the Balance adequate. Ideally, consent would thus prove complete control over data, allowing each individual to make up their own mind up over where they judge the Balance adequate to their sensibilities.

An emphasis on control can be found throughout various parts of the GDPR: beyond the consent mechanism, the rights of the data subject strengthened in the GDPR - such as the right of access (now to be provided free of charge within a month)[491], the strengthened right to object[492], and a renewed duty to provide information in a way that is clear and transparent[493] - all show the focus on giving tools to data subjects to take control of their data.

These tools are aimed at preventing harm by shifting the balance of power towards the data subject through a set of control-based guarantees. The focus on control in the GDPR shows that the EU regulator intends to use control as a core element of this guarantees-based approach.

The idea of individual control over data as a central guarantee is not just found in the GDPR itself. Attempting to rectify the balance by empowering the data subject has been suggested in the literature, usually using technological tools[494]. A section of tools aimed at achieving this has been dubbed “Privacy Enhancing Technologies” (PETs)[495]. Some tools create “noise” to limit the accuracy and thus the value of data obtained about individuals (an example is TrackMeNot, a tool developed by Helen Nissenbaum which

[490] Lynskey (n.127)
[491] GDPR, Article 15
[492] GDPR, Article 21
[493] GDPR, Articles 14 and 15
[494] Quelle, C. (2016). Not just user control in the general data protection regulation. In *IFIP International Summer School on Privacy and Identity Management* (pp. 140-163). Springer,
[495] Hansen, M., Berlich, P., Camenisch, J., Clauß, S., Pfitzmann, A., & Waidner, M. (2004). Privacy-enhancing identity management. *Information security technical report*, *9*(1), 35-44.

performs random searches to muddle the profile created by the user's activity[496]). However, as we will show through the example of consent, ensuring control and autonomy over data is becoming more and more difficult.

As we have shown, ''risk'' is a major focus of the GDPR's new approach to data protection. A major component of ''risk'' is whether ''control'' is taken away from individuals. This link between ''risk'' and ''control'' places control at the centre of this new approach: under the GDPR, informational privacy is at its core about protecting individuals against misuses of their data, and giving them control over it. As we will show, though this new, ''risk-based'' approach shows a move towards a framework based on what Information is available and the measures taken to mitigate risks stemming from the use of that Information (what we call ''Guarantees''); and under the framework developed for this work of a balance between Information and Guarantees, ''control'' is simply one type of Guarantee and not an end in and of itself. However, the GDPR linking risk, and thus informational privacy, to control and consent as a core goal brings a limitation: giving data subjects control is very difficult, and as such it may not always be possible to ensure effective data subject control over data.

2. The Limitations of ''Informed'' Consent

i. The Importance of Informing Data Subjects

The central basis of giving power and control to the individual to manage their privacy is that it requires that the individual be informed and rational and be able to consent to the different forms of collection and processing of personal data[497]. However, it is a fact that most individuals do not use the options that exist to protect their own privacy. Many reasons have been given for this, including the argument that the need to read through long texts discourages users[498].

A difficulty has been that finding effective ways to communicate information to individuals is difficult, and individuals do not necessarily understand the information given to them[499].

[496] See "Peddinti, S. T., & Saxena, N. (2010, July). On the privacy of web search based on query obfuscation: a case study of TrackMeNot. In *International Symposium on Privacy Enhancing Technologies Symposium* (pp. 19-37). Springer, Berlin, Heidelberg" for a study of the tool, or consult https://cs.nyu.edu/trackmenot/ for information about the tool itself.

[497] Berbers, Y., Hildebrandt, M., & Vandewalle, J. (2018). Privacy in an age of the internet, social networks and Big Data. *Position paper 49b, Royal Flemish Academy of Belgium for Science and the Arts*.

[498] Ibid

[499] Ibid

What's more, the problem is not necessarily that doctors - or in the data protection context data controllers - do not try to enlighten individuals, but instead it seems that even honest attempts at information have been unsuccessful. Both attempts by individual entities and legislative efforts for informed consent seems to have failed to produce a significant success[500].

An example of this phenomenon is the "informed consent" required in medicine for certain actions by the physician. A study showed that only 9% of doctors met the criteria for informed decision-making, which includes informing the patient of his role in the decision-making process, the nature of the decision, possible alternatives, pros and cons, uncertainties associated with the decision, an assessment of the patient's understanding of the decision, and an exploration of the patient's preferences[501].

Under traditional individual theory, individuals are forward looking, utility maximizing beings who are fully informed and base their decisions on probabilities coming from known random distributions[502]. This idea is common in the policy debate, where it is assumed that individuals and organizations only need the ability to manage their information and make choices without regulatory intervention. However, several studies have shown that there is an apparent dichotomy between privacy attitudes and actual behavior. Individuals exchange large amounts of information for little reward, and are rarely willing to adopt privacy protective technologies[503].

Many factors hamper individual decision-making. Incomplete information, and information asymmetries between a well-informed data controller and a non-informed individual make privacy decisions unlikely to be accurate. What's more, individuals cannot properly evaluate possible externalities (third party transfers); risk (privacy risks are not easy to estimate), and uncertainties about possible payoffs from their information all factor into making calculating the costs and benefits of privacy-intrusive actions difficult[504]. Worsening the effect is the fact that privacy does not have a quantifiable monetary cost.[505] Even if individuals did have access to complete information, they could not process and act on the data optimally.

---

[500] Ben-Shahar, O., & Schneider, C. E. (2011). The failure of mandated disclosure. *University of Pennsylvania Law Review, 647-749.*
[501] Ibid
[502] Acquisti, A., & Grossklags, J. (2005). Privacy and rationality in individual decision making. *IEEE Security & Privacy, 3(1),* 26-33.
[503] Ibid
[504] Ibid
[505] Ibid

Even if individuals had access to all the information, rationality has its limits. Many deviations from rationality affect decision-making, in particular self-control problems and the many social norms at play in the making of those decisions. Finally, the ignorance of individuals of the dangers that come from privacy invasions is a factor[506].

This would not be an issue if the individual could accurately assess harm, cost, and the magnitude and probability of the loss. However this is not the case. Consumers lack the information necessary to make this decision, and are in a disadvantageous position to decide when and how to protect themselves from the harms inherent in behavioral targeting[507].

Consumers also lack information about what information can be obtained about them by data controllers. Their inability to assess the magnitude of loss and understanding the possibilities of data collection available to a data controller suggests that consumers would be more upset if they understood all those factors - such as phone calls being monitored.

ii. The Paradox of Informing Users

It would be easy to conclude that data controllers are intentionally keeping users uninformed for their own benefit. However, as we will, show, this increasing gap, the same gap blurring the concept of "personal data", is present even where efforts are made to keep users up to date. The main paradox is called the "transparency paradox", where the more information is shared, the less understandable this mass of information is[508]. Because of the complexity of online tracking and surveillance, it is practically impossible for privacy policies shared by data controllers to be both accurate, and understandable by every user[509].

The consequence of this overload of information, intrusive pop-ups screens and lack of choice is "consent desensitisation": users are no longer used to making active, informed

[506] Ibid
[507] Politou, E., Alepis, E., & Patsakis, C. (2018). Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. *Journal of Cybersecurity*, *4*(1)
[508] Schermer et al. (n.460)
[509] Van Alsenoy, B., Kosta, E., & Dumortier, J. (2014). Privacy notices versus informational self-determination: Minding the gap. *International Review of Law, Computers & Technology*, *28*(2), 185-203.

choices when confronted with a consent situation[510]. This will only be exacerbated with the added requirements from the General Data Protection Regulation.

Another difficulty comes from the scale of Big Data - the number of entities collecting and sharing data for individuals to keep track of. A survey has shown that the average US citizen, for example, visits almost a hundred websites a month, doing business online and offline with countless companies, each of which has the potential to hold, use, transfer, or sell personal data[511]. This extends to data that individuals might not even be aware exists, such as metadata for phone calls or websites tracking every click and second spend on their site by each individual[512].

This means that even if the option existed to protect one's privacy, and even if every company provided informed and reasonable choices to manage and protect one's data, this challenge would persist: individuals do not have the time or energy or willpower to manage all of the entities holding their data[513].

As an example of the difficulty individuals face in using their ability to consent, one study explained that it would cost 781 billion dollars in lost productivity if everyone were to read every privacy policy they visited in a one-year period[514].

This is the main issue with the idea of consent as an expression of one's true intentions: attempts to inform the public and lead to pertinent decision-making are not effective. Individuals do not read the information provided to them. If they read it they do not understand it. If they read it and understand it, they do not have the knowledge to make an informed choice[515]. If they read it, understand it, and have the knowledge to make an informed choice, that choice is very likely skewed by obstacles to decision-making.

In conclusion, the GDPR creates stronger consent requirements, demanding individuals be more and more informed in order for consent to be given. However, it is fast becoming impossible for individuals to understand the issues they are faced with because of the

---

[510] Bergemann, B. (2017). The Consent Paradox: Accounting for the Prominent Role of Consent in Data Protection. In *IFIP International Summer School on Privacy and Identity Management* (pp. 111-131). Springer

[511] Ibid

[512] Ibid

[513] Solove, D. J. (2012). Introduction: Privacy self-management and the consent dilemma*. Harv. L. Rev., 126, 1880.*

[514] McDonald, A. M., & Cranor, L. F. (2008). The cost of reading privacy policies. *ISJLP*, *4*, 543

[515] Custers, B., van Der Hof, S., Schermer, B., Appleby-Arnold, S., & Brockdorff, N. (2013). Informed Consent in Social Media Use-The Gap between User Expectations and EU Personal Data Protection law. *SCRIPTed*, *10*, 435.

complexity both in the processing of the collected data, and in what information may be created by data processing later on (too complex and too unpredictable). This shows at the very least that consent as an instrument fails to ensure the protection of autonomy, but we would go further and argue that because of the ever-widening gap between individual knowledge and the complexity of data processing, free autonomy is going to become increasingly challenging to prove. This also challenges the conception of informational privacy as ‘’control’’, since achieving such control is difficult, or even impossible.

We argue that any attempt to move European data protection towards adapting to the Big Data age will mean a move towards top-down regulation, instead of a focus on fixing consent and reinforcing the control of individuals. Though such efforts are important and have a place (as we will show later on) they are likely to fail if data processing practices remain the same as they are now. As we have shown in our third Chapter, data processing is increasingly elaborate and complex, worsening the existing limitations developed in this section.

In this Chapter so far, we have gone over how the technical challenges involved in the rise of Big Data technology have affected two pillars of EU data protection regulation: anonymisation (and with it the ‘’personal/non-personal data’’ paradigm) and consent (and with it ‘’control’’ as a focus of data protection). We have shown that these technologies provide tremendous power to parties which can use it to their benefit, and that this power involves threatening informational privacy by giving more power to data controllers without giving sufficient Guarantees to data subjects. Meanwhile, “personal data” is hard to define, and limitations in fields such as anonymisation and consent both in defining personal data and handling the changes that come from its further processing will get more prominent as dat processing continues to increase in complexity.

We have now identified some key challenges in European data protection regulation. We will now synthesize and analyze the deeper issues which have led to this problem, in order to understand the core reason behind these limitations. We will also study how the European framework of privacy is also showing signs of being affected by the changes Big Data has brought but has handled them in an approach which supports the ‘’Information/Guarantees Balance’’ approach.

B. Challenges in Informational Privacy Protection: Transparency and Obscurity

Over the course of this thesis, we have shown that the appearance of Big Data has changed what we understand about traditional conceptions of how data is created and spread, while core pillars of EU Data Protection - consent (and with it the ‘’control’’ focus of data protection) and anonymisation (and with it the ‘’personal data’’ focus of data protection) - are facing important challenges. We will now analyse what we argue to be the core of this limitation: “Identifiability” and the binary vision of data it creates, as well as the ‘’public/private distinction’’ of privacy, and its similarly-binary vision. However, as we will see, unlike data protection law, privacy law is showing signs of moving beyond this binary vision and towards adapting to the age of Big Data. As we will show, this change is slower in the context of data protection, because of the importance of the ‘’identifiability’’ test.

I. The Limits of “Identifiability”: Data and Information

1. The ‘’Identifiability’’ Test

In the GDPR, “personal data” is defined as “information relating to an identified or identifiable person”[516]. An “identifiable person” is one who can “be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity”[517].

Guidance on “identifiability” can be found in the UK Information Commissioner’s Office, according to which, “identifiability” is the first criteria of determining whether data is “personal”[518]. If it is not fulfilled, then the data is not personal data for the purposes of the UK’s Data Protection Act.

[516] Directive 95/46/EC, OJ L 281 of 23.11.1995, Article 2(a)
[517] Article 29 Data Protection Working Party, Opinion 4/2007 on the concept of personal data, WP136
[518] Information Commissioner’s Office (2012), Determining what is personal data, accessed at https://ico.org.uk/media/for-organisations/documents/1554/determining-what-is-personal-data.pdf on 13/6/2018

The ICO defined "identified" by whether "you have distinguished that individual from other members of a group.[519]" For guidance, the ICO mentions that a name, if not unique, might not be enough and might need to be aggregated with more data (such as an address) to satisfy the "identification" threshold. According to the ECJ, the name of a person combined with "his telephone number or information about his working conditions or hobbies" will satisfy the "identifiable" requirement[520].

In Recital 26 of the GDPR, guidance was given on how to determine wheter a person is "identifiable": "to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person"[521].

That wording, "means likely reasonably to be used", is where the "identified" requirement sees its biggest limitation. As we will see, and as we saw when discussing anonymisation, powerful means are becoming more and more widespread, rapidly reaching the point where it is reasonable for a data controller to use tools which might identify individuals from very little data. At that point, every data gathering could potentially be fully protected, and the link between data protection and privacy protection would be weakened. Whether or not it is possible for data to lead to personal information does not necessarily mean it will. This difference is an issue in the current regulatory framework.

The "identifiable" requirement was indeed useful when the information creation process was difficult and limited and it took extensive means to obtain relevant information, but this is becoming less and less true as Big Data becomes a part of day-to-day life.

The European data protection framework comes into effect when the data is "identifiable". The limitation is that though data which does not fall under this requirement is considered "non-personal", it can be used to create new information, which itself might have a "personal" quality. An example is in the practice of profiling. As discussed in the previous chapter, through aggregating data on an individual, each piece being either not personal or having been given with consent or another legal ground, a data controller can piece together a much more revealing picture than what each single data point might.

[519] Ibid
[520] Bodil Lindqvist v Aklagarkammaren i Jonkoping – Case Commissioner - 101/01 – European Court ofJustice delivered on 19 September 2002
[521] GDPR, Recital 26

At that point, this “new information” has already passed the tests under the European Data Protection system - aggregation of data invalidating the previous legal bases for processing would be impractical, and more importantly unenforceable. As such, this “raw data”, despite not being personal, is free to be used to create personal “information”.

At the same time, even when personal information is gathered, the danger it represents to informational privacy can be slight or nonexistent. An example is a website monitoring customers’ debit card purchases to find out which products are selling the most: the “personal” part of that data is trimmed out of the analysis as irrelevant, and the information created at the end of the process has no impact on privacy.

Finally, data gathering can impact privacy even with no identification of individuals. The best example is the practice of group profiling: information about groups (based on area, post code, population, etc) which identifies no one individual inside that group. In that situation, no one is “identified”. Nevertheless, assumptions about individuals built on that data can impact their informational privacy[522]. In 2012, a Wall Street Journal report found that many major companies would use information on the physical locations of their users to display different online prices to different consumers, with potential implications for discrimination and detrimental impact on consumer welfare. In particular, better deals were offered to high-income locations[523]. Yet group profiling is not effectively addressed by the GDPR[524] because there is no “identification”.

There are many circumstances where there is a very tenuous link between “personal data” and a practice breaching informational privacy because of this recent transformation, which has been brought about by the advent of new technologies. These processes are continuously improving and expanding, further eroding the link between the two concepts. We need to change our perception of “personal data” and “personal information”: they are two different (though related) concepts, and need to be considered as such.

2. The Contextual Approach to ‘’Identifiability’’

[522] Wachter, S. (2018). Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR. *Computer Law & Security Review*, *34*(3), 436-449.
[523] Valentino-Devries, J., Singer-Vine, J. and Soltani, A., Websites Vary Prices, Deals Based on Users' Information, *Washington Street Journal,* 24 December 2012, available at: https://www.wsj.com/articles/SB10001424127887323777204578189391813881534
[524] Hildebrandt, M., & Gutwirth, S. (2008). *Profiling the European citizen* (pp. 17-45). Dordrecht: Springer.

In order to understand “identifiability”, it is important to outline the types of data that are recognized by the GDPR and how they relate to this requirement. We have studied the main types of data understood by the GDPR: “identified data” which directly identifies an individual, “identifiable data” where the individual’s identity is not clear but a link to it can be reliably found (such as pseudonymous data) - Article 11 de-identifies data which could identify the individual but where the data controller has no means to do so - and anonymous data where the individual cannot be identified. Under the GDPR, only that last type is immune from the Regulation.

A difficulty in “identifiability” is in what information can identify an individual or not. Examples have been given at various stages by regulatory or advisory authorities, but no clear rule has appeared to decide what constitutes “identifiable” or “identified” data. Instead, we can see that this distinction is highly contextual. Data that are not identifiable for one person may be identifiable for another, and some individuals hold some data as having a more sensitive status than others.

An additional challenge is that the status of data changes based on the other data it is linked with: the aggregation of two ‘’non-personal’’ pieces of information can create identifiable information in many, often unforeseen, cases. Because of how much data can be collected in the age of Big Data, knowing what the context actually consists of - who knows what and what does it mean - can be very difficult. We will show this by going through the various steps of data processing that we have previously identified (collection, aggregation, storage, processing, mining and usage) to show this development and the problem this poses with regards to the GDPR.

In the collection stage identified and identifiable data can be gathered, and will need a legal basis for collection and processing. Any such data can be de-identified (using anonymization, pseudonymization, or other techniques). If that data is de-identified sufficiently to be considered “anonymised data” (or is not linked to individuals whatsoever, such as weather sensor data), it can then be sold and transferred to any third party.

However, at the aggregation stage, the combination of these data points can create new information, even if the data points involved are completely non-personal data. Weather patterns in an area can be aggregated with certain risks that allow identification of populations which are more likely to suffer a flood - which can affect insurance

premiums[525]. This can happen accidentally as well - as we have mentioned when discussing anonymisation, not only is it possible to re-identify data, but it can happen without even looking for it when the right patterns emerge.

The data is then curated through the storage and processing phases into a form that will be useful for the model that is being used. That is important, because all of the personal data collected is not necessarily used in the model. If personal data is not useful to the exercise the data controller is performing, then the curated data will not contain that data. As such, the Article 29 Working Party has made a distinction on these factors[526]: depending on the purpose of the processing, data can be considered as personal data or not. According to the Working Party, if the goal is to identify individuals and treat them a certain way, data should be considered identifiable because the controller will make a reasonable effort to identify individuals[527].

This is not the only time EU authorities have shown themselves to be amenable to a more flexible approach to identifiability. Advocate General Campos Sánchez-Bordona in the Breyer case hints at this possibility by asserting that in order to decide whether certain IP addresses are personal data, the context is crucial[528]. In his opinion in the case, at paragraph 68, the Advocate General argues that when it comes to "third parties" being considered as capable of re-identifying data for the purposes of the data controller's "means likely to be used" test, that "This will not occur when contact with those third parties is, in fact, very costly in human and economic terms, or practically impossible or prohibited by law. Otherwise, as noted earlier, it would be virtually impossible to discriminate between the various means, since it would always be possible to imagine the hypothetical contingency of a third party who, no matter how inaccessible to the provider of services on the Internet, could — now or in the future — have additional relevant data to assist in the identification of a user."

Additionally, the Court of Justice of the European Union expressly refers to the paragraph of the Advocate General's Opinion and excludes identifiability "if the identification of the data subject was prohibited by law or practically impossible on account of the fact that it

[525] Borgesius, F. Z., Gray, J., & Eechoud, M. V. (2015). Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework. *Berkeley Tech. LJ*, *30*, 2073.
[526] Article 29 Working Party (2007), Opinion 4/2007 on the Concept of Personal Data, WP136, Adopted on 20 June 2007
[527] Ibid
[528] Case C-582/14, Breyer v Bundesrepublik Deutschland, (2016) ECLI:EU:C:2016:779

requires a disproportionate effort in terms of time, cost and man-power, so that the risk of identification appears in reality to be insignificant."[529]

3. The "Relating to" Element of Personal Data

The last aspect we will mention when it comes to the wide scope of "personal data" and "identification" is that beyond the fact that much more data than before can be considered to "identify" an individual, similarly more circumstances than ever before can allow data to be "related to" the identified individual. What is "related to" the individual has been developed, to some extent, by the Article 29 Working Party in its Opinion 4/2007 on the concept of personal data[530], explaining that data is "related to" an individual when it is "about that individual" - which remains relatively unclear. It then lists some situations where it does apply: one's individual file in a HR department, or an "image of a person filmed on a video interview". The elements taken into account are three-fold - a "content" element, a "purpose" element, and an "effect" element. In short, data can be related to a person if the content is about a person, if the data is likely to be used to evaluate or treat the individual in a certain way, or influence their status or behavior, or if the data is likely to have an impact on the individual[531]. Emphasis is to be placed on the "or" in this set of criteria. Any of them is enough to qualify data as "relating to" the individual.

This seems to open the way for a very broad definition. And indeed, a recent decision showed this trend continuing. In the Nowak v Data Protection Commissioner case[532], a trainee accountant who had failed an exam four times wanted to have access to all personal data held about them by the exam authority, eventually making his way to the Court of Justice of the European Union. According to the Court, the purpose of an exam is to "identify and record the performance of a particular individual". As such, an examiner's comments incorporate some element of personal data about a candidate. That would be the case even if the candidate's name were not on the script but s/he were instead identified by a unique reference number or barcode[533].

[529] Stalla-Bourdillon (n.438)
[530] Article 29 Working Party (2007), Opinion 4/2007 on the Concept of Personal Data, WP136, Adopted on 20 June 2007
[531] Ibid
[532] Judgment of 20 December 2017, *Nowak v. Data Protection Commissioner*, C-434/16, ECLI:EU:C:2017:994
[533] Article 29 Working Party (2007), Opinion 4/2007 on the Concept of Personal Data

This expansion not only shows how far the reach of the notion of what constitutes ‘’personal data’’ has gone, but also shows how much this will continue to develop over time. Legal decision after legal decision, recommendation after recommendation, a point will be reached where many usual business activities will have the potential to collect all sorts of sensitive information because of the creep of the definition.

As we can see overall, what is “personal data” is far from a clear distinction. The age of Big Data brought more and more Information, which is more susceptible to context and biases. Because context changes over the course of the life of data (for example, data becomes processed by different entities for different purposes using different curated data), whether it is “identifiable” is similarly changing. Because it is possible to re-identify data in the later stages of processing, a data protection regulatory system focused on the collection stage (especially for the consent test) might be unable to adapt to the rise of Information: it is necessary to take into account the data at every stage and the surrounding context. Because the nature of the data can be so changing as to become unrecognizable, such an analysis needs to look beyond what can be directly linked to an individual.

This is true of all personal data, and not just “identified” or “identifiable” data. As we have seen when discussing anonymisation, all data, even anonymised data, has the potential to, depending on context, to link back to individuals and endanger informational privacy. Because the lines between the types of data are very blurry, making sure that anonymised data are also considered in the equation is necessary.

This phenomenon explains why anonymisation is failing: identifiability is not binary. It is also why consent is failing: the data involved is often collected without actually being identified, which means it bypasses the consent requirement.

Now that we have identified this limitation of the EU data protection regulatory framework, we will show the other major way in which the changing technological environment cannot be dealt with by the current European data protection conception: the public/private distinction, and how it affects data protection.

II. The Limitations of a Public/Private Distinction: Privacy's Binary Test

Introduction: When Private Goes Public

Informational privacy can be endangered in a number of ways. As we will show, while informational privacy was originally relatively well divided into a public and private sphere, this distinction is increasingly blurred.

The first example we will use is a story that made the news in 2005, the story of "dog poop girl"[534]. A young woman in South Korea was filmed not picking up after her dog defecated on a train, and the footage went viral online. Concerns for the woman's privacy quickly became secondary to her "crime", and very quickly her personal information was spread online and she was subjected to threats and harassment[535]. Acts committed in public view are held as public, and so there would seem to be no privacy issues in the footage being released online. Nevertheless, this led to her informational privacy being invaded. This shows both the fact that "public" can have a variety of meanings, and that it is not just from companies and governments that privacy invasions can come. They can come from random strangers on the Internet, and be the most damaging form of informational privacy invasion - in this case even including death threats.

Privacy concerns of individuals are generally considered in terms of their "private life", while public surveillance tends to be less noticed. This is due to a variety of reasons. One is that the main way privacy is understood publicly is as a "private sphere" that one should be able to have control over, and when that sphere is breached we feel a very strong and sudden aversion - that is why stories of police breaking into homes reach us more than police surveillance of CCTV footage[536]. The idea of a distinction between the "private" and "public" sphere is prevalent, and has a long history, such as a perception that the "private" sphere is the one of family and intimacy while the "public" sphere is the

[534] Solove, D. (2007). The future of reputation: Gossip, rumor, and privacy on the Internet. *Yale University Press.* - Part 1
[535] Ibid
[536] Nissenbaum, H. (1998). Protecting privacy in an information age: The problem of privacy in public. *Law and philosophy*, *17*(5), 559-596.

realm of work, citizenship, and community[537]. This division is even essential to the legal system itself, which distinguishes between "public" and "private" law. By definition, "privacy" would seem to be applicable to the "private" sphere only.

A link however can be made to privacy in public thanks to the American perspective of privacy. As we have mentioned earlier in this thesis, the American perspective is based on the opposition between the state and the individual, with the sanctity of the "home" as a central concept. However, because government interference in individual privacy can be found in places other than the home, jurisprudence has been developed that surrounds where individuals should or should not have privacy. This doctrine is known as the "Reasonable Expectation of Privacy", and has had many interesting developments in its "native" land of America. We will study this doctrine, and why its developments prove that not only is privacy in public a valid legal concept, it is one that has been - albeit implicitly - adopted by regulators.

This will inform our developments over ''informational privacy'' by showing that protecting informational privacy can be achieved, and has been achieved in the US with some success, through a balance between the information which can be obtained about individuals, and the guarantees protecting that information from being collected or used wrongfully.

1. Reasonable Expectations of Privacy in the US: inspiration for the transparency/obscurity distinction.

Before we study the American test of a "reasonable expectation of privacy", we need to establish the fundamentals of that test.

There is always a balance to be kept between different rights and obligations. In the case of privacy in the US, this right primarily conflicts with the right to freedom of expression and the right to freedom of the press when it comes to private actors, and conflicts with government intrusion when it comes to state actors[538]. Because any contact with another person, with the outside world, will expose some information about oneself to the world, there needs to be a limit to how much of one's life can be recorded and exposed. That

[537] Ibid

[538] Gorman C. (2011), Is Society More Reasonable than You?The Reasonable Expectation of Privacy as a Criterion for Privacy Protection *LLM Law & Technology Masters Thesis, Tilburg University*, p.7

limit is the “Reasonable Expectation of Privacy”, and this balance approach, especially holding freedom of expression so high, is characteristically American.

The strongest provision in American privacy law is the Fourth Amendment, however it is limited in scope, and only protects someone against government searches which violate a “reasonable expectation of privacy”[539]. Under the Fourth Amendment, such an expectation of privacy requires both “objective” and “subjective” expectation of privacy: one needs to actually expect privacy, and society as a whole needs to deem that expectation legitimate. This rule comes from a United States Supreme Court 1967 decision, Katz v United States[540]. This case holds that entering a telephone booth, closing the door, and making a call, entitles a person to privacy in the sense that the government cannot record what that person said without a warrant. This was because Katz considered himself to be under a certain expectation of privacy, and the idea is usually expressed in the form of “the Fourth Amendment protects people, not places”[541].

Under the American conception, the important distinction on whether privacy was breached, due to their torts-based view of privacy, relies on whether the information was “knowingly exposed” to a third party[542]. This is important because unlike the European conception which usually relies on a distinction between “private” and “public” information, the US approach is based on the context of the exposition of the information, and not the nature of the information itself. As such, the question of the reasonable expectation of privacy then becomes what is considered “reasonable”, and what constitutes “knowingly exposed”. Some Supreme Court cases seem to have gone against individual rights since Katz and toward a more government-friendly approach[543].

From the case law, two main approaches can be used which establish what is “reasonable” and “knowingly exposed”. In the first, the criterion of “knowingly exposed” is used on the basis of an assumption of risk - with a person not being able to maintain their expectation of privacy where some information will be revealed to the public. This approach has been criticized because it allows Courts to restrict privacy whenever a third party is involved[544]. An example is in telephone calls, where one is considered to have “assumed the risk” of publicly exposing data and that the metadata - who made the call,

[539] Fourth Amendment, Constitution of the United States
[540] Katz v. United States, 389 U.S. 347 (1967)
[541] Ibid
[542] Wilkins, R. G. (1987). Defining the reasonable expectation of privacy: an emerging tripartite analysis. *Vand. L. Rev.*, *40*, 1077.
[543] Gorman (n.538)
[544] Ibid

from where - was going to be shared from the third party to the government[545]. This means that whenever a third party is involved, privacy seems to end. On the other hand, in the European system, this is exactly where privacy begins, considering the high level of obligations put onto data controllers and processors.

The other approach is a balancing test between "the need to search" against "the invasion which the search entails"[546]. This balancing act is noteworthy because unlike the first approach, there is not a distinct point where privacy ends - instead, the importance of the right to privacy is included in the analysis. This approach was originally provided for in Katz, but then narrowed by subsequent decisions. Outside of the Fourth Amendment and the Katz line of case law, there exist limited actions in torts which use a test of "reasonable expectation of privacy", and which follow a similar balancing approach[547].

From the use of these approaches in the case law, a pattern emerges of how privacy in public is understood in the United States. In United States v. Knotts in 1983[548], government agents tracked a drum of chemicals to track the defendant's movements. Comparing public highways to open fields, the Court held that monitoring the defendant's location does not implicate the Fourth Amendment because the tracking was taking place on public roads, where the defendants could have been physically tracked by the agents - seeing the tracking device simply as a proxy for the agents' gaze[549]. Along this line of thinking, extending the capacities of data gathering in public is not held as increasing the privacy implications of the activity.

In United States v. Karo, just a year later[550], this issue appeared again except that the tracking device was then transported into a private residence - and the action was deemed a "search" since it concerned the interior of the residence. This shows the unrelenting focus on the "home" as the core of all privacy issues, as well as the paramount importance in US law of the public/private distinction. The tort of "intrusion upon seclusion" creates liability when individuals transgress into privates spaces. It is

[545] Surveillance Self-Defense, *Reasonable Expectation of Privacy*, accessible at https://ssd.eff.org/your-computer/govt/privacy
[546]United States v. Knottes 460 U.S. 276, 278 (1983).
[547] Gorman (n.538)
[548] United States v. Knottes 460 U.S. 276, 278 (1983).
[549] Casey, T. (2007) Electronic Surveillance and the Right to be Secure. *UC Davis L. Rev.* 41: 977.
[550] 468 U.S. 705, 721 (1984).

possible for that to happen in public, but US Courts rarely afford liability for images obtained in public[551].

This was then developed in the Kyllo decision, which illustrated the increasing difficulties in including new means of surveillance in classic scenarios[552]. In that case, the use without a warrant of a heat sensor to detect marijuana cultivation was struck down as not respecting the reasonable expectation of privacy[553]. Justice Scalia of the US Supreme Court held that when "the government uses a device that is not in general public use . . . [then it is] unreasonable without a warrant."[554]

This means that the meaning of "reasonable expectation of privacy" in the US is based on whether the technology is commercially deployed. In other words, whether the technology is reasonably likely to be used by a third party, which can be likened to the European anonymisation test of "means likely to be used" by the data controller. Like that test, what is considered likely to threaten the individual's privacy depends not on the motivations or purposes of the processing, but simply on whether the technology to do so is likely to be used or not. There is an important distinction, however: in the US expectation of privacy conception, once something becomes likely to be used, privacy becomes weaker as individuals have a lower expectation of privacy, while in the EU anonymisation conception, once something becomes likely to be used, privacy becomes stronger as it puts the responsibility for privacy on data controllers.

The similarity between the two concepts allows us to see the same challenges to the expectations of informational privacy: what becomes "reasonable" to use is changing in the age of Big Data, and eventually the means likely to be used to invade one's privacy will become so cheap, powerful, and widespread that there will be no reasonable expectation of privacy left.

As such, under this conception, the reasonable expectation of privacy is bound to be profoundly transformed. But another standard has been developing in different contexts, as identified by Joel Reidenberg in "Privacy in public."[555] In that work, the change in information flows was seen as being taken into account in some Court analyses. An example is "U.S. Department of Justice v. Reporters Committee for Freedom of the

[551] Kaminski, M. E. (2015). Regulating Real-World Surveillance. *Washington Law Review, Vol. 9, No. 113*
[552] Reidenberg (n.85)
[553] Kyllo v. United States, 533 U.S. 27, 40 (2001).
[554] Kyllo v. United States, 533 U.S. 27, 40 (2001).
[555] Reidenberg (n.85)

Press", where the Court denied access to a "rap sheet", even though every item in that sheet was publicly available information[556]. This meant that in this case, whether the information was "public" was not the issue - instead, the information itself was the problem, regardless of how it was created. As we will see, this is evidence of the growing trend that this thesis argues should be developed further.

The changes identified by Joel Reidenberg that led to the need for this new way of understanding privacy in public are mainly based on how not only has information changed but how this change has not reached individuals. Originally, whether data was public or private was actually whether it was accessible or not. If the data could not be gathered, it was "private". This was acknowledged by the Supreme Court in United States v. Jones: "In the pre-computer age, the greatest protections of privacy were neither constitutional nor statutory, but practical. Traditional surveillance for any extended period of time was difficult and costly and therefore rarely undertaken."[557]

Not only is it possible to discover new information in unexpected ways, but what types of data can lead to very personal information can be impossible to predict and depend on the different contexts. This is compounded by the increasing complexity of technology and data creation, which multiplies the sources of data and how they can combine each other in a way which becomes closer and closer to being completely impossible to predict.

For example, in Gonzales v. Google, the US government obtained from Google some user search queries.[558] This was quashed by the government on the privacy grounds that "[a]lthough the Government has only requested the text strings entered . . . basic identifiable information may be found in the text strings when users search for personal information such as their social security numbers or credit card numbers .... The Court is also aware of so-called "vanity searches," where a user queries his or her own name perhaps with other information."[559]. A vanity search is an obvious red flag - yet less obvious ones such as Thelma Arnold's searches on homes for sale near her in the AOL search query incident can be just as revealing yet hard to predict[560].

As such, the obscurity of information plays a powerful role in what social expectations of privacy are. That conclusion follows from the fact that when transparency becomes the

[556] DOJ v. Reporters Comm. for Free Press, 489 U.S. 749 (1989)
[557] United States v. Jones, 132 S. Ct. 945, 963 (2012) (Alito, J., concurring).
[558] Gonzales v. Google, Inc., 234 F.R.D. 674 (N.D. Cal. 2006).
[559] Gonzales v. Google, Inc., 234 F.R.D. 674 (N.D. Cal. 2006).
[560] Schwartz (n.17)

norm (under the “general public use” test found in Kyllo) the legal expectations of privacy are reduced[561]. By the same logic, increasing practical obscurity would increase the legal protections of privacy. By this logic, the difference between the “public” and “private” space is primarily to do with obscurity: privacy in the home was derived from the fact that it was possible to obscure one’s actions in one’s home.

In the same way, in public, perfect transparency was never assumed. Instead, it was always a balance between what data was transparent or obscured while in the public space. The anonymity of the crowds is no longer a given, and the ubiquity of data-gathering devices such as CCTV has created a broad accessibility to personal data, which was acknowledged by the US Supreme Court in Whalen v. Roe, where the Court considered a challenge to a database where pharmacies reported the prescriptions made by physicians. Justice Brennan wrote that “[t]he central storage and easy accessibility of computerized data vastly increase the potential for abuse”, while Justice Stevens argued that “[w]e are not unaware of the threat to privacy implicit in the accumulation of vast amounts of personal information in computer data banks.”[562]

Ease of access to information has long been understood as an important factor - regardless of whether that information is public or private, such as with the “U.S. Department of Justice v. Reporters Committee for Freedom of the Press” case referenced above. In that case, the Court had stated that “there is a vast difference between the public records that might be found after a diligent search of courthouse files, county archives, and local police stations throughout the country and a computerized summary located in a single clearinghouse of information.”[563]

These cases show the transformation of the “reasonable expectation of privacy” test - a test that is no longer adapted in a world where so much data is likely to be collected by so many means - into a test which is based on appropriate levels of transparency, less dependant on whether the individual is inside or outside the home, or whether the data is public or private.

In conclusion, this series of cases has shown that the public/private distinction is actually an obscurity/transparency distinction, driven not just by law, but also by practical means. As practical means change and evolve to make walls and curtains exceedingly obsolete as obscurity tools, and as tools of transparency in public become more and more efficient

[561] Reidenberg (n.85)
[562] Whalen v. Roe, 429 U.S. 589, 591 (1977)
[563] DOJ v. Reporters Comm. for Free Press, 489 U.S. 749 (1989)

and make anonymity in a crowd just as obsolete as obscurity tools, the public/private distinction becomes antiquated and in need of replacement. This is extremely important when it comes to the European understanding of data protection, as this distinction is what the entire doctrine of "identifiability" is based on: whether data is ''identifiable'' depends on whether the data required to identify the person is available, i.e. ''transparent''. In the same way, trying to prevent identifiability is increasingly difficult as more information is available to more parties. The focus should not be on that information not existing, but on the tools used to obscure it.

2. Reasonable Expectations of Privacy in the EU: A Privacy/Data Protection Divide

A "reasonable expectation of privacy" test in the EU was first used in the landmark case of Lüdi v. Switzerland[564], though the words "reasonable expectation" were not directly used, in a case where a Swiss national was convicted in Switzerland of drug trafficking primarily based on a written testimony by an undercover agent and phone recordings. The Court established that there was a violation of the defendant's Article 8 right to private life due to the telephone interception, but that the goal of "prevention of crime", which was necessary in a democratic society, was more important. The Court stated that once Mr Lüdi proceeded to participate in a criminal activity, he "assumed the risk" of encountering an undercover police officer. Thus one can assume that committing certain actions (such as criminal activity) means that a lesser expectation of privacy can be expected.

This has some similarity to the American approach, which also uses certain contexts to determine that a person has relinquished their reasonable expectation of privacy when certain actions are performed. It is interesting to point out that this "action" in the US can be simply using a phone, while in the European context it is pursuing an illegal activity.

The appearance of the "reasonable expectation of privacy" test in Europe was developed after Lüdi in relation to employer/employee privacy issues. In the US, localization tools are common practice in a company setting, as well as computer and communications monitoring, including e-mails and phone communications, but also software recording

[564] Lüdi v. Switzerland, 1992, ECHR, Series A, No. 238, PN 2004-135

every keystroke and mouse click[565]. On the other hand, the EU Courts - and especially the French courts - do not allow a fraction of those employee-monitoring measures[566]. In particular, certain European Union member states (including the United Kingdom, Germany and the Netherlands) strictly prohibit the monitoring of employee communications and under normal circumstances electronic monitoring is not permitted[567].

In the Halford v United Kingdom case[568], an employer was monitoring an employee's telephone transmissions. This was held to be part of one's private life under Article 8, and that there existed a "reasonable expectation of privacy" when making these calls[569]. In relation to the phone calls the plaintiff was performing, the Court noted that she would "have had a reasonable expectation of privacy for such calls, which expectation was moreover reinforced by a number of factors." In this case, it is important to note the broad range of the test, when compared to how strictly it is applied in the American system: in Halford the test seemed to protect all Article 8 rights including family life, home, and correspondence[570], which may have been intentional.

That is not to say the European vision of the "reasonable expectation" test is a wall of protection that is only breached in overwhelming circumstances. In the case of P.G and J.H v. United Kingdom[571], the next case to reference a "reasonable expectation" test, covert listening devices were planted in the plaintiff's home, and though the Court ruled the measure illegal, it did add that "there are a number of elements relevant to a consideration of whether a person's private life is concerned by measures effected outside a person's home or private premises. Since there are occasions when people knowingly or intentionally involve themselves in activities which are or may be recorded or reported in a public manner, a person's reasonable expectations as to privacy may be a significant, although not necessary conclusive, factor"[572].

This case shows that the European conception of "reasonable expectation of privacy" is a balancing test, and not a binary decision. This shows an interesting difference between

---

[565] Determann, L., & Sprague, R. (2011). Intrusive monitoring: Employee privacy expectations are reasonable in Europe, destroyed in the United States. *Berkeley Technology Law Journal*, *26*(2), 979-1036

[566] Ibid

[567] Ibid

[568] Halford v. United kingdom (20605/92) 1997 ECHR 32, PN 2003-49

[569] Gomez-Arostegui, H. T. (2004). Defining private life under the European convention on human rights by referring to reasonable expectations. *Cal. W. Int'l LJ*, *35*, 153. p.10

[570] Ibid

[571] P.G. and J.H. v. United Kingdom, (44787/98) [2001] ECHR 546 2001 PN 2004-199

[572] P.G. and J.H. v. United Kingdom, (44787/98) [2001] ECHR 546 2001 PN 2004-199

privacy and data protection European law: while privacy law recognises the need for balancing interests and contextualising different cases, data protection is still using a binary approach for what is or is not personal. This approach was restated in Peck v The United Kingdom[573], in which CCTV footage of the applicant's attempted suicide was used in news programs without his consent. The Court noted that "the applicant's reasonable expectations of privacy, among other factors"[574] would be relevant to an Article 8 analysis.

A major development came in von Hannover v. Germany[575]. In this case, reasonable expectations were reinforced as the benchmark for a violation of Article 8, discussed as a "legitimate expectation" of protection of one's private life. In this case, the Court ruled that one has such an expectation despite being a person in the public eye. The ''balancing test'' between the right of the public to obtain that information and the right of the individual to privacy fell on the side of the individual and it was held that the plaintiff had a ''legitimate expectation'' of privacy. In a dissenting opinion, Judge Barreto stated that the princess of Monaco's "legitimate expectation" should not be valid where pictures were taken of her in a beach club swimming pool, because there was no possibility of entertaining a "reasonable expectation of not being exposed to public view"[576]. Judge Zupančič was even more vocal in asking for the reasonable expectations test as the standard for privacy cases, as it "permits a nuanced approach to every new case"[577].

However, for this test to be valid, it is important that the exact specification of what constitutes "reasonableness" - whether a subjective, objective, or combined approach - be cleared by the Court. A possible answer can be found in a UK House of Lords case, Campbell v MGN[578]. In this case, Lord Hope discussed the concept of "reasonable expectation of privacy", explaining the House of Lords' position: "the question is what a reasonable person of ordinary sensibilities would feel if she was placed in the same position as the claimant and faced with the same publicity". This objective view is a good indication of the European position. What's more, the European vision of privacy as a "social good" would point towards the same objective approach[579].

[573] Peck v The United Kingdom, (44647/98) [2003] ECHR
[574] Ibid
[575] von Hannover v Germany [2004] EMLR 379; (2005) 40 EHRR 1
[576] von Hannover v Germany [2004] EMLR 379; (2005) 40 EHRR 1
[577] von Hannover v Germany [2004] EMLR 379; (2005) 40 EHRR 1
[578] Campbell v Mirror Group Newspapers Ltd [2004] UKHL 22
[579] Custers, B., van der Hof, S., & Schermer, B. (2014). Privacy expectations of social media users: The role of informed consent in privacy policies. *Policy & Internet*, *6*(3), 268-295.

So we have seen there is at least some understanding of a reasonable expectation of privacy, which impacts every space including the work space (Halford v. United Kingdom[580]), the private space (P.G and J.H v. United Kingdom[581]), and the public space (von Hannover v. Germany[582]). We will now see how that expectation works in the public space.

An ECtHR case in 2000, Rotaru v. Romania[583], confirmed its previous jurisprudence over the reasonable expectation of privacy when it comes to a file stored by agents of the state by stating that this file fell within the scope of Article 8[584]. As to whether public information can be considered part of one's "private life", the Court noted that "Moreover, public information can fall within the scope of private life where it is systematically collected and stored in files held by the authorities. That is all the truer where such information concerns a person's distant past."[585]

This is particularly relevant because this is the same approach taken to "reasonable expectation of privacy" as in the American conception, with the same conclusions to the concepts of transparency and opacity. The conclusion was reinforced in the aforementioned P.G. and J.H. v The United Kingdom, where the Court defined that "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'."[586] Additionally, the Court added elements to decide whether one's private life is affected in public places[587]:

"Since there are occasions when people knowingly or intentionally involve themselves in activities which are or may be recorded or reported in a public manner, a person's reasonable expectations as to privacy may be a significant, although not necessarily conclusive, factor. A person who walks down the street will, inevitably, be visible to any member of the public who is also present. Monitoring by technological means of the same public scene (for example, a security guard viewing through closed-circuit television) is of a similar character. Private-life considerations may arise, however, once any systematic or permanent rec

[580] Halford v. United kingdom (20605/92) 1997 ECHR 32, PN 2003-49
[581] P.G. and J.H. v. United Kingdom, (44787/98) [2001] ECHR 546 2001 PN 2004-199
[582] von Hannover v Germany [2004] EMLR 379; (2005) 40 EHRR 1
[583] European Court of Human Rights, Judgment of 4 May 2000 (Rotaru v Romania), no. 28341/95.
[584] Nouwt, S. (2008). Reasonable expectations of geo-privacy. *SCRIPTed*, *5*, 375.
[585] European Court of Human Rights, Judgment of 4 May 2000 (Rotaru v Romania), no. 28341/95.
[586] P.G. and J.H. v. United Kingdom, (44787/98) [2001] ECHR 546 2001 PN 2004-199
[587] Nouwt (n.584)

ord comes into existence of such material from the public domain. It is for this reason that files gathered by security services on a particular individual fall within the scope of Article 8, even where the information has not been gathered by any intrusive or covert method…"[588]

Despite this acknowledgment of privacy in public in a similar vein as in the American conception, a very big exception based on surveillance is permitted by the ECtHR jurisprudence. This includes the case of Perry v. The United Kingdom[589], in which the Court concluded that camera surveillance in public places without recording visual data was fulfilling a legitimate purpose and was not a breach of Article 8.

Overall, as we can see, privacy law in the EU provides a concept of "reasonable expectation of privacy" in public spaces, taking into account the context and in particular potential data-gathering and data aggregation by private or public entities. This contextual approach to how much something is ''private'' or ''public'' and likely to represent a breach of informational privacy has been effective in adapting informational privacy to new technological and social trends, yet a similarly contextual approach to whether data processing represents a breach of informational privacy has not found its way to data protection, despite signs of it with the ''risk-based'' approach.

Indeed, on the Data Protection side, the challenges of the European approach to data protection regulation remain true in the public context. Though there remain barriers to gathering data from individuals in public, it can be done quite well in certain contexts where the infrastructure allows it. Besides the CCTV common in public spaces (especially in the UK), collecting information can be done increasingly well by those with enough motivation. This includes loyalty cards to track what people buy in stores, transport cards that track journeys by the user, and any other IoT-enabled device that allows the monitoring of user data throughout their day as they spend time in public spaces[590].

As we have shown, there is recognition, in privacy law, that informational privacy needs to be protected by an opacity/transparency balance of which factors can be used to obtain the information. We have also shown that there are signs that the GDPR is showing hints of moving towards such an approach. However, as analysed earlier in this Chapter, the core pillars of EU data protection law are not adaptable to this new type of

[588] P.G. and J.H. v. United Kingdom, (44787/98) [2001] ECHR 546 2001 PN 2004-199
[589] Perry v The United Kingdom: ECHR 17 Jul 2003
[590] Treacy (n.312)

approach, which is leading to a conflict between those two approaches in how to protect informational privacy.

In conclusion, both the European and American privacy law systems, despite their strong differences at both the conceptual and practical levels, have developed an understanding of a reasonable expectation of privacy. However, in both jurisdictions, this reasonable expectation is separate from the normal system of protection of individuals and not part of the main currents of thought and jurisprudence, and in particular in the EU is not taken into account in the data protection side of informational privacy. We have seen the development of an approach where instead of “public” and “private” spaces, a focus on “transparency or opacity” has appeared through judicial decisions brought about by the fundamental changes that public spaces are going through. In particular, a recurring factor is the aggregation of data into new ways of understanding data, showing that the distinction between “personal” and “non-personal” data is increasingly obsolete and a similar ‘’transparency/opacity’’ view is necessary and is showing signs of being adopted.

3. Public Spaces as Privacy-Neutral Environments: The Example of Smart Cities

We have identified that the public/private distinction is quickly becoming less and less useful, in particular with the development of new ways to observe individuals online. We will now show how the development of “smart cities” - alongside the general development of technologies which make public surveillance easier - means that traditional understandings of the public/private distinction will face exponentially greater challenges. We will also show how to understand the role of these new spaces - dubbed “private-public-places”[591].

As we have seen previously, the reason there has never been a strong need for privacy in public is because of the inherent opacity of the individual in a crowd. It used to be that following someone around took up tremendous time and resources - not to mention risk - and there was little personal information to be gleaned from it. This is changing with the “Smart City” phenomenon: increasingly, anonymity in public is disappearing.

[591] Edwards (n.326)

Because smart cities are by definition in the public space shared by all residents, there is no choosing whether or not one is involved in the data gathering activities used by the city. One does not choose to be the target of CCTV cameras, reducing the autonomy that is often heralded as a major theme in the privacy debate. In the European context, this leads to the question: if consent is necessary in private but the government and its third party partners in the smart city space do not need consent to gather data, where does the value of consent stand? This is another highlight of the dissonance that undermines the effectiveness of data protection regulation.

4 . The "Public" Cybersphere: Towards Reasonable Expectations of Privacy Online

The understanding of a "public space" implies several characteristics, which have implications over how privacy functions in those "public spaces". As we will see, the essentials characteristics that impact these spaces exist both in the physical and non-physical space. Indeed, there are characteristics in common between shopping centres and websites[592] which can enlighten us on both. These "pseudo-public spaces"[593] reinforce the idea that it is the practical implications of a situation, and not whether it is "public" or "private" under the mundane conception of those terms, that should form the basis of the applicability of privacy protections.

The spaces in which society evolves in are important, as many human rights are organized around them, with examples such as political expression being a feature of public spaces (with demonstrations as a further example) and intimate spaces such as the home being vital to privacy and personal development. When these spaces undergo an upheaval, society is therefore affected in profound and often unexpected ways. As the digital space becomes an increasingly-important part of our lives, the way we communicate and act changes, and this change needs to be taken into account[594].

[592] Mac Sithigh, D. (2012) Virtual walls: the law of pseudo-public spaces. *International Journal of Law in Context, 8 (03).* pp. 394-412.
[593] Ibid
[594] Ibid

“Public-private spaces” in the digital and physical realms both share a sense of community and human interaction, while being under public sector control. They both play roles that have been important to governance through providing a place where private enterprise and public rights co-exist, and allow public and private broadcasters to interact with viewers[595].

Because of this, the same rising issues that appear in public spaces do the same in private ones - social media being a particularly well-known example. In the same way that there is increasing publicity of public actions due to technology, social media is becoming more and more analysed and public - an example being the increasing use and study of public Twitter tweets for a variety of purposes: the monitoring of individuals in a “public” space leading to compilations of data which are as problematic as the ones European and American courts have held to lead to a reasonable expectation of privacy.

As such, the conclusions we have reached earlier in this Chapter in terms of the need for a reasonable expectation of privacy based on transparency are applicable to every space - physical and digital. We can see that a public/private distinction is a misnomer: there are spaces where individuals are mostly obscured and gathering data is either difficult, or pointless, and as such we have dubbed these spaces “private”. Meanwhile, there are places where transparency seemed evident, while in fact individuals were obscured by the anonymity of the crowd (now no longer the case), and these spaces were dubbed “public spaces”. Moving forward, it is important to take note of these nuances, and avoid being limited to a public/private distinction.

Over the course of this Chapter so far, we have seen that using “identifiability” as a way of determining what is “personal data” is meeting with ever-increasing challenges. This is due to a variety of factors brought about by the rise of Big Data. Because all of the European data protection regulatory system is based on the binary personal/non-personal distinction in which the “identifiability” test is absolutely essential, its limitations mean that the whole regulatory system is in a situation that cannot be easily fixed. We have shown that anonymization and consent, two vital tools of the data protection regulatory apparatus, are failing because of this central limitation, and that despite both of these instruments being made stronger and more complete with the GDPR, the core issues stemming from the limitations of the “identifiability” standard has not been resolved. However, we have seen that EU authorities are becoming open to the idea of

[595] Ibid

an identifiability standard which takes into account various practical factors, a more flexible approach which could be the beginning of a solution.

We have then seen that on top of “personal data” becoming a muddled idea, so is the idea of “personal space”. The public/private distinction in terms of public or private spaces is increasingly obsolete, mainly because various environments, both physical and digital, combine aspects of both. This has challenged traditional understandings of how privacy spaces are decided, but at the same time has uncovered interesting underlying truths that have laid hidden until this upset. The most important of them is that “public” and “private” primarily actually mean “transparent” and “opaque”. The transparency/opacity vision of privacy has emerged, and been acknowledged in the new conceptions of the “reasonable expectations of privacy”, showing a move of privacy law towards an approach based on Information and whether there are enough Guarantees (enough ‘’opacity’’) that it will not be used in ways which breach informational privacy.

We have shown that what decides whether something is a “personal” space or “personal data” is not just dependant on the data that can be collected by “anyone”, but by those who are ‘’reasonably likely‘’ to collect it. This subjective element was shown in Whalen v. Roe where the ease of obtaining data, and not the data itself, was problematic as well as in the privacy concerns arising from third party involvement in the creation of smart cities. This was also shown by the very nature of pseudonymised data: depending on the data held by the one holding the pseudonymised data, the same data can be both personal or anonymised in that person’s eyes.

The subjective nature of data is a source of difficulties for regulatory authorities. Is data personal if there exists one holder (apart from the individual the data pertains to) which can identify the individual in the World? Our argument is that there are actually two concepts at play here: “data”, and “Information”. While data is the same regardless of who observes it, “Information” is the combination of the data and the observer’s perception of it. Because any observer will have their own set of data, biases, and preconceptions, the same piece of data can create different “Information”. As we will see, understanding this difference, and using “personal Information” instead of ‘’personal data’’ as the basis for data protection, allows a balance-based approach in data protection.

C. The Legal Argument for Processed Data as a Separate Construct

The idea that data changes over the course of processing is well-established. What we intend to show in this thesis is not that data changes. It is that data becomes a wholly separate construct, with its own rules, biases, and overall a new legal status. We argue that this new construct, "Information", cannot be held to be inscribed in the continuity of what existed in previous stages. This, we will argue, means that any rules that follow that data, intending to continue binding it as it is changing, are akin to attempting to take a river boat over the delta of a river and into the ocean, and being surprised at being out of one's depth.

I. Informational Privacy as the Core of EU Data Protection Law

The move towards a more flexible approach in European Data Protection regulation can be seen in comments by European officials, such as the December 2013 comments by Neelie Kroes, Vice-President of the European Commission responsible for the Digital Agenda:

"Sometimes, full anonymisation means losing important information, so you can no longer make the links between data. That could make the difference between progress or paralysis. But using pseudonyms can let you to analyse large amounts of data: to spot, for example, that people with genetic pattern X also respond well to therapy Y. So it is understandable why the European Parliament has proposed a more flexible data protection regime for this type of data. Companies would be able to process the data on grounds of legitimate interest, rather than consent. That could make all the positive difference to big data: without endangering privacy."[596]

This comment, which aims to show the need for pseudonymised data, shows the difference between "data" and "Information". In this case, the pseudonymised data is useless in the hands of anyone except those who can link the pseudonym back to the

[596] Kroes, N. (2013), Data isn't a four-letter word, *IAPP Europe Data Protection Congress/Brussels,* 11 Dec. 2013. Available at http://europa.eu/rapid/press-release_SPEECH-13-1059_en.htm.

individual. They have the data, but cannot obtain the Information because of a missing piece.

More generally, the understanding of privacy in public and the “personal data” binary are shown in the doctrine of reasonable expectations of privacy and its recent developments. Specifically, this test was used in numerous cases where the data gathering itself seemed to be innocuous, but where other circumstances made the “transparency” of the data too high, such as where a tracker on a vehicle provided a large amount of data that could still have been gathered by law enforcement by traditional means with enough time and effort[597], or a database that provided too much insight into the lives of individuals, even when that database was composed entirely of public data[598]. In these cases, the issue is not the data itself and its status as personal or not, but the easiness with which Information can be created. Even if the data is anonymised, hidden, or otherwise obfuscated, the question remains as to whether the opacity is enough, and what Information can be created. The same data held by another party or processed by another tool creates different Information. This is recognized by the endorsement of the “opacity/transparency” paradigm as the direction in which the “reasonable expectation of privacy” is developing.

Importantly, because the “reasonable expectation of privacy” test is not bound by the “personal data” test of data protection, it is only is interested by the various factors that affect whether a reasonable expectation of privacy is present. This allows for a more flexible approach, but more importantly there is no need for the data to have any specific status, bypassing the “identifiability” criteria of personal data. Only the Information, and who can obtain it, is relevant. This shows that it is possible to protect informational privacy without the need for the concept of “personal data”.

In conclusion, the idea that the transformation of data makes it difficult to attach any legal obligations to the point at which it is first collected is recognised in privacy law with the increasing traction of a “reasonable expectation of privacy”. However, this development in how to protect informational privacy found in privacy law has not seen an equivalent in data protection, despite hints of it such as the ‘’risk-based’’ approach of the GDPR.

Now that we have shown the role of Information, as well as the fact that it is distinct from ‘’data’’ while being both very powerful and inadequately regulated, we will study how the

[597] United States v. Knotts, 460 U.S. 276 (1986)

[598] DOJ v. Reporters Comm. for Free Press, 489 U.S. 749 (1989)

GDPR is showing signs of moving towards the protection of not just ‘’personal data’’ but instead what we have identified as ‘’personal information’’.

II. Linking ‘’Information’’ to Informational Privacy

Not all Information has an impact on ‘’informational privacy’’. In order only to target Information relevant for our purposes, we need to identify what Information needs to be in scope. In order to capture all relevant data, EU data protection has the notion of ‘’personal data’’, which needs to be protected without having to define what particular societal interests are achieved by protecting it. We argue that this approach is valuable, as we have shown in our first Chapter the difficulties of fully defining the scope of the right to privacy. However, because ‘’personal data’’ is binary it is a limited concept. It is too restrictive, meanwhile the fact that the full extent of the GDPR applies to anyone holding any ‘’personal data’’ means that it also too broad, as many holders of data have to comply with a burden for which they do not have the resources to, particularly when dealing with data which may be combined to create special categories of personal data, even where the holder of data has no reason ever to make those connections.

However, a conception which only burdens to the extent of the Information one holds and takes into account all the relevant contextual factors allows for a scope which captures data that is not relevant or problematic. That kind of data allows little Information creation, and as such needs no or few Guarantees. The GDPR acknowledges that, indirectly, by putting out of scope any data processed ‘’in the course of a purely personal or household activity and thus with no connection to a professional or commercial activity.’’[599] By targeting only professional or commercial activity, the GDPR is implying that some data processing activities exist in social contexts that do not present a risk to informational privacy. This, as we will show, fits in with our conception that Legal Guarantees appear when others, such as Social Guarantees, are unable to effectively limit the ability to process data.

The conception developed in this thesis, thus, has a wide scope, since it needs to include all situations involving ‘’Informational privacy’’. This is why my thesis argues for a move from ‘’personal data’’ towards ‘’personal information’’, a concept which keeps most of the

[599] GDPR, Recital 18

elements of personal data but removes the element of ''identifiability''. Data protection is already headed towards a pragmatic, contextual, risk-based approach, and is limited by the focus on ''identifiability'', and the focus on ''data'' instead of ''Information''. As such, attaching all these tools developed in the GDPR to ''personal information'' instead of ''personal data'' allows for a ready-made regulatory framework.

The only thing missing is linking that ''Information'' back to the interests that need protecting. ''Personal data'' does so by being linked to ''an identified or identifiable natural person''[600]. This goes back to the ''identifiability'' criteria that we have shown is no longer fit for purpose. How does one link ''Information'' back to individuals without falling for the same limitations? The answer might be in the GDPR itself.

As we mentioned when discussing the limitations of ''personal data'', the ''relating to'' element of the definition[601] means that data can be related to an individual not only if it is ''about'' them[602], but also where ''data are used or are likely to be used, taking into account all the circumstances surrounding the precise case, with the purpose to evaluate, treat in a certain way or influence the status or behaviour of an individual''[603] or where ''[the use of the data] is likely to have an impact on a certain person's rights and interests, taking into account all the circumstances surrounding the precise case. It should be noted that it is not necessary that the potential result have a major impact. It is sufficient that the individual be treated differently from other persons as a result of the processing of such data''.[604]

There are many instances where non-personal data may be used to evaluate or influence individuals, or to have an impact on a certain person's rights and interests. However, if that data is not ''identifiable'', it is outside the scope of the GDPR. This ''related to'' requirement, influenced by the risk-based approach, is what the focus of European data protection should be on, looking at the use and impact of the information, instead of trying to build a link, however tenuous, to an ''identifiable data subject''.

This linking of ''personal data'' to the consequences that may affect the individual can be found in other places throughout the Regulation, including the data subject's right not to

[600] GDPR, Article 4
[601] Article 29 Working Party (2007), Opinion 4/2007 on the Concept of Personal Data, WP136, Adopted on 20 June 2007
[602] Judgment of 20 December 2017, *Nowak v. Data Protection Commissioner*, C-434/16, ECLI:EU:C:2017:994
[603] Ibid
[604] Ibid

be subjected to automated decision-making in Article 22 of the GDPR which refers to any decisions ''which produces legal effects concerning him or her or similarly significantly affects him or her''[605], a right which does not mention ''personal data'' at all. The fact that the GDPR is creating new rights which set aside the concept of ''personal data'' is strong evidence of its limitations.

Another important way in which the GDPR's approach shows this trend is the reinforcement of the ''principles'' of the GDPR, and particularly its focus on governance and ''purpose limitation''.

The seven principles are the core of the obligations set out in the GDPR for controllers[606]. They ensure that controllers, when processing data, follow some general rules which extend to all personal data processed by them. The first five principles are mainly articulated around identifying the ''purpose'' of processing, and then establishing why and how the data is processed based on that purpose[607]. The first and second principles are based on having an explicit, transparent, lawful legal basis founded on that purpose (encompassing all of the data included in that purpose), while the third and fourth principles, respectively, are based on only collecting the data necessary to accomplish that purpose (''data minimisation'' principle) and keeping and processing all of the data necessary to ensure the purpose can be accomplished (''data accuracy'' principle). The fifth principle calls for the data to be deleted once the purpose is accomplished (''retention'' period).

But the new principle, the ''accountability'' principle analysed previously, which states that ''The controller shall be responsible for, and be able to demonstrate compliance''[608] with the other principles, is the one which most supports the view that a global, controller-focused approach is inevitable. This principle is the basis for a GDPR compliance framework based on controllers setting up internal policies to ensure the principles are being followed throughout the organisation - a framework which looks at the organisation first and the actual data second. In particular, the Article 29 Working Party recommends a Data Protection Impact Assessment (an assessment of the risks of data subjects)[609] where the data controller is ''matching or combining datasets, for example originating

[605] GDPR, Article 22
[606] GDPR, Articles 5 and 6
[607] Colesky, M., Hoepman, J. H., & Hillen, C. (2016, May). A critical analysis of privacy design strategies. In *Security and Privacy Workshops (SPW), 2016 IEEE* (pp. 33-40). IEEE.
[608] GDPR, Article 5(2)
[609] Article 29 Working Party (2017), Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679, WP 248, Adopted on 4 October 2017

from two or more data processing operations performed for different purposes and/or by different data controllers in a way that would exceed the reasonable expectations of the data subject''. This shows that the same data processed differently may lead to different risks, showing that it is the Information, not the data itself, which should be the focus.

Based on these observations, ''Information'' as expressed in this thesis can be defined as ''any information relating to a natural person'', bypassing the ''identifiability'' criteria altogether. Additionally, the purpose-based approach, based on all of the Information held by controllers and for what purposes they can use that Information is the way forward for the European right to data protection. Controls should be based on the use of the data and regulating those purposes. This is the aim of the ''Guarantees''-based framework developed in the following Chapters: to understand and regulate Information based on what it can be used for, instead of whether it is ''personal''.

Conclusion

Let us recap this thesis so far. In the first chapter, we set the background by explaining that "privacy", in the European context, is based on a common core of central forces, based on the collection, control, and dissemination of information, with control mechanisms being created in order to protect this "Informational Privacy". These control mechanisms can be legal, but can also come from other sources, especially technological sources.

The second Chapter analyses the understanding of the European right to data protection, with a focus on how the European regulatory system functions, which showed it to be a strong, top-down system focusing on having one set of rules across the entire spectrum. The rights-based European approach
, based on a right putting the burden on data controllers and seeing the protection of individuals as the primary goal, recognizes that some contexts require a specific approach, including the personal data of children or health data, but stay within the same overall framework to do so.

As such there exists a certain disconnect between the theorisation of privacy, which is a balance of control measures and new ways of gathering data, and the EU Data Protection regulatory system which is anchored in a binary approach based on "personal data".

We have then, having set the scene for how informational privacy is being protected by both the right to privacy and the right to data protection, explained the factors which have upset the traditional understandings of data and which have exposed certain limitations in the existing regulatory framework. The developments of Big Data have meant that the existing control mechanisms were outmatched, and new situations appeared that they were not prepared for, just as in the 1890s, the USA was not ready for instant photography, but to a much greater degree.

This Chapter has explained that the exercise of data processing is a mixture of technology and human agency, which is affected by a number of biases, limitations, inaccuracies, and change. Because of this, the data one has at the beginning of the process can become something entirely different at the end of the process. More than data, what is actually a risk to informational privacy is the construct we have referred to as “Information”, which is the combination of data and human interpretation and bias, which changes depending on an infinity of factors, especially the purpose for which the data is used. Through studying the transformation of data, we have shown the source of why concepts such as group profiling or anonymisation, though they create “Information”, do not fall into the definition of “personal data”, even though they might affect informational privacy. We have shown not only that an approach which recognises the transformation of data across its processing can truly protect informational privacy in the age of Big Data, but also that an approach with these features has already seen some developments through the increasing risk-based approach taken by the GDPR or the “reasonable expectation of privacy” recognising the need for a flexible approach to tackle issues such as privacy in public.

Through these developments, we have explained what the challenge is, and shown where in the European Data Protection regulatory system the limitations are. However, because the core limitation, the “personal data” binary, is at the centre of the system, moving beyond it to a system in a way that can protect informational privacy in the age of Big Data is necessary. As we can see, the European legislators have accepted this need to move towards a new approach, but cannot fully do so unless the ‘’personal data’’ binary is changed. As shown in this thesis, we propose to move towards a concept of ‘’personal information’’ which is not a binary but made of degrees: any information relating to a data subject is personal information, and if that information is not bound by sufficient guarantees informational privacy is not assured.

This links back to both the beginning, and the end, of my thesis. We mentioned when discussing privacy that it is characterised by a shifting balance between new developments and control mechanisms on them; in other words the "opacity/transparency" recognised in the doctrine of reasonable expectation of privacy, and also part of what makes the difference between "data" and "Information" through the fact that "obscurity" can apply not just to data, but also to the interpretation of it. We have mapped out a way to understand the underlying causes of what creates a need for new control mechanisms, and precisely pinpoint what new effects the changes have on the status quo, considering all the factors that "Information" can take into account but that "data" cannot.

The next Chapter will study the ''Guarantees'' which limit the way Information is used. As we have shown, the EU is already showing signs of moving towards a conception which is aligned with this framework. As such, understanding the Guarantees will allow us to show where the existing measures to protect informational privacy are effective, and where they are not.

Chapter 5. The Guarantees

In this Chapter, we take the previous steps that we have developed and use them to create what we call the Information/Guarantees Balance. In the first part, we go over the scholarly sources that we have used as the foundations for this Balance. We then study other ways in which the challenges of informational privacy have been tackled, and what lessons can be learned from them.

Based on this foundation, we explain the factors which restrict the use of Information, which we call Guarantees. We assert that when Guarantees are balanced, when tools limiting the creation, dissemination, and usage of Information which increase the "transparency" of individuals effectively create a satisfying status quo, informational privacy is protected. We go over these Guarantees and go back to the source of the elements that have changed the status quo, putting all of these different factors and forces under the same terms, to build an understanding of how to protect informational privacy in sometimes unexpected ways.

We will then, in our sixth and last Chapter, go over the Balance itself, first by explaining how it is possible to find an existing, satisfactory status quo in the European context that we can compare with the current climate in order to find out whether data controllers are able to process data in a way that is not limited by sufficient Guarantees. We will then go over how to use various tools in order to rectify the balance, and discuss what we call the "Privacy Toolbox", a set of tools which can be used to built up Guarantees.

As such, we are not intending to propose a replacement for the European data protection regulatory system. Though we are critical of some fundamental aspects of this apparatus, the "Privacy Toolbox" can provide solutions to certain parts of the challenges to informational privacy brought about by the age of Big Data.

A. Scholarly Sources of the Guarantees

I. Lawrence Lessig's Pathetic Dot

In his seminal essay, “Code and Other Laws of Cyberspace”[610], Lawrence Lessig lays out the vision of John Stuart Mill, an important figure of libertarianism who argued for freedom from the government’s power and its power of coercion. His analysis is based on asking what the threat to liberty is, and what means can be used to resist that threat.

Threats can take many forms, one of which in the context of cyberspace Lessig identifies as “Code” itself - the technological environment of cyberspace as a constraint on human liberty. However, that is only one of multiple factors. Lessig looks at these threats, these forces, through the eyes of one regulated by them, represented by a “pathetic dot”[611].

Lessig categorizes four forces, taking the example of smoking. There are legal constraints, on where one is allowed to smoke or not, such as on aircrafts, but that is not the only norm. Restaurants have no-smoking signs, even where there is no law enforcing them. People around the smoker might complain about the smoke, creating pressure on them to regulate their behavior. This type of norm, social norms, can be just as constraining as legal ones.

Another force identified by Lessig is the market, using the example of the price of cigarettes, which can be manipulated by the State through increased taxation to regulate choice. Finally, there are the technological constraints of cigarettes, with health effects and strong odor reducing incentive to smoke certain types of cigarettes or others.

These forces - legal, social, market-related and architectural - are the four forces which bind all actions by individuals according to Lessig. Changing one will change the whole, and each force may be opposed to another.

Lessig then takes this to cyberspace, where the same forces are still present - social norms can be found in what is or is not acceptable to post on Facebook. Legal norms on defamation, intellectual property and hate speech restrict what one can say online. The provision of services and their terms and costs change which ones online users participate in or not. And of course, the technological infrastructure of cyberspace is a major influence on what can be done in it - possibly the most important[612]. These modalities regulating cyberspace have a balance as well, which is influenced by, yet separate from a “real world”’s balance[613]. This disconnect, this balance in a new context

[610] Lessig (n.36)
[611] Ibid
[612] Koops (n.57)
[613] Ibid

that has never seen it before, is a new development, and re-creating the balance found in the real world is a major challenge, not to mention the question of whether even to follow the same rules as those of the real world at all.

Meanwhile, these constraints do not simply appear. They are all, especially in cyber-space, man-made and can be affected using mechanics with various degrees of complexity[614]. In order to exemplify the modification of these forces, Lessig brings up the possibility of theft of radios: to deter thieves, the government may impose a new legal norm such as harsher punishment, but can also make sure radios are more protected by improving their security - such as locking mechanisms if the radio is moved, or a mandatory GPS tracker installed. Depending on what is more practical or a better deterrent, different solutions coming from different sides of the four forces will be needed[615].

Lessig is adamant on one fact, which is the crux of this thesis, that lawmakers do not have to accept the different forces as a given and do not have to limit themselves completely to them. Regulation of the market can regulate the market-led forces on behavior, such as the aforementioned increased tax on cigarettes or a tax break for electric cars. Laws can also change the architecture, for which Lessig uses the example of someone in a wheelchair whom the government can accommodate by legally mandating that wheelchair ramps be placed for their use. Another relevant example for our purposes is the placing of speed bumps on a road to ensure drivers slow down - an architectural guarantee which has the effect of a legal speed limit. Finally, an example which we will draw on later is the use of street lights or cameras in public spaces to reduce crime[616] - this does not prevent one particular crime, or target one particular individual, but attempts to limit all kinds of criminal behavior in the area. As such specific issues can be handled through non-specific solutions.

Meanwhile, laws can also change social norms - through educating citizens, awareness programs, etc. This was an important factor of the anti-smoking campaigns in many western countries: in conjunction with changing the market forces (by increasing the tax on cigarettes) and the legal forces (by making smoking illegal in various public spaces), these campaigns used awareness programs, especially ones targeted at minors, to great effect[617].

---

[614] Ibid
[615] Ibid
[616] Ibid
[617] Gilpin, E, A., et al. (1994), Smoking initiation rates in adults and minors: United States, 1944–1988. *American Journal of Epidemiology* 140.6: 535-543.

There are unseen effects to these types of actions, of course. Because human behavior is not easily regulated, unexpected knock-on effects can occur, and all actions to rectify a situation are not beneficial to all. It adds to the calculations the regulator has to make: should they act directly or indirectly? What are the benefits and costs of the measure? What force to use, and why? What are the potential effects of action?

As one can see, acting on the forces is far from simple. But importantly, while a regulator can add to forces to enforce behavior, the opposite is also true. By removing protection on certain industries, they can be dis-incentivised, while reducing penalties for certain crimes can encourage these behaviors. All of these factors will become highly relevant once applied to the context of informational privacy.

II. Helen Nissenbaum's Contextual Integrity

We have already studied Helen Nissenbaum's ''privacy as contextual integrity'' framework early on in this thesis, as part of a historical study on privacy and its developments over time. We will now develop it under a new light, this time in the context of the Information/Guarantees Balance. There are two reasons to do so. First is to understand the theories that have led us to develop the conception of informational privacy as a balance between Information, and Guarantees limiting the creation and dissemination of that Information. Second, is the fact that the points that contextual integrity shares with our framework, of taking into account factors beyond just the law or just the data, means that some instances of contextual integrity affecting legislation in the US prove the need for such an approach. We will first go over what contextual integrity is and what it brings to this thesis, then go over contextual integrity's significant impact in US law and how it evidences the success of a contextual approach.

In order to understand the idea of identifying a balance, it is useful to go over the theories that have inspired it. Earlier in this thesis, we have argued that the protection of informational privacy is moving away from the binary of public and private spaces, as well as the binary of personal and non-personal data. If these concepts are to be set aside, an updated standard of assessing what Information various parties can obtain and what they can do with it needs to be established.

The theory of contextual integrity was created by Helen Nissenbaum, and is based on the idea of creating not a system of law, but instead a way to try and establish a "foundation for policy and law expressed in terms of moral, political, and social values"[618]. Because of the huge range of interests that privacy encompasses, Nissenbaum's theory attempts not to provide an answer to all questions of privacy, but instead to find a set of principles allowing for the study of privacy without being bound by discussion on what the limits of the right to privacy actually are. Another vital feature of contextual integrity - and the primary reason why it is included in this thesis - is the aforementioned separation from the binary of public and private spaces.

A second principle is the type of information involved and the restriction of its access and transmission. Instead of focusing on who is intruding, the principle is based on the fact that *someone* is. Around that principle lies the idea that people are entitled to have secrets, and that one of the major values of privacy is the ability to decide on the degree of disclosure one is willing to give about those secrets. This leads to the difficult categorization of what data is "personal", "sensitive", or other names for data that is held as entirely controlled by the individual. The United States, like the EU, holds "sensitive" data in a special category, and as such the idea of having certain data being particularly important or sensitive is not a cultural conception but a more wide-ranging one. In particular, any data related to health or children is held to need special protection. As we will see later, this creates interesting dynamics where some "sensitive" data not covered by existing legislation can still be protected by the social stigma associated with collecting it[619].

The final principle identified by Nissenbaum pertains to what is held as the "Personal" or "Private" sphere. In our discussion on the distinction between public and private spaces, we have shown that the distinction between these is in fact made up of shades of opacity. The idea of "opacity" is more difficult to establish in the US conception because of its insistence on the "home" as the central space belonging to the individual, but is nevertheless gaining traction as acknowledged by Nissenbaum who points out the existence of these "gray spaces". Taking these principles as the "skeleton" around which privacy needs to be articulated, Nissenbaum uses them to create the framework of contextual integrity[620].

[618] Nissenbaum (n.102)
[619] Ibid
[620] Ibid

These principles are vital to understand, because contextual integrity is not aimed at an answer to all challenges to informational privacy but to work at the foundations of these principles. The idea of contextual integrity is that all areas of life are governed by "norms of information flow", each inscribed within a certain context. Instead of a "public" or a "private" context, each individual situation is its own context with its own rules, depending on a number of factors.

Nissenbaum uses "appropriateness" as the measure of whether the flow of information is adequate, which allows for great variety and adaptability, with such wide-ranging possibilities as the appropriateness of a doctor-patient relationship or the one between a married couple. Different contexts with different people mean that the norms of appropriateness being violated are enough to characterize a breach, which allows a way to bypass the "personal data binary" while still protecting privacy. In addition to appropriateness, "distribution" is another set of norms on the transfer of information obtained within a context. For example, if one's friend confides very personal information, distributing that information to strangers is a breach of contextual integrity[621].

On top of these norms come "roles", such as the role of patient or doctor, and they dictate the types of acceptable information transmission, and principles of transmission which are the rules governing the transmission of information - not just what information is shared, but why and how[622].

III. The Limitations of a Harms-based Approach to Protecting Informational Privacy

We have shown that what needs to be regulated in order to protect informational privacy is the creation of Information, not just the flows of data. Stronger privacy and data protection laws become necessary when Information becomes created in new and expected ways. We will not attempt to delimit exactly why informational privacy is important, or what "harms" or "risk" to it mean. Each of these questions could be (and has been) the subject of a library's worth of scholarly works. In this part, we will go over

[621] Ibid
[622] Adam, B. (2006), et al. Privacy and contextual integrity: Framework and applications. *Security and Privacy, 2006 IEEE Symposium on*. *IEEE*.

the reasons why we have approached privacy in this way, and the pitfalls that a "balance" approach would avoid.

In this part we will go over the "privacy harm" approach to privacy and why we have chosen not to use it for this thesis. Nevertheless, lessons from it have inspired the way in which the Informational Guarantees framework has developed, and so this approach needs to be mentioned.

We have mentioned before Daniel Solove's framework of privacy harms[623]. However, Solove himself was aware of the limitations of such an approach. Finding ''privacy'' interests and where they are breached is difficult: "Too many courts and policymakers struggle with even identifying the presence of a privacy problem. Protecting privacy requires careful balancing, as neither privacy nor its countervailing interests are absolute values. Unfortunately, due to conceptual confusion, courts and legislatures often fail to recognize privacy problems, and thus no balancing ever takes place[624]. This does not mean that privacy should always win in the balance, but it should not be dismissed just because it is ignored or misconstrued."[625] This goes straight to the core of what the Information/Guarantees framework attempts to do: avoid the very question of finding "privacy problems", and instead find a way to re-establish the same balance as it existed before it was upset.

There is undeniable value to taxonomies of privacy. They allow us to have a complete and nuanced view of the interconnected interests of privacy, by classifying the complex interest that it is. Solove already solves many of the problems of classifying privacy by focusing on a pragmatic approach of ''privacy problems'' instead of the elusive search for the meaning of privacy[626], which shows the American origins of the approach, following the "privacy torts" theorised by Warren and Brandeis.

The limitation of introducing "harm" into informational privacy is that it introduces the idea of finding who has been "wronged" by the activity in question[627]. Finding such injuries in

[623] Solove (n.87)
[624] Ibid
[625] Ibid
[626] Calo (n.20)
[627] Austin, L. M. (2014), Enough About Me: Why Privacy is About Power, Not Consent (or Harm). Forthcoming in Austin Sarat, ed., A World Without Privacy?: What Can/Should Law Do.. Available at SSRN: http://ssrn.com/abstract=2524512

the context of informational privacy is difficult. Beyond the simple fact that proving moral harm is difficult due to its subjectivity[628], finding the right boundary is also difficult.

We have seen how challenging defining the limits of privacy is, which is why we have limited this approach to "informational privacy". A recent paper by Bert-Jaap Koops and others, "A Typology of Privacy"[629], attempts to map out the various interests which apply to privacy, which includes two scales - from freedom to be left alone to freedom for self-development, and from the "personal zone" of solitude to the "public zone" of inconspicuousness. In this conception, there are a great number of dimensions to privacy: behavioral privacy, proprietary privacy, spatial privacy and more[630]. Informational privacy permeates every one of these dimensions. Is there harm in a data controller selling pictures of an individual they uploaded on their server? It is informational privacy, but is it also bodily privacy, proprietary privacy, something else? Even if harm can be characterized, finding which type of privacy is harmed is a whole other question all to itself.

Secondly, there are types of harm not even fully recognized as harm at all. Some argue that privacy goes beyond the individuals, and that there is a "societal" dimension to privacy, where the violation of an individual's privacy not only harms the individual, but also society as a whole[631]. Protection of autonomy and of private spaces are held as vital to democracy and society[632], but this extends as well to places such as lectures theaters, intended to be a place for the free exchange of ideas and designed for academic freedom (as seen in Antović and Mirković v Montenegro in which surveillance of lecture halls was seen as an invasion of privacy)[633]. Finding the "harm" there is difficult, or even impossible: how can one measure the erosion of democracy? How can a value be put on that, especially when the erosion is taking place through millions of individual transfers and gatherings of data?

---

[628] Koppelman, A. (2005) Does obscenity cause moral harm?. *Columbia Law Review* (2005): 1635-1679.
[629] Koops et al (n.24)
[630] Ibid
[631] See "Schwartz, P. M. (1999). Privacy and democracy in cyberspace. *Vand. L. Rev.*, *52*, 1607.", "Reidenberg, J. R. (1994). Setting standards for fair information practice in the US private sector. *Iowa L. Rev.*, *80*, 497".; "Schwartz, P. M. (1994). Privacy and participation: Personal information and public sector regulation in the United States. *Iowa L. Rev.*, *80*, 553." and "Simitis, S. (1987). Reviewing privacy in an information society. *University of Pennsylvania Law Review*, *135*(3), 707-746."
[632] Reidenberg, J. R. (2002), Privacy wrongs in search of remedies, *Hastings LJ* 54: 877.
[633] Antović and Mirković v Montenegro [2017] ECHR 1068

Finally come the challenges coming from what we have called ‘’data sculpting’’. As we have developed, “personal data” becomes something completely different over the course of its life. That something different becomes utterly disconnected from the original data, and tracing back harm to the original individual can be exceedingly difficult. Except for societal harm, all harm eventually has to be traced back to “who” has been harmed. In the Sculptor’s Work, finding that “who” might very well be impossible.

The GDPR has acknowledged this, by using a conception of ‘’risk’’ which does not define ‘’harm’’[634]. As we analysed earlier, guidance by the Article 29 Working Party deliberately stays away from defining in what ‘’risk to the rights and freedoms of data subjects’’ consists, but instead defines types of processing which are inherently ‘’high risk’’. This shows a shift away from “harm” and towards a contextual, balance-based approach: if the flows of Information are limited by insufficient Guarantees, then the processing is high risk.

In conclusion, the harms-based approach to privacy holds some valuable advantages. Most importantly, it allows a way to escape the “personal data” binary, as the “harm” is not reliant on the data itself but is linked back to the individual through the notion of being “harmed”. Nevertheless, the notion of “harm” in the field of privacy is as elusive as privacy itself, and even when narrowing the scope down to “informational privacy” the “harm” is not clear. As we have seen, this was acknowledged by the Article 29 Working Party, which focuses their approach of “risk” not on the possible harm, but on the practice and its Information implications. In this way, if any Information controller is bound by insufficient Guarantees, informational privacy is not protected, regardless of the ‘’harm’’ that might materialize. This has allowed the GDPR to move beyond the difficulties of defining the broad range of interests that may be protected by informational privacy, and focus instead only on informational privacy itself.

B. The Guarantees: A Multidisciplinary Approach

We have gone over the core theorists who have inspired this thesis, why their works are so important, and why we use some of these ideas to build an approach based on balancing Information and the Guarantees on that Information. From Lessig, we take the framework of the four forces and their interactions, to represent the Guarantees which

[634] Macenaite, M. (2017). The “Riskification” of European Data Protection Law through a two-fold Shift. *European Journal of Risk Regulation*, *8*(3), 506-540.

make it so that a data controller's ability to create Information is limited. From Helen Nissenbaum's contextual integrity, we take the idea of creating a balance instead of using a standard which tries to pin down the inscrutable nature of privacy, and looking at each context on its own merits rather than using the "personal data" binary. However, instead of using flows of information, we focus on establishing a set balance and looking at its state in different contexts based not on the flow of information, but on the Information controllers themselves, what Information they hold, and the Guarantees that bind them. The flows that led to them controlling that Information are only relevant in that they bring with them Social, Legal, or other Guarantees on the controller.

In this section, we will explain how these forces can aid us in mapping out informational privacy interests, while showing that the use we aim to make of it, different from the one laid out in Lessig's work, will place the Information creator in the position of "pathetic dot".

In this part, we go over the concept of "Guarantees", the elements which bind every individual, corporation and government and restrict their access to Information and their ability to transform, transfer, or use that Information. As we will show, these Guarantees take a wide variety of forms, from the fact that it is impossible to read minds to different anonymisation techniques. To explain this vital concept and why it constitutes a balance, we will first explain the substitution of Lessig's "modalities" for the word "Guarantees". We will then go over each type of Guarantee and its place in the overall framework. We will pay special attention to the non-Legal Guarantees, as their coercive power can be very important, in particular Market and Technological Guarantees which have seen the greatest upsets over time. As such, we will first discuss Technological Guarantees, before moving on to Market Guarantees, Social Guarantees, and only then will we discuss Legal Guarantees.

I. Understanding the term ''Guarantees''

The first matter to address is the use of the term "Guarantees" to describe the elements which constitute the framework of protection surrounding Information. Lessig's framework talk of "forces", pressures exerted from various sources on the "pathetic dot". These forces pressure each other as well as the dot, who is swayed one way or the other. Guarantees, however, do not sit on both sides or conflict with each other. Instead, every element that makes it so that Information cannot be freely obtained, transmitted and used for whatever purpose its collector wishes is a Guarantee, while any change that makes it easier to create that Information is not an opposite force, but simply the

weakening of an existing Guarantee. The ability to record someone with a microphone is nothing but the removal of the Guarantee that one's speech cannot be recorded, which had always existed but never been in question, until the appearance of that innovation. Understanding the distinction is vital to understanding the framework of Informational Guarantees, which is why we will go over what this means for each of the various types of Guarantees so that the reader can fully grasp the implication of using "Guarantees".

The choice to use these Guarantees is motivated by the directing line of this thesis: finding a way to treat every technology and change under the same rules. For that, a universal core has to be devised which is always true. In this case, the core is the complete inexistence of Guarantees - which is only theoretical, as it would require omniscience. As such, compared to that state of affairs, the informational privacy landscape a hundred years ago and the one today are actually not that different - there are still powerful Guarantees in place, but some have been weakened or disappeared. In short, the question is not which "attacking" or "defending" forces surround Information, but only which Guarantees, how strong they are, and whether the balance is even.

An obvious criticism of comparing Lessig's four forces as though they have the same authority is the fact that one stands out as seeming to have a higher coercive power - the Law. If one ignores social norms, they cannot forcibly be put in prison for it. If one ignored market forces, they will only find themselves poorer, but still free. This is where the word "Guarantees" continues to bring value. As we will show when discussing the specific types of Guarantees, the power of a Guarantee does not lie in direct coercive power. The word "Guarantee" is to highlight that just because something is a guarantee does not mean it is certain. A "force" tends to evoke the image of sustained, constant pressure. A "guarantee", on the other hand, is just as real but can also happen to fail. Trusting in a friend that they will not reveal your secret is a Guarantee; you trust they will keep your confidence, but cannot know that they will for sure or be able to have a hard remedy such as legal action if they do not. This follows the recurring theme of moving from a binary approach to one which follows the same kind of idea as the one hinted at with the "identifiability" requirement: the "means likely to be used". No Guarantee is perfect and no Guarantee can provide protection with 100% certainty. Instead, a Guarantee only has to make it "reasonably likely" that it will be successful in doing what it promises to do. And that distinction allows for a much greater umbrella of practices and tools, which is why using the term "Guarantees" instead of "forces" is valuable.

With the meaning of "Guarantees" in mind, we will now go over each type of Guarantee and their implications.

II. Technological Guarantees

In order to understand what we mean by Technological Guarantees, it is vital to imagine a hypothetical world where there would be no such Guarantees whatsoever. If there were no Legal Guarantees, then individuals would see no protection for their data coming from the State. If there were no Social Guarantees, there would be nothing to stop those who know one's secrets from sharing them at their will. Without Market Guarantees, those interested in gathering one's data would be able to deploy all the available resources to do so. All of these are impossible scenarios and each of them would prove to be almost impossible to deal with. However they are still one level under what would happen without Technological Guarantees.

Technological Guarantees are all the ways in which Information is difficult or otherwise impossible to create, whether it is by limiting what data can be collected, whether it can be processed, stored, mined, accessed and transferred. As an example, most browsers have an "incognito mode" which limits what data is stored on one's computer or given to websites one visits, and this constitutes a Technological Guarantees] - less data is gathered which means less Information is produced.

However, another Technological Guarantee is the fact that it is impossible to read minds. That might seem obvious, but 100 years ago it was impossible to track someone through a gadget in their pocket; the technology simply did not exist[635]. As long as the technology to do so does not exist, then one's thoughts remain entirely private information. All throughout history ways of finding out what people are thinking have been attempted due to the value of that information - from lie detection[636] to behavioural prediction[637], the field of "thought identification", as it is called, is a strange field often on the fringe of pseudo-science. Nevertheless, there is no reliable way to read minds, but if there were it would create immensely valuable Information. There is a deceptively high amount of Information which similarly cannot be created because the technology is not there. Dreams, emotions, conversations away from microphones, actions away from cameras,

[635] Koops (n.57)
[636] Granhag, P. A., & Hartwig, M. (2008). A new theoretical perspective on deception detection: On the psychology of instrumental mind-reading. *Psychology, Crime & Law*, *14*(3), 189-200.
[637] Baumann, C., Burton, S., Elliott, G., & Kehr, H. M. (2007). Prediction of attitude and behavioural intentions in retail banking. *International Journal of Bank Marketing*, *25*(2), 102-116.

and so on. Despite technology seemingly being ubiquitous, there are still many limits to what can be achieved using it.

The inability to even to obtain the data is an important Technological Guarantee, and taking this "barriers to omniscience" approach allows us to understand the Information/Guarantees Balance better. When a new technology appears that allows for collection of data previously uncollectable, it is nothing but a change in the balance of Guarantees: a Technological Guarantee that existed previously (the inability to collect the data) has disappeared, which has made the balance uneven and needs to be corrected, as Warren and Brandeis did in reaction to gossip columns and instant photography.

The example of Warren and Brandeis leads us to the next point in terms of Technological Guarantees, which regards the steps following data collection. In the same way that data which cannot be collected has no value, data which cannot be stored, transferred or processed is similarly useless. For example, human memory is a "data storage device" – albeit a biological one. But because the "storage" is imperfect and limited, there are many types of data that we can "collect" (through sight, smell, or other senses) but cannot store in a meaningful fashion. The most obvious example is crimes with no witnesses but the victim - the data is there and exists in the minds of both the criminal and the victim, but cannot be stored, transferred, processed, or accessed by any realistic means.

Thus, the fact that we cannot (or rather, could not) store and process various kinds of data means that little or no Information can be extracted from it, which constitutes a Technological Guarantee in and of itself. When looking at the history of the Balance throughout time, this perspective allows for a well-aligned view of the Guarantees involved. The development of new data-gathering technologies is not an outside force requiring Guarantees to control it. Instead, it is the disappearance of existing Guarantees and the subsequent need for new Guarantees to appear or existing ones to be reinforced. That is what happened with Warren and Brandeis, and what happened with the events which led to the development of the first data protection legislation efforts. In the same way, any barriers created to limit once again the capabilities of data gathering[638] (such

[638] de Koning Gans, G., Hoepman, J. H., & Garcia, F. D. (2008, September). A practical attack on the MIFARE Classic. In *International Conference on Smart Card Research and Advanced Applications* (pp. 267-282). Springer, Berlin, Heidelberg.

as tools to help users protect or manage their privacy) are also new, man made Technological Guarantees[639].

III. Market Guarantees

So far we have shown that the absence of a technology is a Technological Guarantee in itself. But many technologies now exist to record various types of data, from our heartbeats to our sleep patterns. So if that is the case, why is it that this data is not available and gathered by every individual, corporation and state? Setting aside for a moment the intervention of the Law and Social norms which limit various ways in which data can be gathered and Information created, a major intervening factor are the Market Guarantees and how they interact with Technological Guarantees.

The crux of this argument comes from the requirements for identification in the GDPR, which takes into account "the means likely the be used" for identification by the data controller or processor[640]. We have shown when discussing anonymisation that with enough time and effort, any anonymisation can be broken. Does that mean that anonymisation is pointless? No, in fact it is a very powerful tool and is widely used and relied upon in a variety of fields and does so well. This is where the "means likely to be used" test comes in. Even though the technological means exist to penetrate anonymisation, not everyone has the resources to do so. Moreover, not everyone with these resources to do either because the benefit of doing so might not be worth the expense. Finally, deploying data gathering tools is an expense as well, which means it also might not be worth it. To explain what we mean we will go through a few examples.

- A corner shop gathers data on its customers' purchases. That data could be used to create Information pertaining to the customers' personal lives, such as their religion (if they buy halal products for example) or their family life (such as buying child diapers) but attempting to ascertain for sure if those assumptions are correct is not particularly useful for that local shop. As long as the tastes and preferences of the individual are established, looking deeper into their personal life is an unnecessary expense. As such, no reasonable business would go to those

[639] Notario, N., Crespo, A., Martín, Y. S., Del Alamo, J. M., Le Métayer, D., Antignac, T., ... & Wright, D. (2015, May). PRIPARE: integrating privacy best practices into a privacy engineering methodology. In *Security and Privacy Workshops (SPW), 2015 IEEE* (pp. 151-158). IEEE.
[640] GDPR, Recital 26

lengths, and so the “means likely to be used” by that supermarket are limited by that fact.

- A social media platform records all messages between individuals. It could attempt to find out how many individuals lie to each other by comparing their messages and other data that might be gathered. Nevertheless, that would have almost no value to the social media platform itself, so it is not likely to use any means to obtain that data.
- A jealous co-worker wants to find dirt on their high-achieving colleague to mar their chances at a promotion. However, they cannot afford a private investigator to spy for them. The “means likely to be used” for their investigation is limited by their financial means, and so the Market Guarantee of that data being too expensive to gather remains.
- A voice-activated home assistant device collects data to check whether its activation phrase is spoken and what the commands of the individual are. While it could be always-on, collecting data from conversations and spying on private moments, that would mean consuming a tremendous amount of bandwidth, which could have prohibitive costs. In this case, the cost of bandwidth constitutes a Market Guarantee.

Through these examples, one can see the effect that Market Guarantees have. On the one hand, they might seem weaker or less binding than Legal Guarantees. On the other, them and their interactions with the limitations of Technological Guarantees are what has shaped much of our current environment, and every Legal Guarantee developed to protect informational privacy has done so in response to what is essentially a change in the “means reasonably likely to be used” by various entities.

These Guarantees actually constitute most of the informational privacy environment. There is no law for the theft of large buildings, because there exist no means to steal such a building. In the same way, before the 1960s, there was no law for the collection of huge amounts of personal data, because there existed no means to do so. The change we are seeing now is that the law was designed for an environment where these Guarantees seemed set in stone, and did not account for their erosion, or at least not enough to keep up with unexpected developments. The main reason behind working on this framework of Informational Guarantees is to provide a way to find the core forces creating this balance, allowing us to find where the change in Guarantees is and what Guarantees are needed to rectify it.

We have now gone over the two major types of Guarantees which dictate the environment in which these rules operate. Understanding the importance of Market and Technological Guarantees is vital to understanding why any change in these Guarantees is so impactful. This goes back to our previous discussion on the "opacity/transparency" conception of privacy, and reinforces the idea that socio-legal rules on privacy only appear in response to new types of transparency. With that in mind, we will now go over Social Guarantees, and their highly-contextual place.

IV. Social Guarantees

1. Social Guarantees in Society

Social Guarantees are often the most fleeting of Guarantees, as a Social Guarantee exists where a party is both able and legally allowed to reveal, discover, or transfer Information, but chooses not to do so. One might equate it to "trust", which covers the most common uses of the Social Guarantees, but they go beyond that.

A simple example is trust between friends. There is no law for friends not to reveal embarrassing or private facts about each other to strangers, and there is no true way of making sure that your secrets are not being shared. Social contexts are complicated and follow many different norms - as highlighted in Helen Nissenbaum's work on contextual integrity which looks at the anatomy of these various flows of information[641]. Social Guarantees can come from duty (such as a priest towards confessors), love (such as a couple), trust (between friends), or simple respect (between coworkers). They might seem to have little binding force, but in our day-to-day lives we expect these Guarantees, not the Law, to ensure our privacy.

The reason why these Guarantees and their overall implications have taken a backseat in this thesis (unlike the Legal, Market and Technological Guarantees which have been examined in the first half of this thesis) is because the upsets that have shaken the world of privacy have not affected every context equally. While individual/business, individual/government and government/business relationships have changed dramatically, individual/individual relationships remain largely the same.

[641] Nissenbaum (n.102)

Social Guarantees are more complicated when it comes to companies and governments. Companies are primarily driven by profit and have little social obligations towards individuals. This is only becoming more of an issue when taking into account the "Sculptor's Work" we discussed earlier as well as the fact that data can be transferred to a long chain of controllers over the course of its life. As these factors get more prevalent, social ties become increasingly weaker, to the point that any Social Guarantee disappears. Nevertheless, as we will show, even though traditional individual/business relationships have few Social Guarantees, such Guarantees can appear from other sources.

2. Social Guarantees on the Internet

Despite the fact that businesses would seem to be immune to most Social Guarantees - as they have very little social obligations towards their customers - there is still some influence in the form of public outrage. Indeed, a data controller could potentially create Information on individuals in all legality, but which might be considered immoral or abusive, leading to backlash. Examples abound for social media, from Facebook performing research into targeting emotionally vulnerable teenagers[642] or manipulating news feeds in order to influence the emotions of its users[643], or Twitter selling the public tweets of millions of users for marketers and researchers to mine data[644]. These examples can be compounded with the Snowden revelations[645], in which it was shown that under the PRISM program, the United States government had been secretly gathering data from large Internet companies[646].

Does public outrage matter? Well, coming to mind is the idea that bad publicity would make the business lose customers, or motivate customers to increase their own privacy settings. The exposition of these practices being akin to the removal of a Social

[642] Whigham, N., Leaked document reveals Facebook conducted research to target emotionally vulnerable and insecure youth. News.com.au, 1/5/2017, available at http://www.news.com.au/technology/online/social/leaked-document-reveals-facebook-conducted-research-to-target-emotionally-vulnerable-and-insecure-youth/news-story/d256f850be6b1c8a21aec6e32dae16fd

[643] Booth, R., Facebook reveals news feed experiment to control emotions, 30/6/2014, The Guardian, available at https://www.theguardian.com/technology/2014/jun/29/facebook-users-emotions-news-feeds

[644] Mejova, Y., Macy, M. W., Weber, I. (2015), Twitter: A Digital Socioscope, *Cambridge University Press*

[645] Beckett, L. (2017), Big Data Brokers: They Know Everything About You and Sell it to the Highest Bidder, Gizmodo, 18 March 2013, available at: http://gizmodo.com/5991070/big-data-brokers-they-know-everything-about-you-and-sell-it-to-the-highest-bidder

[646] Preibusch, Sören. Privacy behaviors after Snowden. *Communications of the ACM* 58.5 (2015): 48-55.

Guarantee (there is a loss of the Guarantee that these practices are too immoral for businesses to consider), another Guarantee has to fill in the gap, either a Market one (the loss of customers) or a Technological one (increased measures taken by individuals), and the Balance would remain even. But is that actually the case?

Research by Sören Preibusch looked into privacy-preserving behaviours after the Snowden revelations. It showed that mistrust by German Internet users of government and corporate data processing increased by 9% between 2011 and 2013[647], but with only a minority showing they actually changed how they managed their personal data[648]. Meanwhile, Google Trends showed that searches that might arouse the suspicions of the US Government declined in various countries[649]. Nevertheless, the researcher agreed that measuring this data is difficult. Through studying multiple patterns - privacy settings in browsers, new web pages created around the topic of government surveillance, and measuring visits to privacy-related pages, they showed that though privacy-aware behavior increased in the 30 weeks after PRISM was revealed, it had then faded. Overall, the Snowden revelations did not seem to have the corresponding increase of Guarantees hoped for by privacy campaigners.

In the grander scheme of things there is still a move from Facebook towards other, more privacy-preserving services. This is due to a variety of factors, including individuals not wanting to be connected on social media with their parents[650] or being annoyed by Facebook's features, with half of current Facebook users saying they'd considered leaving the website[651]. Nevertheless, across the major factors behind leaving Facebook, a Pew Research Center survey identified privacy concerns as the main issue for 4% of users considering quitting Facebook[652].

Still, despite these relatively small numbers, the last few years have seen some strong initiatives from various social media platforms to improve their privacy settings and user control. Facebook has made wide-ranging attempts to create and promote user-friendly

---

[647] BITKOM (Federal Association for Information Technology). (2013), Internetnutzer werden misstrauisch,
[648] Dierig, C., Fuest, B., Kaiser, T., and Wisdorff, F. (2014), *Die Welt* ; http://www.welt.de/wirtschaft/article126882276/Deutsche-unterschaetzen-den-Wert-persoenlicher-Daten.html
[649] Marthews, A. and Tucker, C. (2014), Government Surveillance and Internet Search Behavior. *SSRN Working Paper, Social Science Electronic Publishing, Inc., Rochester, NY*.
[650] Lang, N. (2015), Why teens are leaving Facebook: It's 'meaningless', *Washington Post*
[651] Stieger, S., Burger, C., Bohn, M., & Voracek, M. (2013). Who commits virtual identity suicide? Differences in privacy concerns, internet addiction, and personality between Facebook users and quitters. *Cyberpsychology, Behavior, and Social Networking*, *16*(9), 629-634.
[652] Rainie, L., Smith, A., & Duggan, M. (2013). Coming and going on Facebook. *Pew Research Center's Internet and American Life Project*.

privacy controls for users[653], even as recently as 2017[654]. Meanwhile, WhatsApp is offering full end-to-end encryption for its messages, documents and calls for its hundreds of millions of users to promote their privacy[655], while Facebook and Instagram, following news that police gained access to their platforms to track protesters, updated their rules so that developers could no longer "use data obtained from us to provide tools that are used for surveillance"[656].

But an even more powerful example is the fact that certain brands are actually trying not simply to reassure users through giving them new Technological Guarantees, but actually working towards creating an entirely new type of rapport, a "social contract", with their users to build this "trust" element which would otherwise seem to be absent in the relationships between consumers and the amoral nature of businesses[657]. This very much shows the way in which the disappearance of a Guarantee finds its own need for replacements eventually.

These examples seem to be responding not just to data protection legislation, but to social outrage and pressure by users. In the last example, especially, the increased Technological Guarantees are a direct response to the public outrage over government surveillance. There is difficulty in tracking exactly to what extend the attitudes to privacy and demand for privacy by users are pushing forward the development of other Guarantees, but there is no doubt they have had an impact.

The impact of Social Guarantees despite the absence of traditional social obligations thanks to social indignation is a useful tool on the side of the data subjects. Nevertheless, such indignation will only be a relevant force as long as these types of Information creation do not become the norm.

---

[653] Stern, J., Facebook's New Privacy Controls Roll Out to All U.S. Users, ABC News, 20/12/2012

[654] Newcomb, A. Facebook's New Tools are Designed to Put Privacy in Your Pocket, NBC News, 26/1/2017, available at http://www.nbcnews.com/tech/tech-news/facebook-s-new-tools-are-designed-put-privacy-your-pocket-n712246

[655] Groome, I. What is WhatsApp encryption?, *Metro News UK,* 27/3/2017, available at http://metro.co.uk/2017/03/27/what-is-whatsapp-encryption-6535899/

[656] Levin, S. Facebook and Instagram ban developers from using data for surveillance, The Guardian, 13/3/2017

[657] Hosea, M. Why brands are creating 'social contracts' to build trust around data use, Marketing Week, 4 October 2016, available on https://www.marketingweek.com/2016/10/04/why-brands-are-creating-social-contracts-to-build-trust-around-data-use/#content, accessed on 16/05/2017

The value of informational privacy as a social good is an important Social Guarantee, and its disappearance would create a strong imbalance. It is important to remember when it comes to the shifting of Social norms that this thesis takes a certain point in time as the even balance, and works under the hypothesis that having an even balance is objectively a good thing. If society cared not at all for privacy and threw out all Guarantees completely, there would be nothing wrong in subjective terms, but the Information/Guarantees Balance would be uneven.

As we have seen, as long as wide-ranging surveillance is not "normal", there will be the Social Guarantee of outrage and pressure on services to adopt policies which match the privacy preferences of the general public. However, this "normality" of privacy is itself in flux. Like every value, privacy is subjective and affected by a number of currents and transformations. As the new generations are formed with social media in their day-to-day lives, there are concerns that the way they define their generational identity is different from previous ones[658]. Perception of the value of social media is changing, becoming so central that losing privacy to continue using it becomes a logical tradeoff[659]. Meanwhile, social media presents its own version of privacy, being able to connect and talk and relate to others in a bespoke context, chosen, and separate from one's home life and the gaze of parents and teachers. This makes online activity not merely more "public" than offline activity, but also the creator of different Information, under different Guarantees. Many on the Internet would not care if a stranger had Information on them, but would care if those they know had. As such, the ability to have a separate online life has value and is not necessarily a problem in terms of the Information/Guarantees Balance.

An important consideration is the question of whether dissociating the online and offline contexts is valuable in the first place. On the anonymity side of the discussion are figures and groups stemming primarily from the tech community, from hackers to privacy advocates. On the other side stand those who believe there should be no divide between the online and offline contexts and that one person has one identity, with Facebook as one of its major proponents[660]. Many States follow the idea of separating online and

[658] Martos, C. M. (2008). The Transformation of Intimacy and Privacy through Social Networking Sites. *Institute Of Communications Studies, University Of Leeds, Regulation & Governance*, *2*, 425-444.
[659] Tene, O., & Polonetsky, J. (2013). A theory of creepy: technology, privacy and shifting social norms. *Yale JL & Tech.*, *16*, 59.
[660] Zhao, S., Grasmuck, S., & Martin, J. (2008). Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in human behavior*, *24*(5), 1816-1836.

offline activity, as can be seen with the expansion of surveillance powers by various state such as the US with its Patriot Act[661] or the UK with its Investigatory Powers Bill[662].

A change in the social guarantee that informational privacy, regardless of whose privacy, has value and should be controlled by the individual, would have rippling effects on the rest of the Guarantees, especially Legal Guarantees, as the Law reflects the social norms of society. There is great debate as to the value of privacy and informational privacy and as to whether individuals actually care about their privacy or not. This debate is outside of the scope of this thesis, which instead continues to work under the assumption that informational privacy does have an important value, highlighted by the myriad efforts to create new Guarantees or reinforce existing ones whenever one disappears. Ethical behaviors are rewarded by consumer loyalty, which in itself is a Market Guarantee (if there is more profit to be made through establishing trust with consumers, then this ethical approach will be more common)[663]. In this way, though businesses remain amoral, they adopt Social Guarantees because they influence the Market Guarantees that truly motivate them.

As we can see, despite the fact that Social Guarantees are very subjective, influenced by hard-to-predict trends and outrages, dependent on a public with various amounts of technological literacy and a fickle attention span, they still have an important place and need to be taken into account. In particular, their power in relationships between individuals and their governments and between each other still have a huge importance, though it tends to be left by the wayside as it is not where the major upheavals in Guarantees are found: secrets between friends remain secrets no matter what year it is, and the Balance is (mostly) maintained.

V. Legal Guarantees

We have approached each type of Guarantee in turn. In this way, we have shown how despite what it might seem, the Guarantees are primarily made up of the existing social,

[661] Osher, S. Privacy, Computers and the Patriot Act: The Fourth Amendment Isn't Dead, But No One Will Insure It. *Fla. L. Rev.* 54 (2002): 521.
[662] Akdeniz, Y., Taylor, N., & Walker, C. (2001). Regulation of Investigatory Powers Act 2000 (1): Bigbrother. gov. uk: State surveillance in the age of information and rights,[2001]. *Criminal Law Review*, 73-90.
[663] McMurrian, R., & Washburn, J. H. (2008). Branding: a social contract between a business and its customer. In *Contemporary Thoughts on Corporate Branding and Corporate Identity Management* (pp. 5-22). Palgrave Macmillan UK.

economic, and technological environment, with Law only filling the gaps where Guarantees are imbalanced. However, as we have seen, Guarantees are becoming increasingly weaker, which increases the role of Legal Guarantees to restore the balance. The way in which Law is a control mechanism seems pretty clear, however the unique way in which Law is a multi-dimensional tool will be explored to show how it has the best ability to deal with specific and targeted issues.

We have gone over how restricting access to Information is a vital tool in ensuring informational privacy. For this, the Balance is effective at pointing out which data controllers obtain Information they did not have before, and which Guarantees were changed to lead to that situation. However, another factor in ensuring informational privacy is whether that Information is used. An example comes from a government's obtaining of massive amounts of data under the justification of national safety. With that data, powerful Information could be created, used for example to track protestors or discriminate against groups which hold certain characteristics or ideas. What limits informational privacy is that the legitimate justification for the government holding such data is strong: the government does need to protect its citizens, and does have a responsibility to deploy tools to further that goal. As such, Guarantees are naturally weakening as governments' ability to gather data improves, while any strengthening of Guarantees hits the obstacle of national security. This is a limitation of Legal Guarantees: Legal Guarantees have to balance different interests.

Protecting informational privacy means limiting either Information from being obtained, or the use of that Information to make decisions[664]. We will go over a few examples to show how Legal Guarantees can create various types of other Guarantees, or simply act as Guarantees in their own right:

- Technological Guarantee: A legally-mandated transparency tool allows users to know what profiles are constructed by the data controller and what products are recommended to them based on these profiles. Though the Information to create and use the profile is already held by the data controller, the Social Guarantee of being perceived as protecting and caring towards users makes it less likely that these uses will violate the informational privacy.
- Market Guarantee: Detailed profiles are created that snoop into very personal topics. They cost a lot of create due to legislation on data minimization and

[664] Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J. H., Metayer, D. L., Tirtea, R., & Schiffner, S. (2015). Privacy and Data Protection by Design-from policy to engineering. *arXiv preprint arXiv:1501.03726*.

consent, but are worth little as buyers are interested in targeting advertising and would rather get more data that is more relevant, even if less revealing. This lack of demand means that the profiles remain unused.

- Social Guarantee: The government uses data it obtained through intelligence services to create profiles on journalists to analyze those who are likely to be critical of the government. Thank to the (legally-enforced) freedom of the press, citizens are outraged and the program is cancelled. This is comparable to France's "SAFARI" system which tracked individuals using data without informing them, which led to a wide public backlash once it was revealed[665].
- Legal Guarantee - Privacy Law: A government employee is able to use data-gathering tools implemented by intelligence services to spy on their spouse (a practice so problematic it was given a name due to its use in the NSA - "LOVEINT"[666]). As stalking is an offense against the right to privacy, the employee is prosecuted, leading others to be more wary of their usage of the data.
- Legal Guarantee - Criminal Law: An employee at a discreet dating service finds out a client is using it to cheat on their spouse and considers blackmailing the client using that Information. As blackmail is illegal, the employee is too afraid to go through with this plan.

These are a few examples as to how to restrict usage and collection of Information using Legal Guarantees. They are noticeably more certain than previous Guarantees which restrict access to Information, as they can act as powerful deterrents.

Conclusion

Using the Guarantees as a measure for protecting informational privacy allows for a multidisciplinary approach, taking into account any factors which become relevant. Using Lawrence Lessig's taxonomy shows how those who control Information only act in ways which are dictated by the Guarantees, which emphasises how every type of Guarantee has an important impact.

---

[665] Burkert (n.4)

[666] Peterson, A. LOVEINT: When NSA officers use their spying power on love interests, *Washington Post,* August 24, 2013, accessed 23/5/2017 at https://www.washingtonpost.com/news/the-switch/wp/2013/08/24/loveint-when-nsa-officers-use-their-spying-power-on-love-interests/?utm_term=.6f11788480b8

Companies obtaining and sharing large amounts of personal data in a way that might lead to a large data breach or misuse of data would not do so if Market Guarantees made it unprofitable, if Social Guarantees made it unacceptable, if Legal Guarantees made it illegal, or if Technological Guarantees made it impossible.

In the same way, a company which does minimal, careful data processing might consider starting invasive Big Data processing activities if a lack of Market Guarantees makes it profitable, a lack of Social Guarantees makes it acceptable, a lack of Legal Guarantees makes it legal, or a lack of Technological Guarantees makes it possible.

In our last Chapter, we will establish how to find where the balance of Guarantees is equal and informational privacy is protected. We will also develop the ‘’Informational Privacy Toolbox’’: how the different Guarantees have been used, and can be used, to protect informational privacy despite the transformations brought about by Big Data. To this end, we will also study the various transformations that have taken place in the field of informational privacy, what Guarantees have been attempted to fix them, and how successful they have been, up to and including the GDPR itself.

Chapter 6. The Information/Guarantees Balance

In the last Chapter, we have gone over the essential elements of the “Guarantees”: various elements which work together to create the context in which Information is created, transferred, processed, and used. As we have shown, when the Guarantees are balanced, privacy is protected. Because of the “Sculptor’s Work” we discussed earlier in this thesis, the Guarantees change step by step based on who is holding the Information and how the Information was transformed.

The advantage of the Balance comes from the fact that one Guarantee of any type can technically replace any other, which allows for a plurality of solutions to the rising transformations surrounding informational privacy. But this thesis has a specific aim: to use the Balance to provide solutions for the changes in informational privacy. As we have shown, the biggest limitation in this apparatus comes from the “identified” requirement which does not account for the “Sculptor’s Work” or for the differences between “data” and “Information”, as well as the ‘’public/private’’ distinctions in privacy law. Another limitation is the focus of data protection on ‘’control’’ over personal information, despite having evidenced that meaningful ‘’control’’ can be difficult or impossible to obtain. These core issues are unable to deal with the age of Big Data, though there are hints of a more flexible approach being considered, as well as existing provisions on the “means likely to be used” for identification opening the way for such approaches. In this part, we bring together the strands developed over the course of this thesis - the core value of Informational Privacy as understood in the European conception, the limitations of the existing European regulation, the “Sculptor’s Work” and its implications, and the ever-present Guarantees underlying the entire informational privacy environment. We study how they interplay, and how solving the crisis of informational privacy can be achieved by balancing the Guarantees.

The balancing of the Guarantees itself will be approached with a certain goal in mind. The aim of this thesis is to lay out a set of tools - something we call the “Privacy Toolbox” - with which to influence the various Guarantees to rectify the Balance. These tools can be used by various entities - individuals, governments, companies - to influence the Guarantees. Detailing the tools for every type of entity would be the work of an entire book. Instead, the goal of this thesis is to lay out what actual effect on Information comes from various Legal Guarantees, and how to influence the other types of Guarantees

using Legal Guarantees, in the same way that tax legislation on alcohol or tobacco products are legal tools which create Market Guarantees. We will also devote some work on tools for individuals to protect their own privacy preferences, but will limit it to tools that can be provided, promoted, or incentivised by regulatory tools.

We will start our analysis of a practical Balance by attempting to find the default state of the Balance - a time and context where the Guarantees were even, allowing a “default” state from which the current situation can be assessed. This will be an example of an even Balance - it is likely that depending on the type of Information controller, the country, the industry, or a number of other factors a different time can be found for an ‘’even Balance’’, but this analysis will both show how we propose to establish an even Balance, and give a general view of where the Balance was relatively even.

A. Defining an Even Balance: A Historical Analysis

If the right to informational privacy depends on a balance between Information and the Guarantees, it is important to find what an even balance looks like. We have mentioned how the existing crisis in data protection comes from the fact that the Information-creating forces have grown at a speed that has not been met by a fast-enough rise in Legal Guarantees, but how imbalanced are these forces?

In order to establish what should constitute the proper equilibrium of forces, there needs to be a standard to hold as the “starting point”, a point where the Guarantees and the Information are properly balanced. The difficulty of course is that there is no clear single point where the Balance was unmoving for a long enough time that might establish a clear balance, at least not recently. The Guarantees and Information constantly shift, with increasing speed, and identifying a status quo is challenging.

We have established in Chapter 1 that a rise in Information that starts to get out of control is generally met with a rise of legal Guarantees, such as instant photography creating motivation for Warren and Brandeis to write their essay “the Right to Privacy”. We would argue that because of this dynamic, and its consistency, absence of imbalance can be found by the absence of the dynamic. In other words, if there was a time when there was no effort to put forward new Guarantees, that means the balance was even. In addition, as we are looking into the context of European law, only such a balance in the European context would be relevant. Lessons from the US or other legal systems are useful, but

the balance of forces is very different in the US than it is in the EU - we have gone over the differences in privacy conceptions and data protection regulation, but additionally the social context, market, and even the architecture are very different. As such, the basic balance has to be contextual by nature.

As such, in order to find the closest thing to an even balance to base the assessment of the current situation on, the goal is to identify the most recent time the legal Guarantees remained unchanged for a significant period of time, in the European context.

Going back to our Chapter 2 study of the history of EU Data Protection regulation, the first data protection studies were started in the later 1960s, with the Council of Europe's Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data[667] being signed in 1981. This Convention was a relatively weak instrument, explicitly mentioning the right to privacy; its Article 1 stating that "The purpose of this Convention is to secure in the territory of each Party for every individual, whatever his nationality or residence, respect for his rights and fundamental freedoms, and in particular his right to privacy, with regard to automatic processing of personal data relating to him"[668]. It had already started to use the "identifiability" criteria (showing that the criteria is almost 40 years old, a possible factor for its increasing obsoleteness). After the Convention came a number of increasingly-stronger provisions, always responding to a rising problem. However, as we have shown with criticism of the Data Protection Directive and General Data Protection Regulation, these developments failed to re-establish the balance.

Meanwhile, looking at the evolution in that period, not of data protection but of privacy law, an interesting observation can be made. Instead of any speedy development to deal with new ways of infringing on privacy, privacy law itself actually moved very little[669]. In fact, most privacy cases brought to the European Court of Human Rights under Article 8 which produced significant jurisprudence, dealt with issues related to data protection[670], with some issues related to the reasonable expectation of privacy and government intrusion. The fact that outside of data protection there has been comparatively limited development of privacy legislation, under our hypothesis, is a sign that the balance has

[667] Council of Europe, Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, Jan. 28, 1981, Eur. TS. No. 108
[668] Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, Article 1
[669] Kokott, J., and Sobotta, C.. The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR. *International Data Privacy Law* 3.4 (2013): 222-228.
[670] Ibid

shifted little in these areas, and only in the area of informational privacy - which overlaps with data protection - has there been a massively significant shift.

This is interesting because if the modern privacy issues brought about by technology can be categorised almost entirely inside the purview of informational privacy, then this means that this area is where the imbalance in the “Information/Guarantees Balance” has appeared. It would mean that tracing the most recent point of an even balance would be when informational privacy concerns started to appear.

We have mentioned in our section on the patterns of privacy throughout history in Chapter 1 that using data to make decisions has always existed. Information has always been created through data. The question is not when that process started, but when it started to expand and create an imbalance against the other Guarantees that caused a Legal Guarantee response.

An additional challenge in Europe is that the context is different from country to country. Nevertheless, the ECHR and other EU conventions have allowed for a (relative) alignment of the conceptions of privacy in the various European legal systems, and if anything the number of countries only gives us more evidence as to where various initiatives have gained traction. Of course, with these conceptions, the most recent even balance will change from Member State to Member State (as the Guarantees will change between them and as such so will the point at which data protection becomes necessary). Nevertheless, we will be looking at the general European context, which will lead to a less precise answer, but still give an understanding of the point at which new provisions in the protection of informational privacy (which took the form of data protection) became necessary in the EU.

The very first true data protection law in Europe was passed by Germany in 1970. The Hesse Data Protection Act[671] was a law attempting to standardise local uses of these new computing technologies, but not a true national law[672]. The fear of local communities was the centralization of power through the means of these new technologies. Concerns over the power of the state and the confidentiality of personal information became prominent enough to set the law into motion. The Legal Guarantees it put into place by that law were quite weak - basic access and correction rights and possible actions for

[671] Hessisches Datenschutzgesetz (The Hesse Data Protection Act), Gesetz und Verordungsblatt I(1970), 625.
[672] Burkert (n.4)

unlawful processing - but were still a breakthrough at the time[673]. This is also an interesting perspective when thinking of who the law was targeting: the central government, and not companies or individuals, because it was the government which first had the means to carry out this kind of collection. A Technological and Market Guarantee - the central government not having the ability to do that processing - disappeared, leading to the need for a law dealing specifically with that imbalance without targeting companies.

It was Sweden that passed the first true national-level data protection legislation, which was because of its unique situation - Sweden had for decades a personal identification number system, ahead of other countries. The rise of computerised filing systems, combined with that existing system, created too problematic a combination and led to Sweden's early data protection law[674].

The French data protection laws were, similarly, motivated by the development of a government-led digital filing system, in this case the Interior Ministry's "SAFARI" system (named after an activity in which individual animals in the wild are tracked and hunted, which makes, if anything, for an ominous name)[675]. As in Sweden, this project relied on an existing database of citizen data. More ominously, the project was kept secret until it was exposed by an article in Le Monde in 1974, whose writer later became the first president of the French Data Protection Authority, the CNIL (Commission nationale de l'informatique et des libertés). Meanwhile, the French administration had started a similar program to track the personal information of children from their birth throughout their schooling to keep track of their medical data. The discovery of these projects caused the rapid development of the French data protection bill, which went into effect in 1980, giving the CNIL great powers[676].

Finally, other countries such as the UK only introduced such legislation when trading of data with countries that already had them became problematic - which led eventually to harmonisation efforts including the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, the Data Protection Directive, and most recently the General Data Protection Regulation.

[673] Ibid
[674] Ibid
[675] Ibid
[676] Ibid

This shows an interesting pattern - the countries that were first to create data protection regulation did so not just because a new type of technology - large-scale computing databases - became available, but because that technology combined with an existing architecture that could make use of it to create very powerful Information. This shows the interplay between factors that might be unexpected (the fact that Sweden and France happened to implement the data-gathering programs that rendered their later projects possible) which creates an imbalance.

Where does that leave us in deciding the “point of balance”? We would argue that we can go back to the flexible “means likely to be used” test. In this case, the goal is finding the point at which the means likely to be used by data controllers in order to create Information become imbalanced. As such, even if the various Member States that only added data protection legislation later did not actually implement data-collecting systems such as those found in France and Sweden, the first time they had the means to do so would be the moment when the balance became uneven.

In conclusion, we can track the upsetting of the balance in the EU, and as such the “default” state of the Balance of informational privacy to a specific time. That time is when the development of computing technologies allowed all governments the means to gather disproportionate amounts of data on their citizens (followed later on by companies, which had less access to data and fewer resources at their disposal compared to governments). That time, as seen by the sudden appearance of data protection legislation in EU Member States, is the mid-1970s.

This is relevant because it is not just the Technological Guarantees that changed in this period. Social, Market, and even Legal Guarantees beyond the scope of EU law, have all changed in unexpected way, and re-establishing a balance will need to take into account these changes.

Nevertheless, a task that seems insurmountable can be made much easier by the fact that many of these factors are relatively straightforward. The existing Legal Guarantees of data protection and privacy law, and their limitations, have already been discussed above. The Social Guarantees, as we have mentioned, are relatively weak due to the nature of businesses as amoral entities. And the Market Guarantees are the easiest to quantify, as a simple currency value can be attached to them. It is mostly the interplay between the Technological Guarantees and the consideration of new forms of Guarantees to make up for the increased level of Information which will need to be taken

into account to re-establish the balance. In other words, how can the Law incentivise Information-creation through affecting primarily Market Guarantees?

The Balance is now in place. Its goals, based on the interests of informational privacy, are set. Its target, Information, is identified. Its tools, the Guarantees, are laid out. And the point at which the Balance is even has been found. Now that the main framework is laid out, we will analyse a variety of tools that have been developed to deal with certain challenges related to this balance, and explain how we should look to fix the core imbalances.

B. The Privacy Toolbox: Guarantees from Different Disciplines

There are many tools available, which can be used by companies, governments or individuals in order to change the balance - either imbalancing or balancing it. Something that is important to note is that in some instances an imbalance - or a breach in the right to data protection - will be justified because of other societal interests being considered more important. A basic example is public surveillance measures being justified by public and national security. The GDPR provides for a variety of such situations in its Article 23, including national security, defence, public security, public health and social security[677], allowed based on the ''necessary and proportionate measure in a democratic society'' test used in EU law to justify breaches by the State of human rights[678]. On these occasions the balance cannot be even, but using tools to get close is still an objective worth pursuing. Nevertheless, this analysis will focus on instances where the right to data protection is not limited by another right. This analysis will comprise a combination of tools, most of which already exist, including some measures already in the GDPR: as we have stated throughout this thesis, despite its faults the GDPR is the best data protection instrument developed so far.

[677] GDPR, Article 23

[678] Büschel, I., Mehdi, R., Cammilleri, A., Marzouki, Y., & Elger, B. (2014). Protecting human health and security in digital Europe: how to deal with the "Privacy Paradox"?. *Science and engineering ethics*, *20*(3), 639-658.

In short, there are two levels of action which can be undertaken to protect the balance. One, is to ensure that controllers do not obtain the Information which gives them the power to upset the balance. Two, is to ensure that controllers who have obtained that Information, do not use it for any purposes which may unduly impact individuals. In the world of Big Data, the first aspect is difficult to maintain. This is why I argue that the second is the most important, which is where the various Guarantees come into full effect.

Before going over these actions and how existing efforts regarding the various types of Guarantees either exist or can be put into place, a mention needs to be made of security, and its place in this framework.

Security is recognised as being of paramount importance in the GDPR, being the sixth principle of the Regulation[679], and is a centerpiece of the ''risk-based'' approach to data protection. Security in the GDPR is directly tied to risk, as defined under Article 32: "Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk."[680]. As we studied earlier on in this thesis, ''risk'' is very hard to estimate. In the context of the Information/Guarantees balance, however, the reason why data has to be kept secure is because once it is lost, and either released to the public or obtained by an unknown third party, the Balance is impossible to re-establish. Social Guarantees no longer apply, because those who obtained the data have no social obligation towards individuals. Legal Guarantees no longer apply, because finding and binding those individuals with legislative instruments is extremely difficult. Cyber-crime convictions rates are very low[681]. As such, legal instruments are unlikely to be a Guarantee against cyber-criminals using that data for whatever purpose they choose. Meanwhile, Market Guarantees become unknown, because blackmail and sell-offs of data become options, while Technological Guarantees also are an unknown, since the technological means of those holding the data are similarly unknown. In short, security is important because it ensures that only those who are subjected to quantifiable Guarantees hold Information.

[679] GDPR, Article 5
[680] GDPR, Article 32
[681] Smith, R., Grabosky, P., & Urbas, G. (2004). Cyber criminals on trial. *Criminal Justice Matters*, *58*(1), 22-23.

In order to limit the creation of Information, there are two core options. One is ensuring that there is not enough data to create it. The other is to create fake or inaccurate or incomplete data, in order to devalue the Information created from it. This can be done in various ways, some examples of which we will outline here based on the four types of Guarantees.

I. Technological Guarantees

Technologies aimed at allowing individuals to get control over their data are growing exponentially[682], as academics, engineers and businesses all innovate and develop new tools to allow control over information[683].

A Guarantee we have developed earlier in this thesis is removing parts of the data, in order to allow the data controller to obtain the Information they seek while minimising how much of that Information relates to individuals. Meanwhile, unlinkability (the ability to use various services without it being possible to link that usage to one natural person)[684] and unobservability (the ability to use a resource without others being able to tell it is being used)[685] are all techniques which can be used to limit the creation of Information.

Meanwhile, embedding tools which follow the principles of the GDPR - especially the third principle of data minimisation[686] and the fifth principle of data retention[687] - both can help with this development. Various tools, both internal to software and hardware (the ‘’privacy by design and by default’’ mandated by Article 25 of the Regulation[688]) can limit the creation of Information. Meanwhile, limiting Information to make sure that it cannot be shared even within the same organisation, using access controls, can have the same

[682] D'Acquisto, G., Domingo-Ferrer, J., Kikiras, P., Torra, V., de Montjoye, Y. A., & Bourka, A. (2015). Privacy by design in big data: an overview of privacy enhancing technologies in the era of big data analytics. *arXiv preprint arXiv:1512.06000*.
[683] Wright, D., Gutwirth, S., Friedewald, M., Vildjiounaite, E., & Punie, Y. (Eds.). (2008). *Safeguards in a world of ambient intelligence* (Vol. 1). Springer Science & Business Media.
[684] Pfitzmann, A., & Hansen, M. (2010). A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management.
[685] Ibid
[686] GDPR, Article 5
[687] GDPR, Article 5
[688] GDPR, Article 25

effect. There are many ways to do so - from physical locks to passwords and biometrics[689].

Some more complex tools can try to limit the use of data by using artificial intelligence tools. Detecting certain types of information and filtering them out, detecting possible links with existing data, adapting to local privacy standards[690], or using language translations to check whether some cross-language Information could be created. Ironically, sometimes it is necessary to create Information in order to know how not to hold it.

II. Social Guarantees

Social Guarantees see their core power in how social pressure can affect companies. Earlier on in this thesis, we explored the new principle introduced in the GDPR, the principle of ‘’accountability’’. This idea of holding companies accountable to certain social norms is growing to a greater degree, through ideas like ‘’Corporate Social Responsibility’’[691]. This increasing pressure, from employees, partners, and consumers all[692], ensures that efforts are made to fulfill those expectations.

A move towards this is what has been described as the ‘’creepiness factor’’[693] - beyond the legal obligations required for data protection, there are consequences to being seen as a corporation which does not pursue that goal. A simple example is becoming a target for data subject complaints for being seen as a problematic company. The GDPR’s Article 87 allows for an action for compensation for a breach of one’s rights, and class-action lawsuits for breaches of privacy and data protection are developing in the European context[694].

Avoiding the collection and processing of data for unexpected or problematic purposes is not just a benefit to avoid getting fined or sued. Obtaining the trust of partners allows for business advantages. This can be done through standards such as the various ISO

[689] Sagiroglu (n.294)
[690] Wright et al. (n.683)
[691] Jones et al (n.205)
[692] McGlone (n.208)
[693] Leonard, P. (2013). Customer data analytics: privacy settings for ‘Big Data’business. *International Data Privacy Law*, *4*(1), 53-68.
[694] Sapronov, W., & Srouji, J. (2017). Class Consciousness: Class Action Arbitration under US and EU Privacy Laws. *YB on Int'l Arb.*, *5*, 83.

standards based around data security (ISO 27001, ISO 22301, ISO 27005 and more)[695] in order to show accountability and demonstrate that the trust of partners is well-placed.

Meanwhile, trust marks and seals are safeguards of increasing importance. They are guarantees provided for by independent entities to gain the trust of consumers[696]. Some of these efforts are driven by the industry, but some are pushed by consumer-protection agencies, with various initiatives by academics or associations developing these tools[697].

In greater scope, educating data subjects on what data processing can involve has an impact on what they find acceptable or not. From initiatives for technological literacy to poignant science-fiction making individuals aware of possible futures, there are many ways of increasing awareness, leading to data controllers having to be transparent and accountable. That is one of the reasons why part of the first principle of data processing in the GDPR is ''transparency''[698] - a legal tool aimed at reinforcing Social Guarantees.

III. Market Guarantees

While Social Guarantees have a strong impact on incentivising controllers, acting on their ability to obtain financial gains based on Information is also important. Data has been called ''the oil of the information age''[699]. As such, any efforts to limit the value that Information has will have the effect of limiting incentives to obtain it. Consumer-led efforts to devalue data come about through various means, one of the most common ones being either removing themselves from data processing (such as using privacy-protective tools and software) or creating new, fake data which makes the Information resulting from it worthless. This ''noise'' creation can be done in a number of ways, from fake Google searches to entirely fake profiles designed to fool databases[700].

Another way proposed by some theorists has been to change the paradigm altogether, by having individuals sell their data (and control that sale) directly. Many names have

[695] Tankard, C. (2016). What the GDPR means for businesses. *Network Security*, *2016*(6), 5-8.
[696] Wright et al. (n.683)
[697] Hall, T. S. (2014). The Quantified Self Movement: Legal Challenges and Benefits of Personal Biometric Data Tracking. *Akron Intell. Prop. J.*, *7*, 27.
[698] GDPR, Article 5(1)
[699] Tene, O., & Polonetsky (n.8)
[700] Howe, D. C., & Nissenbaum, H. (2009). TrackMeNot: Resisting surveillance in web search. *Lessons from the Identity trail: Anonymity, privacy, and identity in a networked society*, *23*, 417-436.

been given to this idea, such as data vaults[701] or data marketplaces[702]. The idea remains the same - when services have been provided at the customer's expense so far, this would instead have individuals in charge of their own information, and able to sell it on to whoever they choose. Introducing this kind of tool may allow the value of data to be shared with data subjects as much as it is with data controllers.

Market Guarantees are powerful because they are relatively simple to influence through legislation. An example is the GDPR's heavy fines, up to 4% of the entity's annual global turnover or 20 million euro, whichever is higher[703]. That is a huge cost for an organisation, and it is more economically viable to protect data subjects - as being profit-driven is the core of any business, Guarantees which speak directly to that core are most effective. In fact, most Social Guarantees are only effective because they affect Market Guarantees, such as standards that companies are worried about being seen observing, or pro-consumer initiatives to avoid bad press or appease upset data subjects.

IV. Legal Guarantees

We have evoked the subject of Legal Guarantees in a few places, and the GDPR is nothing except the most extensive Legal Guarantee for informational privacy yet. The advantages of legal Guarantees can be seen all over this thesis, from companies changing their ways to avoid fines to enshrined standards verified by Data Protection Authorities and under their supervision[704].

A wide variety of such tools can be put into place to implement various Guarantees, from forcing data controllers to implement accountability measures (a core theme of the GDPR), which is a way to create Social Guarantees artificially. The GDPR mentions changing the entity's ''culture'' many times, and so works to change that culture by legal mandate.

---

[701] Shilton, K. (2009). Four billion little brothers?: Privacy, mobile phones, and ubiquitous data collection. *Communications of the ACM*, *52*(11), 48-53.
[702] Moiso, C., & Minerva, R. (2012, October). Towards a user-centric personal data ecosystem the role of the bank of individuals' data. In *Intelligence in Next Generation Networks (ICIN), 2012 16th International Conference on* (pp. 202-209). IEEE.
[703] GDPR, Article 83
[704] Grierson, J., PlayStation data hack: Sony fined £250,000 for 'preventable' breach, 24 January 2013, accessed on 28/9/2017at http://www.independent.co.uk/news/business/news/playstation-data-hack-sony-fined-250000-for-preventable-breach-8464651.html

Similarly, through fines and legal actions, through authorising transfers and codes of conduct, through setting norms and limitations, Market Guarantees are affected by the GDPR. Meanwhile, by creating the concepts of ''data protection by default'' and ''data protection by design'' enshrined in Article 27 of the GDPR[705], the GDPR is clearing the way for a host of Technological Guarantees to be embedded in organisations.

In fact, seeing how the GDPR is built, the use of every type of Guarantee is very apparent. But in fact, creating every tool possible to achieve the needed outcome is not a new concept - but rarely is it pushed to such an extent. The GDPR is built around Guarantees, using a vast array of them to accomplish its goals, and as we noted earlier in this thesis these Guarantees are mostly innovations of the GDPR not found in the Data Protection Directive.

In short, European data protection law is in a transition period, between the past - a prescriptive, rights-based, personal-data-based approach - and the Information/Guarantees Balance-based approach. This thesis outlines this change, and the road ahead to accomplish this change completely. This new approach, the ''Informational/Guarantees Balance'', is both a useful way forwards and one already being adopted by EU regulatory authorities, and recognising it would allow for a better way forwards to protect informational privacy.

We argue that the transformation towards this approach will and should continue, which involves primarily continuing to move away from consent as well as the notion of "identifiability", and continuing to move towards a conception of "risk" based on the types of processing and the capabilities of the data controller and not one based on "harm". Taking into account all applicable Guarantees as well as the Information-creating capabilities of the particular controller allows for a conception which creates a burden proportional to the abilities of the controller instead of one where every controller has the same burden no matter what the actual risk involved.

C. The Information/Guarantees Balance in Practice: Scenarios

[705] GDPR, Article 27

In order to complete the development of this approach which is developing in EU data protection law, we will go through three scenarios which will allow us better to highlight the advantages of such an approach. These cases will specifically be based on situations where the current tools to protect informational privacy have difficulty in doing so. However, the study will remain relatively high-level: matters such as cost-benefits analyses and the complexities of psychology or sociology will not receive extensive development. This is due to the fact that these case studies are primarily aimed at showing examples of the mindset involved in this approach. These case studies are:

Scenario One: The Neighborhood Watch

In this scenario, the data controller is a neighborhood watch in England. A new family has moved in for the first time in a decade, and members of the watch become convinced that learning as much as possible about these new members is vital to the neighborhood's peace of mind. Using a wide array of tools available to them, the watch will attempt to learn as much as they can about their neighbor. This will highlight the issue of privacy in public and the role of Social Guarantees

Scenario Two: The At-Home Carer Company

Here the controller is a British company which delivers at-home care services to elderly and often disabled individuals. In the course of their activity, carers obtain a huge amount of sensitive information, easily used for a variety of profitable purposes. Nevertheless, a number of effective Guarantees ensure that the client's' privacy is safeguarded. This will show how a company able to obtain a huge amount of very sensitive data may still be limited by powerful Guarantees.

Scenario Three: The Multi-Billion Group Profiling Company

This scenario involves a massive multi-billion dollar company, with access, through an aggregation of services and partner companies, to essentially all data gathered online as well as the financial means to create huge amounts of Information, in particular profiles in order to sell Information to a variety of corporations looking to target their products better. This will highlight the fact that even organisations with great Information-creating powers can still be under very significant Guarantees, but also how to justify and implement Guarantees linked to group profiling.

Scenario Four: The Self-Driving Car Company

Our final scenario will focus on the impact of new technologies on the existing legal framework. Self-driving cars are a growing technology which relies on obtaining a very large amount of data, collected in public spaces, analysed in a Big Data environment, in order to function. As we will see, this technology will prove to be a challenge to a binary-test view of data protection.

For each of these cases, we will go through the steps of the Information/Guarantees Balance. First, we will identify the even Balance, as well as the Guarantees that have changed since the last time that Balance was even. We will then identify where the imbalance lies, and what Guarantees can be put in place in order to ensure informational privacy. As we will see, the GDPR does provide, in the new approach we have shown it is heading towards, interesting new Guarantees which have a positive impact on the balance. However, we will also see how the challenges identified in this thesis impact these scenarios.

Scenario One: The Neighborhood Watch

What can one reasonably expect neighbors to know about oneself? Though that varies widely depending on context, in the UK there are signs pointing to the answer being ''not much''[706]. Social norms do not encourage neighbours to know each other particularly well. Some bonds can be created by necessity (asking someone to take care of a pet or hold a key, for example) which creates a level of trust. This trust is in itself a Social Guarantee: that bond gives some assurance that the person can be trusted with one's personal information (as well as belongings or pets). This means that there is a Social Guarantee in that trying to pry into one's neighbor's lives is not seen as normal or acceptable.

Meanwhile, privacy and data protection laws have established Legal Guarantees in that context. Privacy from neighbors has a long history in case law - most recently in the UK a case involving a dispute between neighbors showed where one's right to process data for ''purely personal or household activities'' was limited to monitoring non-public spaces[707], specifically in the case of CCTV monitoring.

[706] A study by social network Nextdoor showed that 60% of UK residents do not know their neighbors well or at all (BBC News (2017), Britons 'should know their neighbors', accessed on 4/10/2018 https://www.bbc.co.uk/news/uk-40811530)
[707] Woolley & Woolley v Akbar or Akram [2017] SC EDIN 7

As for Market Guarantees, the neighborhood will generally have limited financial means, which means only mass-market-available tools, skills and technologies are likely to be used, limiting the Information which may be gathered.

It is possible for the watch to observe comings and going in the house, possibly even following around the residents. However, this would require a significant time and cost investment, one unlikely to be available to a neighborhood watch. As such, without the addition of new technologies and tools, the Balance is relatively even. There is no absolute certainty that informational privacy will be respected, but sufficient Guarantees to ensure the risk can be tolerated.

There are a few Technological Guarantees which have seen a change in the last two decades, but the main one, especially in the UK, is the ease of availability of CCTV. Even though, as we have seen, CCTV capturing a neighbor's property is a breach of data protection law, it seems like this Legal Guarantee has not been wholly effective as disputes related to this type of data collecting are common[708]. From accusations of paedophilia if a camera captures children playing to stalking when zoom lenses are used, a variety of privacy invasions are now a common risk due to the ease of obtaining of CCTV. This change in Technological Guarantees had a wide impact which Legal Guarantees have not seemed to address effectively: making these practices unlawful is ineffective unless enforcement of that ban is consistently applied, which does not seem to be the case in this context.

As such, we have identified the main source of imbalance: easily-obtained CCTV without sufficient Guarantees to stem it.

We saw that the core imbalance in the context of a neighborhood watch association was the presence of CCTV cameras in public, as well as the fact that existing Legal Guarantees on monitoring the premises surrounding one's home have not stopped this phenomenon. As such, limiting the Information which can be obtained using that technology is the main way of reestablishing a Balance.

[708] Moss, E. (2014), Neighbourhood watch: how domestic CCTV is sweeping the UK, The Guardian, accessed on 4/10/2018 at: https://www.theguardian.com/world/2014/dec/19/neighbourhood-watch-domestic-cctv-sweeping-uk

Legal Guarantees are already in place, but not entirely effective as shown by the fact that the problem persists despite their presence[709]. As such, there are two ways in which Legal Guarantees could be used: either the existing ones could be enforced more consistently, for example with systematic fines imposed on those who breach the GDPR by ''monitoring data subjects in the Union''[710] through home-based CCTV, or new laws and regulations could be put into place. This could include making home-based CCTV altogether unlawful, or requiring its installation to be made by an approved entity or professional who would be responsible for the CCTV not capturing anything beyond the premises of the individual's property.

Social Guarantees could be applied, such as by a public awareness campaign on the importance of privacy in the home[711]. Public awareness campaigns have a mixed result in a number of fields[712] but have nevertheless seen many uses, from environmental issues such as recycling to wearing seatbelts.

Market Guarantees could consist of making it less affordable to place such CCTV. This can be done through taxing CCTV equipment, for example.

Finally, Technological Guarantees can also be applied. CCTV systems which can be easily set up to block or blur part of the field of vision to make protecting informational privacy easier. Meanwhile, individuals can set up basic Technological Guarantees such as a higher fence or drawn curtains.

Any of those measures, or a combination of them, could be effective in limiting the Information which can be obtained by the neighborhood watch association.

Scenario Two: The At-Home Carer Company

The carer industry is a growing market - as the population throughout the western world trends toward an older average age as well as more disabilities due to obesity[713], more professional carers are needed than ever. The job of a carer involves obtaining data

[709] Ibid
[710] GDPR, Article 3 on the scope of the GDPR
[711] Wilson, D., & Sutton, A. (2004). Watched over or over-watched? Open street CCTV in Australia. *Australian & New Zealand Journal of Criminology*, *37*(2), 211-230.
[712] Fletcher, A., McCulloch, K., Baulk, S. D., & Dawson, D. (2005). Countermeasures to driver fatigue: a review of public awareness campaigns and legal approaches. *Australian and New Zealand Journal of Public Health*, *29*(5), 471-476.
[713] Wang, Y. C., McPherson, K., Marsh, T., Gortmaker, S. L., & Brown, M. (2011). Health and economic burden of the projected obesity trends in the USA and the UK. *The Lancet*, *378*(9793), 815-825.

about individuals which could not be more sensitive - medication, health, intimate private information about the patient's daily life, but also information that is necessarily gleaned by spending a significant amount of time performing task for someone in their home, from family and friends to financial information. This puts the carer, as a data controller, in a position of possessing data about an individual which even Google could not be able to match. Without the right Guarantees, what is to stop that data from being used for wrongful purposes and breaching informational privacy?

Social Guarantees tend to be weaker in a business-to-consumer context than in a personal one, though trust and reputation are an important part of any healthcare-related organisation, whether for-profit or not[714]. Meanwhile, codes of conduct and industry regulations are in place to protect the elderly[715]. As such, as far as the company itself is concerned, there are very strong Social and Market Guarantees (reputational damage and loss of trust) against breaching informational privacy. Additionally, the care sector is heavily regulated in and of itself, adding Legal Guarantees.

Despite these Guarantees however, new challenges have appeared because of the rise of the Information Age. Information is now centralised by these organisations and accessible by a broader set of individuals than before, and new powerful tools are used to monitor patients. The main issue comes therefore from the confidentiality and accessibility of that information: carer companies have extremely sensitive data, and though they still have no real motivation for infringing their patient's informational privacy anyone who would get their hands on that Information could. This shows us that in this context, restoring the Balance is primarily a question of information security and accessibility.

In this case, the data possessed by the organisation is extensive. However, it is unlikely that powerful Information would be created in a way which would endanger the informational privacy of the data subjects, instead of being created for their continued well-being. A carer company is under very strong Legal and Social Guarantees, as well as having limited means which constitute Market and Technological Guarantees. Thus, the core danger to individuals' informational privacy is that data is being obtained by entities or individuals that are not subject to these Guarantees. This means that both ensuring security, and regulating transfers of data to third parties, are the core way of protecting informational privacy in such a context.

---

[714] Shore, D. A. (2005). *The trust prescription for healthcare: Building your reputation with consumers*. Health Administration Press.
[715] Hughes, M. (2004). Privacy in aged care. *Australasian Journal on Ageing*, *23*(3), 110-114.

The GDPR already provides Legal Guarantees in this context, by requiring a level of security appropriate to the ''risk'' to data subjects. As we saw earlier in this thesis, the notion of ''risk'' is itself based on what Information can be obtained with that data. If the GDPR is effectively applied, Legal Guarantees should ensure the carer company will protect the data effectively. Other Legal Guarantees could be applied, such as prosecuting particularly severely employees who lose or steal patient data, or legally requiring certain measures to be taken to avoid data breaches.

Social Guarantees are already present in the carer-patient relationship, but awareness campaigns and pressure from patients and their families could have a positive impact.

Market Guarantees are also present - the GDPR prevents data from being used for purposes others from the ones for which it has been collected, including selling it for a profit[716]. Additionally, carer companies have generally limited means, and as such are unlikely to create very powerful Information. Beyond government-mandated measures, self-regulation by the industry can be developed, such as standards and codes of conduct aimed at obtaining and maintaining the trust of clients to protect the industry as a whole. This ''trust'' element links back to Social Guarantees as well.

Finally, Technological Guarantees can be put in place. The GDPR's preventive approach, which imposes a fine on data controllers for not having the right security measures in place even where no data breach has taken place, pushes controllers to implement Technological Guarantees. Measures implemented to ensure carers do not lose data given to them by clients can include encryption or anonymisation, or simply moving from a paper format to data hosted in a digital format.

Overall, this scenario shows that carer companies are already under very strong Guarantees, and thus may not need to be subjected to the rigorous standards of GDPR compliance that the data it collects may require. However, when it comes to security and accessibility of data, Guarantees need to be put into place.

Scenario Three: The Multi-Billion Group Profiling Company

[716] GDPR, Article 5(1)(b) on the purpose limitation principle

The ability for companies such as Google or Microsoft to aggregate data through their various services is one of the core phenomena that characterises the ''Big Data'' age. However, when data is aggregated and then turned into a group profile, that profile is not subjected to the GDPR - this is because the data is not ''identifiable'', and thus is not personal data despite having an impact on data subjects (thus "relating to" individuals due to the "result" element of the processing). Though data protection as a Legal Guarantee is mostly not present, other rights do apply, particularly the right to non-discrimination[717].

It was always possible, through research, to create group profiles. However, the profiles created by large data aggregators have much more data feeding into them than ever before, and tools allowing better Information, which requires appropriate Guarantees.

Nevertheless, there are very few other Guarantees on these controllers. Social Guarantees are limited by the fact that most individuals are not aware of the phenomenon, or at least not informed in a significant way[718]. Meanwhile, the profit which can be generated from that type of activity is so high that there are few Market Guarantees holding such an organisation back. The only limit would seem to be the Technological Guarantees involved with how much data and Information can be obtained[719].

This lack of sufficient Guarantees despite the Information which can be created and its resulting impact on data subjects is one of the limitations of the "personal data"-focused approach we have identified earlier in this thesis. This has led to a variety of organisations such as Google relying on group profiling for a number of purposes, freely obtaining Information which may impact individuals but is not covered by the Legal Guarantee that is the GDPR.

Large group profiling organisations see an imbalance primarily due to the fact that the Information they create is not within scope of the GDPR and can be processed without any of the Legal Guarantees regulating other types of processing being applicable. We have shown how group profiling still ''relates'' to data subjects in the sense that the ''purpose'' element is fulfilled (the data is processed with the goal of influencing

[717] Schreurs et al (n.16)
[718] Ibid
[719] Taylor, L., Floridi, L., & Van der Sloot, B. (Eds.). (2016). *Group privacy: New challenges of data technologies* (Vol. 126). Springer.

individuals) and the "effect" element is achieved (the processing will impact individuals, what they see and how they are influenced). This shows a limitation of the "identifiability" focus of data protection law. Under an Information/Guarantees Balance approach, we can see that though the company only creates group profiles, these still constitute personal Information since they "relate" to data subjects. Group profiling is not inherently new (any shop owner will have identified patterns in their customer base) but its extent and sophistication of them have never been seen before. As such, what Guarantees can be put in place to restore the Balance?

A Legal Guarantee would be to include Information which relates to data subjects within the scope of the GDPR, even if done through a group profile. Beyond that, however, regulation over the use of those group profiles, especially in areas connected to special categories of data, would limit the use of that Information. The GDPR already includes a right for individuals not to be subjected to "automated individual decision-making, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."[720] which specifically forbids automated decisions based on special categories of personal data[721]. If that right can be applied to group profiling, then that would constitute a possible Legal Guarantee.

Meanwhile, Social Guarantees such as pressure on an organisation to avoid unwanted impacts of data subjects have been seen, for instance, the public outcry following American company Target identifying a teen girl's pregnancy before her parents did[722].

Market Guarantees are limited, since there are huge profits to be made from that kind of analytics. Efforts to add noise to the data, in order to weaken the value of the group profile, have been attempted by the development of a number of tools such as Helen Nissenbaum's TrackMeNot tool[723]. This would potentially limit the incentives for the use of those profiles.

Technological Guarantees against group profiling can take multiple forms. Ensuring that the decisions of these algorithms can be understood and explained would allow more control over them by individuals[724], but the rise of machine learning is making it difficult for even the creators of those tools to understand their decisions: whether this

[720] GDPR, Article 22(1)
[721] GDPR, Article 22(4)
[722] Hill, K. (2012). How Target figured out a teen girl was pregnant before her father did. *Forbes, Inc*.
[723] Howe (n.700)
[724] Goodman (n.28)

explanation is even possible long-term is up for debate[725]. Beyond artificially making the profiles less effective through adding noise, stopping the creation of these group profiles is unlikely: the economic value involved, as well as the fact that the data is out there, available and easily-processed, means that the Guarantees against these profiles being created in the first place are very limited.

In conclusion, this shows that the likely most effective area to focus on for the establishment of Guarantees is not the creation of those profiles, but the purpose of their use.

Scenario Four: The Self-Driving Car Company

The core of the challenge in this scenario is that in order for self-driving cars to be an effective technology, it requires a huge amount of data to be processed into useful Information. A large amount of the data collected by these cars relates to individuals, especially the cameras surrounding the car which are constantly monitoring the area surrounding the vehicle, including passersby or other cars[726].

The issue here comes from the limitations of the GDPR when it comes to that kind of data processing. Capturing someone on video may not always be "personal data" (since it is not always possible to "identify" the individual) but as the technology becomes better, video will inevitably become personal data as tools to identify individuals using that data improve. For example, according to Techrunch, Chinese authorities now have the ability to identify individuals using "smart specs" which could "identify individuals based on their body shape and the way they walk."[727]

Eventually, the Technological Guarantees against video data leading to personal Information will disappear and all self-driving cars will be, by default, collecting not only personal data but "biometric" data, defined in Article 4 of the GDPR as "personal data resulting from specific technical processing relating to the physical, physiological or

[725] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.
[726] Surden, H., & Williams, M. A. (2016). Technological opacity, predictability, and self-driving cars. *Cardozo L. Rev.*, *38*, 121.
[727] Russell, J. (2018), China can apparently now identify citizens based on the way they walk, Techcrunch, accessed at https://techcrunch.com/2018/11/07/china-can-apparently-now-identify-citizens-based-on-the-way-they-walk/ on 09/11/2018

behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data;"[728].

Even though the self-driving car company may not have a particular interest in creating biometric data (being able to uniquely identify passerbys through facial or gait recognition is unlikely to be useful for the purposes of the self-driving car's operation), anyone holding that data would have the ability to apply tools to do so.

Nevertheless, collecting personal data without a privacy notice is in breach of the first principle of the GDPR. Article 5(1)(a) of the GDPR provides that personal data shall be "processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');". The "fairness" aspect of that principle consists of a privacy notice provided to the data subject, which cannot be done in the case of a self-driving car. The third principle, "data minimisation", is also unlikely to be respected: the way these self-driving cars improve as a technology is by obtaining a huge amount of data, and then sifting through it to find useful Information. This is the exact opposite of the principle of ''data protection by default'' prescribed under Article 25 of the GDPR[729].

As such, the GDPR makes it so that self-driving cars would not be compatible with the GDPR, putting a very strong Legal Guarantee against them, to offset the disappearing Technological Guarantees.

On the other hand, there are very strong, existing Guarantees for self-driving cars - in particular, Market Guarantees. Self-driving car companies do not need much non-anonymised data to develop their utility, and the means ''likely'' to be used to identify individuals in a way which may breach informational privacy are quite limited. As such, despite the potential to create powerful Information, in practice that is unlikely to be the case.

This leads to a situation where data protection may impact innovation and the development of new technologies. As we will see, though this does make for an even Balance, there may be ways of doing so which do not have the same negative outcomes. This is another goal of the Information/Guarantees Balance: protecting Informational privacy while attempting to limit the negative impact of implementing Guarantees on other societal goals.

[728] GDPR, Article 4
[729] GDPR, Article 25

The core challenge in the context of self-driving cars is attempting to reach an even Balance without downright outlawing the technology involved. In other words, attempting to put sufficient Guarantees in place, despite the fact that the technology involves a large-scale gathering of public monitoring, while still allowing the technology to be used.

The GDPR already punishes data being used for purposes beyond the ones for which it was collected (purpose limitation principle)[730]. If that is applied effectively, the data should not be used for any purposes beyond the safety and efficiency of the self-driving car. However, the "legitimate interests" legal basis and its wide application may involve the data being used for other purposes which may impact informational privacy.

Social Guarantees could have a powerful impact, as we could see with Google Glass, a technology which also involved gathering data in public with cameras which could be on at all times. A powerful backlash resulted, in part because of the privacy implications of being monitored at all times (restaurants and other establishments banned Google Glass users from their establishments) leading to the product eventually being abandoned[731]. A similar backlash against the monitoring capabilities of self-driving car could have a similarly chilling effect. In this situation, a Social Guarantee would directly impact a Market Guarantee - the backlash making the technology less profitable.

A Technological Guarantee would be applicable if the self-driving car company found a way to improve their product which did not require the processing of personal data. For example, LIDAR, another data-gathering tool used by self-driving cars, can track objects but is as of yet unable to identify individuals.[732]

We have nevertheless seen throughout this thesis that attempting to stop the collection of data is unlikely to be successful in the long term. Instead, focusing on the later stages of processing may be more effective. As such, Legal Guarantees focusing on the later stages of processing are more likely to be effective, while both allowing the technology to be put into place and protecting informational privacy. This can be done by ensuring that the data is not used for any purpose beyond that for which it was collected, through strict access controls and personnel trained in confidentiality. Strict contractual obligations can be placed on any third parties that have access to any Information - even

[730] GDPR, Article 5(1)(b)
[731] Woodside, J. M. (2015, March). Wearable technology acceptance model: Google Glass. In *Society for Information Technology & Teacher Education International Conference* (pp. 1800-1802). Association for the Advancement of Computing in Education (AACE).
[732] Bloom, C., Tan, J., Ramjohn, J., & Bauer, L. (2017, July). Self-driving cars and data collection: Privacy perceptions of networked autonomous vehicles. In *Symposium on Usable Privacy and Security (SOUPS)*.

if it is made up of anonymised personal data. Organisations processing such data can be obligated to have stronger security (already established in the GDPR which bases the "security" principle on "risk"[733]) or to show additional evidence that the principles of the GDPR are being followed.

In the same way, boycotts of the self-driving car company itself, but also of companies buying and using that kind of data beyond the purposes of safety can lead to that data not being worth selling - boycotts as a Market Guarantee have been successful in a number of areas[734]. Finally, Technological Guarantees which constitute stronger security and access control measures can also have an impact on the availability of that Information.

Overall, once again we can see that focusing on the collection stage and attempting to filter what data is obtained by large organisations risk being both ineffective and hampering innovation (leading to inevitable non-compliances by opportunistic companies). What tends to be more effective is limiting the purposes for which that data is used, and the access to the Information produced by that organisation.

In this Chapter, we have identified a way to assess where imbalances appear (mainly through finding where a new technological innovation has appeared and what Information it has allowed to be created).

[733] GDPR, Article 5(1)(f)

[734] Pruitt, S. W., & Friedman, M. (1986). Determining the effectiveness of consumer boycotts: A stock price analysis of their impact on corporate targets. *Journal of Consumer policy*, *9*(4), 375-387.

Conclusion

Throughout this thesis, two core, interlinked propositions have been made as to how to conceptualise the interests of informational privacy. The first, is that ‘’personal data’’, based around data and whether it is ‘’identifiable’’, and the ‘’public/private distinction’’, based around a typology of physical and virtual places, should move towards ‘’Information’’ and whether it simply ‘’relates’’ to the individual. The second proposition is that, instead of focusing on rules-heavy regulation based on a binary approach, a Guarantees-based, contextual approach should be the goal, taking into account all of the factors permitting and restricting the ability to create Information.

In order to summarize what was achieved in this thesis, I will go over the research questions laid out in the Introduction, and show how they were answered.

The first question we asked was: ‘’what are ‘privacy’ and ‘data protection’ and what interests are these rights trying to protect?’’ Throughout this thesis, the muddled definitions of ‘’privacy’’ and ‘’personal data’’ were outlined. Privacy is hard to pin down, and many attempts to define it have either failed or ended up with such a long list of interests that it started to overlap with a plethora of other rights. As such we made the argument that focusing on where it overlaps with data protection, and identifying that overlap as the scope of this thesis, was necessary, otherwise this approach would also fall to the same limitations of getting lost in those various interests. In particular, data protection is expanding into various areas, from discrimination to consumer welfare, and as such we defined that particular interest which needs to be protected as ‘’informational privacy’’ - where both rights overlap into one area. Importantly, this overlap is not accidental; we made the argument that most of the new threats to privacy, brought on by

the Information Age, are covered by data protection, leading to a natural overlap between the two: ‘’informational privacy’’. Privacy as a whole is not under particular new threats from the Big Data age, except where ‘’informational privacy’’ is. There may be other interests that are protected by ‘’informational privacy’’, but trying to ascertain them is going back into the impossibilities of defining privacy.

As such, instead, we focus this thesis on a particular interest of data protection, which overlaps with a particular interest of privacy, without attempting to define why those interests are important. This is one of the innovations of creating data protection as a separate right - it allows a protection that is not dependant on defining exactly which interests need to be protected. The approach developed in this thesis supports that view, and bases ‘’informational privacy’’ on it. We also identified that the protection of informational privacy is far from new; in fact a pattern emerged: each time a new social or technological paradigm shift changed the way information was created and propagated, other forces stepped in the fill the gap. The fact that this is the natural pattern so far demonstrates that seeing this right as a balance of forces is the logical next step.

The question continues - how are these rights attempting to protect ‘’informational privacy’’? To answer that, we explored the uniquely European right of data protection as a separate right, as well as the doctrine of ‘’reasonable expectations of privacy’’ in European law. These innovations came from the need to deal with new problems without trying to tie them to old solutions. By way of contrast, the American approach was shown, which created four torts over a century ago, and has not been able fully to move past them since. Meanwhile, the EU created the right to data protection, which has seen some amount of success. Our analysis identified a pattern - almost all of the innovations that the General Data Protection Regulation brought to the data protection apparatus act on forces beyond purely legal ones. This lays the groundwork for the findings we make later on in the thesis, and shows that as Big Data transforms our world, traditional conceptions of how to protect informational privacy have to be left behind, and a system addressing the fundamental forces behind the creation and dissemination of Information is necessary.

Together, this allowed for the setting up of the core of this thesis: there is both a need for a system which acknowledges all the different forces that may impact Information, and an actual implementation of some elements of that system in the GDPR. Data protection is already headed towards that approach, without having identified it (but certainly being influenced by the approaches which have influenced this thesis, such as Helen Nissenbaum’s contextual integrity).

The next question was related to Big Data. How does it challenge the existing framework protecting informational privacy? Why is ‘’Big Data” such an important transformation? To study this, we studied the particular implications of Big Data, and the way it has transformed how we understand data and information. Though this is the latest in a series of transformations that have changed the balance of information creation and control, it is a very powerful transformation, which is changing some fundamental ways in which Information is created and disseminated. We argue that this transformation is one that requires such a massive overhaul of regulation that it has highlighted how much informational privacy is based around the balance of forces. However, we also argue that this is just one more change in the paradigm, like the advent of instant photography or tabloids. The biggest change in history, but the same phenomenon nonetheless.

With this in mind, the next step was understanding what is separating EU data protection from achieving this new approach. The core of this is ‘’identifiability’’: the fact that data is only ‘’personal’’ if it can identify a particular individual. This is increasingly ineffective, and we showed how two core pillars of the GDPR - consent and anonymisation - are undermined by this concept. Meanwhile, privacy law (and, by extension, informational privacy) depends on the idea of ‘’private’’ and ‘’public’’ places, which we showed should be replaced with a ‘’transparency/opacity’’ distinction, which takes into account new ways for Information to be created and disseminated. This links back, once again, to the Guarantees-based system developed in this thesis, and how it encompasses the whole of informational privacy - solving problems in how both data protection and privacy law have dealt with the issue.

This having been set up, we develop the alternative to ‘’personal data’’: ‘’Information’’. Can it answer the limitations of ‘’personal data’’? How can it be linked back to informational privacy in the way data personal data does? How can Information be regulated without the binary approach taken by the GDPR?

We showed the limitations of binary approaches, whether it is personal/non-personal, or private/public. This alternative focuses, not on the data and the context in which it was obtained, but on the knowledge and insights that are created using that data: ‘’Information’’. The focus on ‘’Information’’, as we have shown, allows us to avoid the difficulties of ‘’personal data’’, chief amongst them the fact that, depending on a number of factors, the same data can lead to incredibly different Information. This is because the addition of human intelligence, biases, and perspectives to the raw data creates something altogether new and with new implications, and recognizing this is vital. The

GDPR attempts to address this through the ''purpose limitation'' principle, attempting forcefully to stop data controllers from using the unexpected insights they may obtain. However, without acknowledging that those insights are something wholly new, there is a disconnect between the original data, and the final Information.

As we have shown, the idea of focusing not on the data itself but on the Information is not entirely new, and more and more evidence suggests, in the GDPR and elsewhere, that moving towards a model which takes it into account is the way forward. Despite the fact that ''Information'' as a separate concept does not exist in current data protection or privacy law, it is implicitly recognized in the GDPR.

We then showed how this Information can be linked back to the interests of individuals, through the ''relating to'' test already present when defining what is ''personal data''. We argue that though the ''identifiability'' element of that definition should be left behind, the ''related to'' aspect, which is slowly increasing in scope (including any decision which either is intended to affect an individual, or results in them being affected) is enough to link the Information to the individuals who need to be protected - resulting in ''personal Information''.

Answering this leads to the final part that remains to be answered: how to regulate that Information.

A core assertion of this thesis is that a balance of Guarantees both is replacing, and should replace, the current all-or-nothing approach. Before asserting this, however, we studied some core approaches that were used as alternatives. The two main ones considered - both of which heavily influencing this thesis - are Helen Nissenbaum's ''contextual integrity'' and the ''harms-based approach''. The first one revolutionized conceptions of privacy by focusing on flows of information; however this thesis attempts to place the focus on the Information itself instead of flows. This is primarily due to the ''Sculptor's Work'' identified when studying Information: data can be aggregated, processed, transformed, minimized, pseudonymized so much throughout its life that it may be something entirely different at the end of the flow compared to the beginning, and identifying what may be ''appropriate'' can be difficult. Moving away from contextual flows presents a challenge - because the Balance must have an objective grounding (unlike Nissenbaum's flows which can change over time); there must be a way to assess the ''default'' stage of the Balance. We have made the argument as to how to establish that, based on historical developments of privacy, but also based on the hypothesis

argued above that the Information Age is only another in a list of paradigm shifts, and that the human need for a private space has not changed.

These Guarantees, be they Legal, Technological, Market or Social, all have an important, oft-underestimated part to play in the creation and dissemination of Information, and ignoring it leads to the blind spots we have identified throughout this thesis. As such, we showed that the balance of Guarantees is the way to ensure the future of informational privacy.

Besides all the arguments made in my thesis as to the value and pertinence of this argument, the most important one is thus: all of the most significant innovations of the GDPR appear to be headed towards this approach. From the ''accountability'' principle implementing Social Guarantees as an integral part of the Regulation to the ''relating to'' aspect of personal data including ''purpose'' and ''result'' elements as part of what can make that data personal, EU privacy and data protection law is now stuck between the old and the new, and its pillars - especially the prescriptive, non-contextual approach, the ''identifiability'' criteria, and the ''public/private'' distinction - are preventing that approach from going further. As Big Data continues its transformation of society, acknowledging that approach and implementing it will be vital for the protection of individuals.

Through the journey of making its argument, this thesis has also done more than propose a new way of thinking about informational privacy. It has also reviewed the landscape of approaches that have been proposed, and shown their benefits but also their flaws. There have been many different conceptions, some harms-based, some rights-based, some interests-based[735]. As we have shown, every approach has its limits, namely that harms-based conceptions tend to have difficulty narrowing down what a ''harm'' is, while rights-based and interests-based conceptions tend to become quickly outdated and unable to deal with new technological developments. The Information/Guarantees Balance avoids these limitations - by limiting itself in scope to ''informational privacy'' and by looking at imbalances instead of harms - but comes with its own challenges, such as setting a default scale for the Balance and its limited scope. The argument has been made in this thesis that these challenges are solvable. Because European data

[735] Gellert, R. (2016). We Have Always Managed Risks in Data Protection Law: Understanding the Similarities and Differences Between the Rights-Based and the Risk-Based Approaches to Data Protection. *Eur. Data Prot. L. Rev., 2, 481*

protection regulation is heading towards this approach on its own, that time is fast approaching.

This leads to the conclusion of the primary research question: ‘’Is the EU protection of personal information shifting from an ‘’identifiability’’-focused framework to one based on a balance of Guarantees on those holding the means to create personal Information, and is that framework able to address the limitations of other conceptions as well as the challenges brought on by the age of Big Data?’’

The Information/Guarantees Balance builds on the work done both in the creation of the GDPR, as well as an array of theorists and policy-makers, to create a framework that is technology-neutral, applicable to the European conceptions on privacy and data protection, that protects the informational privacy of individuals without the limitations of European conceptions of privacy and data protection. It sets aside increasingly-meaningless distinctions like ‘’personal’’/’’non-personal data’’ and ‘’public’’/’private’’, instead going towards a wider understanding of Information and how to control how that Information is created, shared, and used.

This thesis only establishes this conception in a limited scope. ‘’Informational privacy’’ is but the small overlap of privacy and data protection, and there is a wide host of interests in both these rights that we have not had the opportunity to explore. Consumer welfare, for example, is an increasing question in data protection, for example with the creation of the right to data portability[736] which is ostensibly aimed at that interest and not privacy[737]. Other rights are being proposed to fulfil various interests, from the right to explanation[738] (to understand increasingly-machine-driven decisions), to the right to be de-referenced from search engines[739] (introduced with the Google Spain case[740] in 2014), to the right to informational self-determination[741] (a fundamental German right that offers a very unique approach to privacy as a social good for the promotion of democracy). As privacy and data protection discover new dimensions, the Information/Guarantees Balance may have value in protecting those as well.

[736] GDPR, Article 20
[737] Swire, P., & Lagos, Y. (2012). Why the right to data portability likely reduces consumer welfare: antitrust and privacy critique. *Md. L. Rev.*, *72*, 335.
[738] Goodman (n.28)
[739] Voss, W. (2016). After google Spain and Charlie Hebdo: the continuing evolution of European Union Data Privacy Law in a time of change.
[740] Case C-131/12, Google Spain SL v. Agencia Espanola de Proteccion de Datos (AEPD), 2014 E.C.R. 317
[741] Hornung (n.146)

Bibliography

Academic Journals and Books

Ackoff, R. L. (1999). Ackoff's best: His classic writings on management*. John Wiley & Sons.*

Acquisti, A., & Grossklags, J. (2005). Privacy and rationality in individual decision making. *IEEE Security & Privacy, 3(1), 26-33.*

Adam, B. (2006), et al. Privacy and contextual integrity: Framework and applications*. Security and Privacy, 2006 IEEE Symposium on. IEEE.*

Agrawal, R., & Srikant, R. (2000, May). Privacy-preserving data mining*. In ACM Sigmod Record (Vol. 29, No. 2, pp. 439-450). ACM.*

Akdeniz, Y., Taylor, N., & Walker, C. (2001). Regulation of Investigatory Powers Act 2000 (1): Bigbrother. gov.uk: State surveillance in the age of information and rights,[2001]. *Criminal Law Review, 73-90.*

Allen, A. L. (1999). Privacy-as-data control: Conceptual, practical, and moral limits of the paradigm. *Conn. L. Rev., 32, 861.*

Alo, E. R. (2013). EU Privacy Protection: A Step Towards Global Privacy. *Mich. St. Int'l L. Rev.*, *22*, 1095.

Altman, M., & McDonald, M. P. (2001). Choosing reliable statistical software. *PS: Political Science & Politics, 34(3), 681-687.*

Ambrose, M. L., & Ausloos, J. (2013). The right to be forgotten across the pond. *Journal of Information Policy, 3, 1-23.*

Ausloos, J. (2012). The 'right to be forgotten'–worth remembering?. *Computer Law & Security Review, 28(2), 143-152.*

Austin, L.(2005), Is Consent the Foundation of Fair Information Practices? Canada's Experience Under Pipeda. *56 University of Toronto Law Journal 181*

Austin, L. (2006). Reviewing pipeda: Control, privacy and the limits of fair information practices. *Can. Bus. LJ*, *44*, 21.

Austin, L. M. (2014), Enough About Me: Why Privacy is About Power, Not Consent (or Harm). Forthcoming in *Austin Sarat, ed., A World Without Privacy?: What Can/Should Law Do..* Available at SSRN: http://ssrn.com/abstract=2524512

Banisar, D., & Davies, S. (1999). Global trends in privacy protection: An international survey of privacy, data protection, and surveillance laws and developments. *J. Marshall J. Computer & Info. L., 18, 1.*

Barth, A., Datta, A., Mitchell, J. C., & Nissenbaum, H. (2006). Privacy and contextual integrity: Framework and applications. Proceedings - *IEEE Symposium on Security and Privacy, 2006, 184–198. doi:10.1109/SP.2006.32*

Baumann, C., Burton, S., Elliott, G., & Kehr, H. M. (2007). Prediction of attitude and behavioural intentions in retail banking*. International Journal of Bank Marketing, 25(2), 102-116.*

Baumer, D. L., Earp, J. B., & Poindexter, J. C. (2004). Internet privacy law: a comparison between the United States and the European Union. *Computers & Security*, *23*(5), 400-412.

Bayardo, R. J., & Agrawal, R. (2005, April). Data privacy through optimal k-anonymization. *In Data Engineering, 2005. ICDE 2005. IEEE.*

Bellovin, S. M., Hutchins, R. M., Jebara, T., & Zimmeck, S. (2013). When enough is enough: Location tracking, mosaic theory, and machine learning. *NYUJL & Liberty, 8, 556.*

Bennett, C. (1992). Regulating privacy: Data Protection and Public Policy in Europe and in the United States, *Cornell University Press*

Bennett, C. J. (2018). The European General Data Protection Regulation: An instrument for the globalization of privacy standards?. *Information Polity*, *23*(2), 239-246.

Ben-Shahar, O., & Schneider, C. E. (2011). The failure of mandated disclosure. *University of Pennsylvania Law Review, 647-749.*

Berkhin, P. (2006). A survey of clustering data mining techniques. In *Grouping multidimensional data* (pp. 25-71). Springer, Berlin, Heidelberg.

Berbers, Y., Hildebrandt, M., & Vandewalle, J. (2018). Privacy in an age of the internet, social networks and Big Data. *Position paper 49b, Royal Flemish Academy of Belgium for Science and the Arts*.

Bergemann, B. (2017). The Consent Paradox: Accounting for the Prominent Role of Consent in Data Protection. In *IFIP International Summer School on Privacy and Identity Management* (pp. 111-131). Springer

Berry, M. J., & Linoff, G. (1997). Data mining techniques: for marketing, sales, and customer support. *John Wiley & Sons, Inc..*

Beyleveld, D., & Brownsword, R. (1998). Human dignity, human rights, and human genetics. *The Modern Law Review, 61(5), 661-680.*

Beyleveld, D., & Brownsword, R. (2007). Consent in the Law. Hart Publishing, Oxford

Bloom, C., Tan, J., Ramjohn, J., & Bauer, L. (2017, July). Self-driving cars and data collection: Privacy perceptions of networked autonomous vehicles. *Symposium on Usable Privacy and Security (SOUPS).*

Borgesius, F. Z., Gray, J., & Eechoud, M. V. (2015). Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework. *Berkeley Tech*. LJ, 30, 2073.

Brimsted, K. (2010), Behavioural Advertising: time to tame the cookie?, Privacy and Data Protection, 11 2 (, 7)

Brownsword, R. (2004). The cult of consent: fixation and fallacy. *King's Law Journal*, 15(2), 223-251.

Brownsword, R., & Yeung, K. (Eds.). (2008). Regulating technologies: Legal futures, regulatory frames and technological fixes. Bloomsbury Publishing.

Brownsword, R. (2009) Consent in Data Protection Law: Privacy, Fair Processing, and Confidentiality' in *Reinventing data protection?* by S. Gutwirth, Y. Poullet, P. de Hert, & C. de Terwangne (Eds.). Dordrecht: Springer.

Buccafurni, D. (2008). Reconsidering the Facilitation of Autonomous Decisionmaking in Genetic Counseling (*Doctoral dissertation, Department of Philosophy, University of Utah*).

Burkert, H. (2000). Privacy-Data Protection, *in Governance of Global Networks in the Light of Different Local Values, edited by Engel, C. and Keller, K.* (pp 43-70).

Büschel, I., Mehdi, R., Cammilleri, A., Marzouki, Y., & Elger, B. (2014). Protecting human health and security in digital Europe: how to deal with the "Privacy Paradox"?. *Science and engineering ethics*, 20(3), 639-658.

Butin, D., & Le Métayer, D. (2014, May). Log analysis for data protection accountability. *In International Symposium on Formal Methods* (pp. 163-178). Springer, Cham.

Bygrave, L., Data Protection Law. Approaching Its Rationale, Logic and Limits, *Kluwer Law International*

Calo, M. R. (2011). The Boundaries of Privacy Harm*. Indiana Law Journal, 86, 1131*. Available at SSRN: http://ssrn.com/abstract=1641487

Camenisch, J., Sommer, D., Fischer-Hübner, S., Hansen, M., Krasemann, H., Lacoste, G., ... & Tseng, J. (2005, November). Privacy and identity management for everyone. In *Proceedings of the 2005 workshop on Digital identity management* (pp. 20-27). ACM.

Carmichael, L., Stalla-Bourdillon, S., & Staab, S. (2016). Data mining and automated discrimination: a mixed legal/technical perspective. *IEEE Intelligent Systems, (6), 51-55.*

Casey, T. (2007) Electronic Surveillance and the Right to be Secure. *UC Davis L. Rev.* 41: 977.

Caulfield, T., & Brownsword, R. (2006). Human dignity: a guide to policy making in the biotechnology era?. *Nature Reviews Genetics, 7(1), 72.*

Cerasaro, E. F. (2017), Accountability principle under the GDPR: is data protection law moving from theory to practice?, *LUISS Law Review*, 2/2017

Clarke, R. (2006) What's privacy. *Australian Law Reform Commission Workshop.* Vol. 28.

Clauβ, S., Kesdogan, D., & Kölsch, T. (2005, November). Privacy enhancing identity management: protection against re-identification and profiling. In *Proceedings of the 2005 workshop on Digital identity management (pp. 84-93). ACM.*

Cohen, J. E. (1999). Examined lives: Informational privacy and the subject as object*. Stan. L. Rev., 52, 1373.*

Cole, P. (1985). New Challenges to the U.S. Multinational Corporation in the European Economic Community*. Data Protection Laws, 17 N.Y.U. J. Iirr'rL L. & POL. 893, 898 n.30*

Colesky, M., Hoepman, J. H., & Hillen, C. (2016, May). A critical analysis of privacy design strategies. In *Security and Privacy Workshops (SPW), 2016 IEEE* (pp. 33-40). IEEE.

Colonna, L. (2013), Mo' Data, Mo' Problems? Personal Data Mining and the Challenge to the Data Minimization Principle, *Conference proceedings of Big Data and Privacy: Making Ends Meet hosted by Stanford Law School and The Center for Internet and Society*

Crawford, K., & Schultz, J. (2014). Big Data and Due Process - Toward a Framework To Redress Predictive Privacy Harms*. BCL Rev., 55(1), 93–128.*

Cradock, E., Millard, D., & Stalla-Bourdillon, S. (2015, May). Investigating Similarity Between Privacy Policies of Social Networking Sites as a Precursor for Standardization. In *Proceedings of the 24th International Conference on World Wide Web (pp. 283-289). ACM.*

Curtin, D. (2018). Does the GDPR Change the World or is the World Changing Beyond the Regulation?. In *Das öffentliche Recht vor den Herausforderungen der Informations- und Kommunikationstechnologien jenseits des Datenschutzes| Information and*

Communication Technologies Challenging Public Law, Beyond Data Protection| Le droit public au défi des technologies de l'information et de la communication, au-delà de la protection des donnés (pp. 35-48). Nomos Verlagsgesellschaft mbH & Co. KG.

Custers, B., van Der Hof, S., Schermer, B., Appleby-Arnold, S., & Brockdorff, N. (2013). Informed Consent in Social Media Use-The Gap between User Expectations and EU Personal Data Protection law. *SCRIPTed, 10, 435.*

Custers, B., van der Hof, S., & Schermer, B. (2014). Privacy expectations of social media users: The role of informed consent in privacy policies. *Policy & Internet, 6(3), 268-295*.

D'Acquisto, G., Domingo-Ferrer, J., Kikiras, P., Torra, V., de Montjoye, Y. A., & Bourka, A. (2015). Privacy by design in big data: an overview of privacy enhancing technologies in the era of big data analytics. *arXiv preprint arXiv:1512.06000*.

Danezis, G., Domingo-Ferrer, J., Hansen, M., Hoepman, J. H., Metayer, D. L., Tirtea, R., & Schiffner, S. (2015). Privacy and Data Protection by Design-from policy to engineering. *arXiv preprint arXiv:1501.03726.*

Danezis, G., & Gürses, S. (2010). A critical review of 10 years of Privacy Technology. *Surveill. Cult. A Glob. Surveill. Soc., 1–16.*

de Koning Gans, G., Hoepman, J. H., & Garcia, F. D. (2008, September). A practical attack on the MIFARE Classic. In *International Conference on Smart Card Research and Advanced Applications (pp. 267-282).* Springer, Berlin, Heidelberg.

Determann, L., & Sprague, R. (2011). Intrusive monitoring: Employee privacy expectations are reasonable in Europe, destroyed in the United States. *Berkeley Technology Law Journal*, 26(2), 979-1036.

De Hert, P. (2007). A right to identity to face the Internet of Things ?, *Strasbourg: Unesco* 1–21.

De Hert, P., & Gutwirth, S. (2006). Privacy, data protection and law enforcement. Opacity of the individual and transparency of power. *Privacy and the criminal law*, 61-104.

De Hert, P., & Gutwirth, S. (2009). Data protection in the case law of Strasbourg and Luxemburg: Constitutionalisation in action. In *Reinventing data protection? (pp. 3-44).* Springer, Dordrecht.

De Hert, P., & Papakonstantinou, V. (2012). The proposed data protection Regulation replacing Directive 95/46/EC: A sound system for the protection of individuals. *Computer Law & Security Review, 28(2),* 130-142.

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608.*

Edwards, L. (2016), Privacy, security and data protection in smart cities: A critical EU law perspective. *Eur. Data Prot. L. Rev. 2 (2016): 28.*

Feinberg, J. (1984). Offense to others (Vol. 2). *Oxford University Press on Demand.*

Ferraris, V. and Bosco, D. and Cafiero, G. and D'Angelo, E. and Suloyeva, Y. (2013), *Defining Profiling*. Available at SSRN: http://ssrn.com/abstract=2366564 orhttp://dx.doi.org/10.2139/ssrn.2366564

Ferretti, F. (2014). Data protection and the legitimate interest of data controllers: much ado about nothing or the winter of rights?. *Common Market Law Review*, *51*(3), 843-868.

Finn, R., Wright, D. and Friedewald, M (2013). Seven Types of Privacy, *Dordrecht European Data Protection: Coming of Age*

Fletcher, A., McCulloch, K., Baulk, S. D., & Dawson, D. (2005). Countermeasures to driver fatigue: a review of public awareness campaigns and legal approaches. *Australian and New Zealand Journal of Public Health, 29(5), 471-476.*

Floridi, L. (2006). Four challenges for a theory of informational privacy. *Ethics and Information technology, 8(3), 109-119.*

Furey, E., & Blue, J. (2018). She Knows Too Much–Voice Command Devices and Privacy. In *2018 29th Irish Signals and Systems Conference (ISSC)* (pp. 1-6). IEEE.

Gellert, R. (2016). We Have Always Managed Risks in Data Protection Law: Understanding the Similarities and Differences Between the Rights-Based and the Risk-Based Approaches to Data Protection. *Eur. Data Prot. L. Rev., 2, 481.*

Gellert, R., & Gutwirth, S. (2013). The legal construction of privacy and data protection. *Computer Law & Security Review*, *29*(5), 522-530.

Gilpin, E, A., et al. (1994), Smoking initiation rates in adults and minors: United States, 1944–1988. *American Journal of Epidemiology 140.6: 535-543.*

Gomez-Arostegui, H. T. (2004). Defining private life under the European convention on human rights by referring to reasonable expectations. *Cal. W. Int'l LJ, 35, 153.*

Goodman, B, and Flaxman, S. (2016). European Union regulations on algorithmic decision-making and a" right to explanation*", arXiv preprint arXiv:1606.08813.*

Gorman C. (2011), Is Society More Reasonable than You? The Reasonable Expectation of Privacy as a Criterion for Privacy Protection, *LLM Law & Technology Masters Thesis, Tilburg University*

Granhag, P. A., & Hartwig, M. (2008). A new theoretical perspective on deception detection: On the psychology of instrumental mind-reading. *Psychology, Crime & Law,* 14(3), 189-200.

Greenleaf, G. (2012), The influence of European data privacy standards outside Europe: implications for globalization of Convention 108', *2 International Data Privacy Law 68 .*

Greenleaf, G. (2017), Questioning 'Adequacy' (Pt I) – Japan (December 7, 2017).150 *Privacy Laws & Business International Report, 1, 6-11; UNSW Law Research Paper No. 1.* Available at SSRN: https://ssrn.com/abstract=3096370

Gutwirth, S., Leenes, R., De Hert, P., & Poullet, Y. (2013). European data protection: Coming of age, Springer

Hall, T. S. (2014). The Quantified Self Movement: Legal Challenges and Benefits of Personal Biometric Data Tracking. *Akron Intell. Prop. J., 7, 27.*

Hansen, M., Berlich, P., Camenisch, J., Clauß, S., Pfitzmann, A., & Waidner, M. (2004). Privacy-enhancing identity management. *Information security technical report*, *9*(1), 35-44.

Harford, T. (2014) Big data: A big mistake?. *Significance* 11.5: 14-19.

Hixson, R. (1987), Privacy in a Public Society*. Human Rights in Conflict . New York, Oxford: Oxford University Press*

Hildebrandt, M. and Tielemans, L. (2013) Data protection by design and technology neutral law. *Computer Law & Security Review 29.5509-521.*

Hildebrandt, M., Backhouse J. (2005), Descriptive analysis and inventory of profiling practices. In FIDIS Project Deliverable 7.2. Available at: http://www.fidis.net

Hildebrandt, M., & Gutwirth, S. (2008). Profiling the European citizen (pp. 17-45). Dordrecht: Springer.

Hintze, M. (2016), Viewing the GDPR through a De-Identification Lens: A Tool for Clarification and Compliance*. Future of Privacy Forum*

Hirsch, D. (2011), The Law and Policy of Online Privacy: Regulation, Self-Regulation, or Co-Regulation? *Seattle University Law Review, Vol. 34, No. 2, 2011.* Available at SSRN: http://ssrn.com/abstract=1758078

Hoepman, J. (2012). In things we trust? Towards trustability in the internet of things. *Constructing Ambient Intelligence*. Springer Berlin Heidelberg. 287-295.

Hoepman, J. H., Hubbers, E., Jacobs, B., Oostdijk, M., & Schreur, R. W. (2006, October). Crossing borders: Security and privacy issues of the european e-passport. In *International Workshop on Security (pp. 152-167).* Springer, Berlin, Heidelberg

Holvast, J. (2008) History of privacy. Springer Berlin Heidelberg.

Hornung, G., & Schnabel, C. (2009). Data protection in Germany I: The population census decision and the right to informational self-determination. *Computer Law & Security Review, 25(1), 84-88.*

Hornung, G. (2012). A General Data Protection Regulation for Europe? Light and shade in the Commission's draft of 25 January 2012*. SCRIPTed*, 9(1), 64–81.

Houghton, D. and Joinson, A. (2010). Privacy, social network sites, and social relations. *Journal of Technology in Human Services 28.1-2*

Howe, D. C., & Nissenbaum, H. (2009). TrackMeNot: Resisting surveillance in web search. *Lessons from the Identity trail: Anonymity, privacy, and identity in a networked society*, 23, 417-436.

Hughes, M. (2004). Privacy in aged care*. Australasian Journal on Ageing*, 23(3), 110-114.

Jones, P., Comfort, D., & Hillier, D. (2005). Corporate social responsibility and the UK's top ten retailers. *International Journal of retail & Distribution management, 33(12), 882-892*.

Kalven Jr, H. (1966).Privacy in tort law--were Warren and Brandeis wrong *Law & Contemp. Probs.* 31 : 326.

Kaminski, M. E. (2015). Regulating Real-World Surveillance. Washington Law Review, Vol. 9, No. 113

Katal, A,, Wazid, M., and Goudar, R. (2013) Big data: issues, challenges, tools and good practices. *Sixth International Conference on Contemporary Computing (IC3), IEEE.*

Kerber, W. (2016). Digital markets, data, and privacy: competition law, consumer law and data protection. *Journal of Intellectual Property Law & Practice, 11(11), 856-866.*

Kokott, J., and Sobotta, C.. The distinction between privacy and data protection in the jurisprudence of the CJEU and the ECtHR. *International Data Privacy Law 3.4 (2013): 222-228.*

Koops, B. J., & Leenes, R. (2005). Code and the slow erosion of privacy. *Mich. Telecomm. & Tech. L. Rev., 12, 115.*

Koops, B. J. (2011). Forgetting footprints, shunning shadows: A critical analysis of the'right to be forgotten'in big data practice. *SCRIPTed*, *8*, 229

Koops, B. J. (2014). The trouble with European data protection law*. International Data Privacy Law, 4(4), 250-261.*

Koops, B. J., & Leenes, R. (2014). Privacy regulation cannot be hardcoded. A critical comment on the 'privacy by design'provision in data-protection law. *International Review of Law, Computers & Technology*, *28*(2), 159-171.

Koops, B. J., Newell, B. C., Timan, T., Skorvanek, I., Chokrevski, T., & Galic, M. (2016). A typology of privacy. *U. Pa. J. Int'l L., 38, 483.*

Koppelman, A. (2005) Does obscenity cause moral harm?. *Columbia Law Review (2005): 1635-1679.*

Korenhof, P., Ausloos, J., Szekely, I., Ambrose, M., Sartor, G., & Leenes, R. (2015). Timing the right to be forgotten: A study into "time" as a factor in deciding about retention or erasure of data. *In Reforming European data protection law* (pp. 171-201). Springer, Dordrecht.

Kristol, D. (2001). HTTP Cookies: Standards, privacy, and politics. *ACM Transactions on Internet Technology (TOIT) 1.2 : 151-198.*

Krumholz, H. M. (2014). Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. *Health Affairs, 33(7), 1163-1170.*

Kuner, C. (2017). The Internet and the Global Reach of EU Law, *LSE Law, Society and Economy Working Papers 4/2018, London School of Economics and Political Science*

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science, 343(6176), 1203-1205.*

Le Métayer, D. (2010). Privacy by design: a matter of choice. In *Data protection in a profiled world* (pp. 323-334). Springer, Dordrecht.

Leonard, P. (2013). Customer data analytics: privacy settings for 'Big Data'business. *International Data Privacy Law, 4(1), 53-68.*

Lessig, L. (1999), Code and Other Laws of Cyberspace*, Basic Books, ReadHowYouWant.com*

Lin, D., & Loui, M. (1998). Taking the byte out of cookies: Privacy, consent and the Web. *Computers and Society*, 28 (1), 39–51.

Lindqvist, J. (2017). New challenges to personal data processing agreements: is the GDPR fit to deal with contract, accountability and liability in a world of the Internet of Things*?. International Journal of Law and Information Technology. Volume 26, Issue 1*, 1 March 2018, Pages 45–63

Macenaite, M. (2017). The "Riskification" of European Data Protection Law through a two-fold Shift. *European Journal of Risk Regulation*, *8*(3), 506-540.

Mac Sithigh, Daithi (2012) Virtual walls: the law of pseudo-public spaces. *International Journal of Law in Context, 8 (03).*

McDonald, A. M., & Cranor, L. F. (2008). The cost of reading privacy policies. *ISJLP*, *4*, 543

Maloni, M. J., & Brown, M. E. (2006). Corporate social responsibility in the supply chain: an application in the food industry. *Journal of business ethics*, *68*(1), 35-52.

Mantelero, A. (2014). The future of consumer data protection in the EU Re-thinking the "notice and consent" paradigm in the new era of predictive analytics. *Computer Law & Security Review, 30(6), 643-660.*

Marthews, A. and Tucker, C. (2014), Government Surveillance and Internet Search Behavior. *SSRN Working Paper, Social Science Electronic Publishing, Inc., Rochester, NY.*

Martos, C. M. (2008). The Transformation of Intimacy and Privacy through Social Networking Sites. *Institute Of Communications Studies, University Of Leeds, Regulation & Governance,* 2, 425-444.

Mayer-Schönberger, V., and Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

Mayer-Schonberger, V., & Padova, Y. (2015). Regime Change: Enabling Big Data through Europe's New Data Protection Regulation*. Colum. Sci. & Tech. L. Rev., 17, 315.*

McGlone, T., Spain, J. W., & McGlone, V. (2011). Corporate social responsibility and the millennials. *Journal of Education for Business*, 86(4), 195-200.

McKay, C. (2015). Next Generation Privacy: The Internet of Things, Data Exhaust, and Reforming Regulation by Risk of Harm*. Groningen Journal of International Law,* 2, 30.

McKeon, M. (2009) The Secret History of Domesticity: Public, private, and the division of knowledge. *JHU Press*

McMurrian, R., & Washburn, J. H. (2008). Branding: a social contract between a business and its customer. In *Contemporary Thoughts on Corporate Branding and Corporate Identity Management (pp. 5-22).* Palgrave Macmillan UK.

Mejova, Y., Macy, M. W., Weber, I. (2015), Twitter: A Digital Socioscope, *Cambridge University Press*

Moerel, L. "Back to Basics: When Does EU Data Protection Law Apply?" *International Data Privacy Law 1, no. 2* (January 24, 2011): 92–110

Mohan, P., Thakurta, A., Shi, E., Song, D., & Culler, D. (2012). GUPT: privacy preserving data analysis made easy. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data (pp. 349-360). ACM.*

Moiso, C., & Minerva, R. (2012). Towards a user-centric personal data ecosystem the role of the bank of individuals' data. In *Intelligence in Next Generation Networks (ICIN), 2012 16th International Conference on (pp. 202-209).* IEEE.

Montgomerie, J., & Roscoe, S. (2013, December). Owning the consumer—Getting to the core of the Apple business model. In *Accounting Forum (Vol. 37, No. 4, pp. 290-299).* Elsevier.

Movius, L. & Krup, N., U.S. and EU Privacy Policy: Comparison of Regulatory Approaches, *3INT'L J. COMM. 167, 173 (2009).*

Narayanan, A. & Shmatikov, V. (2008), Robust De-Anonymization of Large Sparse Datasets, *2008 IEEE Symp. on Sec. and Privacy 111, Feb. 5, 2008*, available at http://www. cs.utexas.edu/-shmat/shmatoak08netflix.pdf.

Nenoist, E. (2008) Collecting Data for the Profiling of Web Users, in *Profiling the European Citizen*. Springer Netherlands

Newell, B. (2011). Rethinking Reasonable Expectations of Privacy in Online Social Networks, *Richmond Journal of Law and Technology Vol.XVII, Issue 4*

Nissenbaum, H. (1998). Protecting privacy in an information age: The problem of privacy in public. Law and philosophy, 17(5), 559-596.

Nissenbaum, H. (2004). Privacy as contextual integrity. Washington law review 79.1.

Notario, N., Crespo, A., Martín, Y. S., Del Alamo, J. M., Le Métayer, D., Antignac, T., ... & Wright, D. (2015, May). PRIPARE: integrating privacy best practices into a privacy engineering methodology. *In Security and Privacy Workshops (SPW), 2015 IEEE (pp. 151-158).* IEEE.

Nouwt, S. (2008). Reasonable expectations of geo-privacy. *SCRIPTed*, 5, 375.

Ohm, P. (2010), Broken Promises of Privacy: Responding to the Surprising Failure of Anonymisation. UCLA Law Review, Vol. 57, p. 1701*; U of Colorado Law Legal Studies Research Paper No. 9-12*. Available at SSRN: http://ssrn.com/abstract=1450006

Osher, S. A. (2002). Privacy, Computers and the Patriot Act: The Fourth Amendment Isn't Dead, But No One Will Insure It. *Fla. L. Rev.,* 54, 521.

Parent, W.A. (1983).Privacy, morality and the Law. *Philosophy & Public Affairs,* 12(4), 269-288.

Pauleen, D. J., & Wang, W. Y. (2017). Does big data mean big knowledge? KM perspectives on big data and analytics. *Journal of Knowledge Management*, *21*(1), 1-6.

Peddinti, S. T., & Saxena, N. (2010, July). On the privacy of web search based on query obfuscation: a case study of TrackMeNot. *In International Symposium on Privacy Enhancing Technologies Symposium (pp. 19-37).* Springer, Berlin, Heidelberg

Pearson, S. (2017). Strong Accountability and Its Contribution to Trustworthy Data Handling in the Information Society. In *IFIP International Conference on Trust Management* (pp. 199-218). Springer, Cham.

Pech L. (2012), The Rule of Law as a Guiding Principle of the European Union's External Action. *CLEER Working Paper 2012/3. The Hague: Asser Institute*

Pelteret, M., & Ophoff, J. (2016). A review of information privacy and its importance to consumers and organizations. *Informing Science: The International Journal of an Emerging Transdiscipline*, *19*, 277-301.

Pierson, J., & Heyman, R. (2011). Social media and cookies: challenges for online privacy. info, 13(6), 30-42.

Peppet, S., Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security & Consent (March 1, 2014). *Texas Law Review,*. Available at SSRN: http://ssrn.com/abstract=2409074

Politou, E., Alepis, E., & Patsakis, C. (2018). Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. *Journal of Cybersecurity*, *4*(1), tyy001.

Preibusch, S. (2015). Privacy behaviors after Snowden. *Communications of the ACM*, 58(5), 48-55.

Pfitzmann, A., & Hansen, M. (2010). A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management.

Rainie, L., Smith, A., & Duggan, M. (2013). Coming and going on Facebook. *Pew Research Center's Internet and American Life Project.*

Pruitt, S. W., & Friedman, M. (1986). Determining the effectiveness of consumer boycotts: A stock price analysis of their impact on corporate targets*. Journal of Consumer policy, 9(4), 375-387.*

Quelle, C. (2016). Not just user control in the general data protection regulation. In *IFIP International Summer School on Privacy and Identity Management* (pp. 140-163). Springer,

Rauhofer, J. (2015). Of Men and Mice: Should the EU Data Protection Authorities' Reaction to Google's New Privacy Policy Raise Concern for the Future of the Purpose Limitation Principle. *Eur. Data Prot. L. Rev., 1, 5.*

Reidenberg, J. R. (1994). Setting standards for fair information practice in the US private sector*. Iowa L. Rev., 80, 497.*

Reidenberg, J. R. (2002), Privacy wrongs in search of remedies, *Hastings LJ 54: 877.*

Reidenberg, J. R. (2014). Privacy in public. *69 University of Miami Law Review 141*

Roch, M. P. (1996). Filling the Void of Data Protection in the United States : Following the European Example. *Santa Clara High Technology Law Journal, 12(1), 71–96*. Retrieved from http://digitalcommons.law.scu.edu/chtlj/vol12/iss1/3

Rouvroy, A. (2008). Privacy, data protection, and the unprecedented challenges of ambient intelligence. *Studies in ethics, law, and technology, 2(1).*

Rotenberg, M. and Jacob, D. (2013), Updating the law of information privacy: the new framework of the European Union, *Harvard Journal of Law & Public Policy, Vol. 36 Issue 2*

Safari, B. A. (2016). Intangible Privacy Rights: How Europe's GDPR Will Set a New Global Standard for Personal Data Protection. *Seton Hall L. Rev., 47, 809.*

Sagiroglu, S., and Sinanc, D. (2013) Big data: A review. *2013 International Conference on Collaboration Technologies and Systems (CTS),* IEEE.

Salter, M. and Mason, J. (2007), Writing Law Dissertations: An Introduction and Guide to the Conduct of Legal Research, *Pearson Education*.

Sapronov, W., & Srouji, J. (2017). Class Consciousness: Class Action Arbitration under US and EU Privacy Laws. *YB on Int'l Arb.*, *5*, 83.

Sarkar, S., Chatterjee, S., & Misra, S. (2018). Assessment of the Suitability of Fog Computing in the Context of Internet of Things. *IEEE Transactions on Cloud Computing, 6(1), 46-59.*

Shank, R. (1986). Privacy: History, legal, social, and ethical aspects. In *Library Trends 35 (1) Summer 1986: Privacy, Secrecy, and National Information Policy: 7-18*

Schermer, B. W., Custers, B., & van der Hof, S. (2014). The crisis of consent: How stronger legal protection may lead to weaker consent in data protection. *Ethics and Information Technology, 16(2), 171-182.*

Schreurs, W., Hildebrandt, M., Kindt, E., & Vanfleteren, M. (2008). Cogitas, ergo sum. The role of data protection law and non-discrimination law in group profiling in the private sector. In *Profiling the European citizen (pp. 241-270)*. Springer, Dordrecht.

Schwartz, P. M. (1994). Privacy and participation: Personal information and public sector regulation in the United States*. Iowa L. Rev., 80, 553.*

Schwartz, P. M. (1999). Privacy and democracy in cyberspace. *Vand. L. Rev., 52, 1607.*

Schwartz, P.., and Solove, J. (2011). The PII problem: Privacy and a new concept of personally identifiable information. *NYUL rev. 86 (2011): 1814.*

Segall, J. (2009), Google Street View: Walking the line of privacy-intrusion upon seclusion and publicity given to private facts in the digital age*. Pitt. J. Tech. L. & Pol'y 10*

Shilton, K. (2009). Four billion little brothers?: Privacy, mobile phones, and ubiquitous data collection. *Communications of the ACM, 52(11), 48-53.*

Shore, D. A. (2005). The trust prescription for healthcare: Building your reputation with consumers. *Health Administration Press.*

Simitis, S. (1987). Reviewing privacy in an information society. *University of Pennsylvania Law Review, 135(3), 707-746.*

Smith, M., Szongott, C., Henne, B., & Von Voigt, G. (2012). Big data privacy issues in public social media. In Digital Ecosystems Technologies (DEST), *2012 6th IEEE International Conference on (pp. 1-6). IEEE.*

Smith, N. C. (2003). Corporate social responsibility: whether or how?. *California management review*, *45*(4), 52-76.

Smith, R., Grabosky, P., & Urbas, G. (2004). Cyber criminals on trial. *Criminal Justice Matters*, *58*(1), 22-23.

Sobolewski, M., Mazur, J., & Paliński, M. (2017). GDPR: A Step Towards a User-centric Internet?. *Intereconomics, 52(4), 207-213.*

Soffer, P. (2010). Mirror, mirror on the wall, can i count on you at all? Exploring data inaccuracy in business processes. *Enterprise, business-process and information systems modeling, 14-25.*

Solove, D. (2006) A taxonomy of privacy*. University of Pennsylvania Law Review, 477-564*

Solove, D. (2007) A Brief History of Information Privacy Law. *PROSKAUER ON PRIVACY, PLI, 2016; GWU Law School Public Law Research Paper No. 215*. Available at SSRN: https://ssrn.com/abstract=914271

Solove, D. (2007). The future of reputation: Gossip, rumor, and privacy on the Internet. *Yale University Press.*

Solove, D. J. (2012). Introduction: Privacy self-management and the consent dilemma*. Harv. L. Rev., 126, 1880.*

Soltani, A., Canty, S., Mayo, Q., Thomas, L., & Hoofnagle, C. J. (2010). Flash Cookies and Privacy. In *AAAI spring symposium: intelligent information privacy management (Vol. 2010, pp. 158-163).*

Stalla-Bourdillon, S., and Knight, A. (2016), Anonymous data v. Personal data–A false debate: An EU perspective on anonymisation, pseudonymisation and personal data. *Wis. Int'l LJ APA*

Stewart KA and Segars AH (2002) An empirical examination of the concern for information privacy instrument*. Information Systems Research 13(1), 36-49.*

Stieger, S., Burger, C., Bohn, M., & Voracek, M. (2013). Who commits virtual identity suicide? Differences in privacy concerns, internet addiction, and personality between Facebook users and quitters. *Cyberpsychology, Behavior, and Social Networking, 16(9), 629-634.*

Surden, H., & Williams, M. A. (2016). Technological opacity, predictability, and self-driving cars. *Cardozo L. Rev., 38, 121.*

Swan, M. (2013). The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data, 1(2), 85-99.*

Sweeney, L. (2000), Simple Demographics Often Identify People Uniquely, *Carnegie Mellon Univ., Sch. of Computer Sci., Data Privacy Lab., Working Paper No. 3, 2000.*

Swire, P., & Lagos, Y. (2012). Why the right to data portability likely reduces consumer welfare: antitrust and privacy critique*. Md. L. Rev., 72, 335.*

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs*. American Journal of Political Science, 50(3), 755-769.*

Tankard, C. (2016). What the GDPR means for businesses. *Network Security, 2016(6), 5-8.*

Tavani, H. T. (2008). Informational privacy: Concepts, theories, and controversies*. The handbook of information and computer ethics, 131-164.*

Taylor, L., Floridi, L., & Van der Sloot, B. (Eds.). (2016). Group privacy: New challenges of data technologies (Vol. 126). Springer.

Tene, O., & Polonetsky, J. (2011). Privacy in the age of big data: a time for big decisions. *Stan. L. Rev. Online, 64, 63.*

Tene, O., & Polonetsky, J. (2012). Big data for all: Privacy and user control in the age of analytics. *Nw. J. Tech. & Intell. Prop., 11, xxvii.*

Tene, O., and Polenetsky, J. (2012), To Track or Do Not Track: Advancing Transparency and Individual Control in Online Behavioral Advertising., *Minn. JL Sci. & Tech. 13: 281.*

Tene, O., & Polonetsky, J. (2013). Judged by the tin man: Individual rights in the age of big data. *J. on Telecomm. & High Tech. L., 11, 351.*

Tene, O., & Polonetsky, J. (2013). A theory of creepy: technology, privacy and shifting social norms. *Yale JL & Tech., 16, 59.*

Treacy, B. and Bapat, A. (2013). The 'Internet of Things' — already in a home near you?*, Privacy and Data Protection, 14 2 (11)*

Tsakalakis, N., Stalla-Bourdillon, S., & O'hara, K. (2016). What's in a name: the conflicting views of pseudonymisation under eIDAS and the General Data Protection Regulation. In *Open Identity Summit 2016*. vol. P-264, 8 pp, pp. 167-174.

Van Alsenoy, B., Kosta, E., & Dumortier, J. (2014). Privacy notices versus informational self-determination: Minding the gap. *International Review of Law, Computers & Technology*, *28*(2), 185-203.

Van der Hof, S., & Prins, C. (2008). Personalisation and its influence on identities, behaviour and social values. In *Profiling the European Citizen (pp. 111-127).* Springer, Dordrecht.

Van Dijk, N., Gellert, R., & Rommetveit, K. (2016). A risk to a right? Beyond data protection risk assessments. *Computer Law & Security Review, 32(2), 286-306.*

Voss, W. (2016). After google Spain and Charlie Hebdo: the continuing evolution of European Union Data Privacy Law in a time of change, *Business Lawyer, Vol. 71, No. 1, 2015/2016*

Wachter, S and Mittelstadt, B, and Floridi, L. (2016), Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation (December 28, 2016). *International Data Privacy Law, 2017.* Available at SSRN: https://ssrn.com/abstract=2903469 or http://dx.doi.org/10.2139/ssrn.2903469

Wachter, S. (2018). Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR. *Computer Law & Security Review, 34(3), 436-449.*

Wang, Y. C., McPherson, K., Marsh, T., Gortmaker, S. L., & Brown, M. (2011). Health and economic burden of the projected obesity trends in the USA and the UK. *The Lancet, 378(9793), 815-825.*

Weber, Rolf., and Weber, Rebecca. (2010). Internet of things: legal perspectives. Vol. 49*. Springer Science & Business Media.*

Wilkins, R. G. (1987). Defining the reasonable expectation of privacy: an emerging tripartite analysis. *Vand. L. Rev., 40, 1077.*

Wilson, D., & Sutton, A. (2004). Watched over or over-watched? Open street CCTV in Australia. *Australian & New Zealand Journal of Criminology, 37(2), 211-230.*

Wilson, S. (2014), The Collision between Big Data and Privacy Law. *Australian Journal of Telecommunications and the Digital Economy, Volume 2 Number 3, October 2014.* Available at SSRN: http://ssrn.com/abstract=2548079

Wing, M. G., Eklund, A., & Kellogg, L. D. (2005). Consumer-grade global positioning system (GPS) accuracy and reliability. *Journal of forestry, 103(4), 169-173.*

Whitman, J. (2014). The Two Western Cultures of Privacy: Dignity Versus Liberty, *Yale Law Journal*

Woodside, J. M. (2015, March). Wearable technology acceptance model: Google Glass. In *Society for Information Technology & Teacher Education International Conference* (pp. 1800-1802). Association for the Advancement of Computing in Education (AACE).

Wright, D., Gutwirth, S., Friedewald, M., Vildjiounaite, E., & Punie, Y. (Eds.). (2008). Safeguards in a world of ambient intelligence (Vol. 1). Springer Science & Business Media.

Zanfir, G. (2014). Forgetting about consent. Why the focus should be on “suitable safeguards” in data protection law. In *Reloading Data Protection (pp. 237-257).* Springer Netherlands.

Zhao, S., Grasmuck, S., & Martin, J. (2008). Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in human behavior*, *24*(5), 1816-1836.

Zhu, B. (2014). Traditional Tort for a Modern Threat: Applying Intrusion upon Seclusion to Dataveillance Observations*, NYUL Rev. 89 : 2381.*

News Articles

Akamai, akamai’s [state of the internet] report (2014), p. 1, at https://www.antel.com.uy/wps/wcm/connect/2e38bd0047ad6c9682d3e7af6890d810/q3-2014-state-of-the-internetreport+(2).pdf?MOD=AJPERES .

Barbaro, M. & Zeller, T. (2006), A Face Is Exposed for AOL Searcher No. 4417749, N.Y. TIMES, Aug. 9, 2006, at Al.

BBC News (2012), Facebook buys Instagram photo sharing network for $1bn, BBC Technology, 10 April 2012

BBC News (2014), Facebook to buy messaging app WhatsApp for $19bn, BBC Business, 20 February 2014

BBC News (2017), Britons 'should know their neighbors', accessed at https://www.bbc.co.uk/news/uk-40811530 on 4/10/2018

Beckett, L. (2017), Big Data Brokers: They Know Everything About You and Sell it to the Highest Bidder, Gizmodo, 18 March 2013, available at: http://gizmodo.com/5991070/big-data-brokers-they-know-everything-about-you-and-sell-it-to-the-highest-bidder

Booth, R., Facebook reveals news feed experiment to control emotions, 30/6/2014, The Guardian, available at https://www.theguardian.com/technology/2014/jun/29/facebook-users-emotions-news-feeds

Bourne, J. (2017), Two in five execs grumble flash technology is too expensive, research finds, CloudTech, 29 July 2016, accessed 20/3/2017, https://www.cloudcomputing-news.net/news/2016/jul/29/execs-grumble-flash-technology-too-expensive-research-finds/

Cellan-Jones, C. (2016) Snoopers law creates security nightmare, BBC News, 29 November 2016

Dailey, K. (2012), Could Google's data hoarding be good for you?, BBC News Magazine, accessed on 17/6/2018 at https://www.bbc.co.uk/news/magazine-16749076

Davies, S (2013), Predictions for Privacy A report on the issues and trends that will dominate the privacy landscape in 2013, LSE Enterprise, the London School of Economics

Dierig, C., Fuest, B., Kaiser, T., and Wisdorff, F. (2014), Die Welt ; http://www.welt.de/wirtschaft/article126882276/Deutsche-unterschaetzen-den-Wert-persoenlicher-Daten.html

Grierson, J., PlayStation data hack: Sony fined £250,000 for 'preventable' breach, 24 January 2013, accessed on 28/9/2017at http://www.independent.co.uk/news/business/news/playstation-data-hack-sony-fined-250000-for-preventable-breach-8464651.html

Groome, I. What is WhatsApp encryption?, Metro News UK, 27/3/2017, available at http://metro.co.uk/2017/03/27/what-is-whatsapp-encryption-6535899/

The Guardian Staff and agencies, Google user data to be merged across all sites under contentious plan, The Guardian, 25 January 2012, available at: https://www.theguardian.com/technology/2012/jan/25/google-merge-user-data-privacy

Hern, A. (2017), Amazon's Snowmobile will let you upload stuff by the truckload – literally, The Guardian, 5 December 2016, accessed 20/3/2017, https://www.theguardian.com/technology/2016/dec/05/amazon-snowmobile-upload-truckload

Hill, K. (2012). How Target figured out a teen girl was pregnant before her father did. Forbes, Inc., available at https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#7a12f2196668 accessed on 16/10/2018

Hosea, M. Why brands are creating 'social contracts' to build trust around data use, Marketing Week, 4 October 2016, available at

https://www.marketingweek.com/2016/10/04/why-brands-are-creating-social-contracts-to-build-trust-around-data-use/#content, accessed on 16/05/2017

Intel IT Center, Planning Guide: Getting Started with Hadoop, Steps IT Managers Can Take to Move Forward with Big Data Analytics, June 2012

Kelion, L. (2017), Wikileaks: CIA has tools to snoop via TVs, BBC News, 7 March 2017

Kroes, N. (2013), Data isn't a four-letter word, IAPP Europe Data Protection Congress/Brussels, 11 Dec. 2013. Available at http://europa.eu/rapid/press-release_SPEECH-13-1059_en.htm.

Lambert, F. Tesla has now 1.3 billion miles of Autopilot data going into its new self-driving program, Electrek, https://electrek.co/2016/11/13/tesla-autopilot-billion-miles-data-self-driving-program/

Lang, N. (2015), Why teens are leaving Facebook: It's 'meaningless', Washington Post

Levin, S. Facebook and Instagram ban developers from using data for surveillance, The Guardian, 13/3/2017

Lomas, N. (2016), WhatsApp to share user data with Facebook for ad targeting — here's how to opt out, Techcrunch, 25 August 2016

Maldoff, G. (2017). The risk-based approach in the GDPR: interpretation and implications. IAPP https://iapp. org/media/pdf/resource_center/GDPR_Study_Maldoff. pdf. Accessed, 12.

Miller, P., Amazon wants to ship your data to the cloud using a literal truck, The Verge, 30 November 2016, Available at: http://www.theverge.com/circuitbreaker/2016/11/30/13797212/amazon-aws-snowmobile-snowball-cloud-storage-truck

Moerel, L., & Prins, J. E. J. (2015). Further Processing of Data Based on the Legitimate Interest Ground: The End of Purpose Limitation, IAPP News, accessed at https://iapp.org/news/a/on-the-death-of-purpose-limitation/ on 11/10/2018

Moss, E. (2014), Neighbourhood watch: how domestic CCTV is sweeping the UK, The Guardian, access on 4/10/2018 at: https://www.theguardian.com/world/2014/dec/19/neighbourhood-watch-domestic-cctv-sweeping-uk

Newcomb, A. Facebook's New Tools are Designed to Put Privacy in Your Pocket, NBC News, 26/1/2017, available at http://www.nbcnews.com/tech/tech-news/facebook-s-new-tools-are-designed-put-privacy-your-pocket-n712246

Patel, P. (2017), Tech Turns to Biology as Data Storage Needs Explode, Scientific American, 31 May 2016, accessed 20/3/2017, https://www.scientificamerican.com/article/tech-turns-to-biology-as-data-storage-needs-explode/

Peterson, A. LOVEINT: When NSA officers use their spying power on love interests, Washington Post, August 24, 2013, accessed 23/5/2017 at https://www.washingtonpost.com/news/the-switch/wp/2013/08/24/loveint-when-nsa-officers-use-their-spying-power-on-love-interests/?utm_term=.6f11788480b8

Russell, J. (2018), China can apparently now identify citizens based on the way they walk, Techcrunch, accessed at https://techcrunch.com/2018/11/07/china-can-apparently-now-identify-citizens-based-on-the-way-they-walk/ on 09/11/2018

Scott, M. and Cerulus, L. (2018), Europe's new data protection rules export privacy standards worldwide, POLITICO, last accessed on 3/2/2018 at https://www.politico.eu/article/europe-data-protection-privacy-standards-gdpr-general-protection-data-regulation/

Stern, J., Facebook's New Privacy Controls Roll Out to All U.S. Users, ABC News, 20/12/2012

Surveillance Self-Defense, Reasonable Expectation of Privacy, accessible at https://ssd.eff.org/your-computer/govt/privacy

Valentino-Devries, J., Singer-Vine, J. and Soltani, A., Websites Vary Prices, Deals Based on Users' Information, *Washington Street Journal*, 24 December 2012, available at: https://www.wsj.com/articles/SB10001424127887323777204578189391813881534

Varghese, S., Cops use pacemaker data to file arson charges, iTWire, 2 February 2017, available at: https://www.itwire.com/data/76677-cops-use-pacemaker-data-to-file-arson-charges.html

Whigham, N. , Leaked document reveals Facebook conducted research to target emotionally vulnerable and insecure youth. News.com.au, 1/5/2017, available at http://www.news.com.au/technology/online/social/leaked-document-reveals-facebook-conducted-research-to-target-emotionally-vulnerable-and-insecure-youth/news-story/d256f850be6b1c8a21aec6e32dae16fd

White Papers

Centre for Information Policy Leadership. (2014). A Risk-based Approach to Privacy: Improving Effectiveness in Practice, 19 June 2014

Article 29 Working Party (2001), Opinion 08/2001 on the processing of personal data in the employment context, WP48, Adopted on 13 September 2001

Article 29 Working Party (2007), Opinion 4/2007 on the Concept of Personal Data, WP136, Adopted on 20 June 2007

Article 29 Working Party (2010), Opinion 1/2010 on the concepts of "controller" and "processor" WP 169, Adopted on 16 February 2010

Article 29 Data Protection Working Party (2011), Opinion 15/2011 on the definition of consent, WP187 Adopted on 13 July 2011

Article 29 Working Party, "Letter from the Article 29 Working Party addressed to Google along with the recommendations" (Brussels, 16 Oct. 2012), at ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2012/20121016_letter_to_google_en.pdf

Article 29 Data Protection Working Party (2013), Opinion 03/2013 on purpose limitation, WP203, Adopted on 2 April 2013

Article 29 Data Protection Working Party (2013), Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, WP217, Adopted on 9 April 2014

Article 29 Data Protection Working (2014), Opinion 05/2014 on Anonymisation Techniques, WP216, Adopted on 10 April 2014

Article 29 Working Party (2013), Advice paper on essential elements of a definition and a provision on profiling within the EU General Data Protection Regulation, adopted on 13 May 2013.

Article 29 Working Party (2017), Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679, WP 248, Adopted on 4 October 2017

Article 29 Working Party (2017), Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, WP251, Adopted on 3 October 2017

Buttarelli, G. (2016). The EU GDPR as a clarion call for a new global digital gold standard, *European Data Protection Supervisor*, accessed at 14/4/2018 at https://edps.europa.eu/press-publications/press-news/blog/eu-gdpr-clarion-call-new-global-digital-gold-standard_en

European Commission (2011), Flash Eurobarometer 225, Attitudes towards data protection and electronic identity in the European Union. Survey conducted by the Gallup Organization Hungary upon the request of the Directorate General Justice, Freedom and Security

European Commission (2011), Special Eurobarometer 359, Attitudes towards data protection and electronic identity in the European Union. Survey conducted by TNS Opinion & Social at the request of Directorate-General Justice, Information Society & Media and Joint Research Centre Survey

European Data Protection Supervisor, Opinion of the European Data Protection Supervisor on the data protection reform package, 7 March 2012

FTC Staff report Internet of Things: Privacy and Security in a Connected World, January 2015 at https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013- workshop-entitled-internet-things-privacy/150127iotrpt.pdf (herafter FTC, 215).

Information Commissioner's Office, Code of Practice on Anonymisation: Managing Data Protection Risk, (2012).

Legislation

Charter of Fundamental Rights of the European Union, 2010 O.J. C 83/02

Consolidated Version of the Treaty on the Functioning of the European Union, 2008 O.J. C 115/47

Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Off. J.L. 281 (Nov. 23, 1995) ("Data Protection Directive")

Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases

Reg (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Dir 95/46/EC (General Data Protection Regulation) 2016

Constitution of the United States

Council of Europe, Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, Jan. 28, 1981, Eur. TS. No. 108

Hessisches Datenschutzgesetz (The Hesse Data Protection Act), Gesetz und Verordungsblatt I(1970), 625.

Cases

European Court of Human Rights

ECtHR, Klass v. Germany, Application no. 5029/71, Judgement of 6 September 1978

ECtHR, Amann v. Switzerkand, Application no. 27798/95, Judgement of 16 February 2000

ECtHR, Halford v. United Kingdom, Application no. 20605/92 Judgment of 25 June 2997, Reports, 1997-III

ECtHR, Rotaru v Romania, Application no. 28341/95, Judgment of 4 May 2000

ECtHR, Malone v. United Kingdom, Application no. 8961/79, Judgment of 2 August 1984

ECtHR, P.G. and J.H. v. the United Kingdrom, Application no. 44787/95, Judgment of 4 May 2000

ECtHR, Copland v. the United Kingdom, Application no. 62617/00, Judgment of 23 April 2007

ECtHR, Société Colas Est and others v. France, Application no. 37971/97, Judgment of 16 April 2002.

ECtHR, Rees v. UK, Application no. 9532/81, Judgment of 25 October 1986,

ECtHR, Cossey v. UK, Application no. 10843/84, Judgment of 27 September 1990, Series A, No. 184

ECtHR, B v. France, Application no. 13343/87, Judgment of 25 March 1992

ECtHR, Christine Goodwin v. the United Kingdom, Application no. 28957/95, Judgment of 11 July 2002

ECtHR, Antovic and Mirkovic v Montenegro, Application no. 70838/13, Judgment of 28 November 2017

ECtHR, P.G. and J.H. v. United Kingdom, Application no. 44787/98, Judgment of 25 September 2001

ECtHR, Peck v The United Kingdom, Application no. 44647/98, Judgment of 28 January 2003

ECtHR, Perry v The United Kingdom, Application no. 63737/00, Judgment of 17 July 2003

European Court of Justice

Case C-101/01, Bodil Lindqvist v Aklagarkammaren, (2003) ECLI:EU:C:2003:596

Case C-131/12, Google Spain SL v. Agencia Espanola de Proteccion de Datos (AEPD), (2014) E.C.R. 317

Case C-582/14, Breyer v Bundesrepublik Deutschland, (2016) ECLI:EU:C:2016:779

Case C-434/16, Nowak v Data Protection Commissioner, (2017) ECLI:EU:C:2017:994

EU National Courts

British Horseracing Bd. Ltd.v. William Hill Org. Ltd.(Chancery Div. 2000)

BVerfGE 65, 1 – Volksz¨ahlung Urteil des Ersten Senats vom 15. Dezember 1983 auf diem¨undliche Verhandlung vom 18. und 19. Oktober 1983

Campbell v Mirror Group Newspapers Ltd [2004] UKHL 22

Gulati & Ors v MGN Limited (confirmed by the Court of Appeal in Representative Claimants v MGN Limited [2015] EWCA Civ 1291

Lüdi v. Switzerland, 1992, ECHR, Series A, No. 238, PN 2004-135

UNMS v. Belpharma Communication –Court of Brussels 16 March 1999

von Hannover v Germany [2004] EMLR 379; (2005) 40 EHRR 1

US Courts

DOJ v. Reporters Comm. for Free Press, 489 U.S. 749 (1989)

Gonzales v. Google, Inc., 234 F.R.D. 674 (N.D. Cal. 2006).

Katz v. United States, 389 U.S. 347 (1967)

Kyllo v. United States, 533 U.S. 27, 40 (2001).

United States v. Jones, 132 S. Ct. 945, 963 (2012) (Alito, J., concurring).

United States v. Knotts, 460 U.S. 276, 278 (1983).

Whalen v. Roe, 429 U.S. 589, 591 (1977)